Exploring molecular mechanisms underlying the role of Non-typeable *Haemophilus influenzae* in COPD

by

Rajendra KC

B.Sc. Microbiology, M.Sc. Biochemistry

Tasmanian School of Medicine | College of Health and Medicine

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy (Medical Studies)

University of Tasmania

May 2020

Declaration of Originality

‘This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.’

Rajendra KC

14 April 2020

Authority of Access

The publishers of the papers in Chapters 1, 2 and 4 hold the copyright for that content and access to the material should be sought from the Journals. The remaining non-published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Rajendra KC

14 April 2020

Statement of Co-Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Mr Rajendra KC (candidate), Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Australia.

Associate Professor Ronan F. O'Toole, Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria, Australia, and Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Australia.

Associate Professor Gunasegaran Karupiah, Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Australia.

Professor Graeme R. Zosky, Tasmanian School of Medicine and Menzies Institute for Medical Research, College of Health and Medicine, University of Tasmania, Hobart, Australia.

Dr Isabel K. Hyland, School of Natural Sciences, College of Sciences and Engineering, University of Tasmania, Hobart, Australia.

Professor Jason A. Smith, School of Natural Sciences, College of Sciences and Engineering, University of Tasmania, Hobart, Australia.

Dr Shakti D. Shukla, Priority Research Centre for Healthy Lungs and School of Medicine and Public Health, Faculty of Health and Medicine, and Hunter Medical Research Institute, University of Newcastle, Callaghan, Australia.

Dr Nicholas M. Harkness, Department of Respiratory and Sleep Medicine, Royal Hobart Hospital, Tasmania, Australia.

Dr Sanjay S. Gautam, National Institutes of Health, Maryland, USA.

Mr Kelvin WC. Leong, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria, Australia.

Professor Philip M. Hansbro, Centre for Inflammation, University of Technology Sydney, Faculty of Science, Ultimo, Australia, and Centenary Institute, Sydney, Australia.

Dr Louise A. Cooley, Department of Microbiology and Infectious Diseases, Royal Hobart Hospital, Tasmania, Australia

Dr Julia Lachowicz, Department of Respiratory and Sleep Medicine, Royal Hobart Hospital, Tasmania, Australia.

Mrs Belinda McEwan, Department of Microbiology and Infectious Diseases, Royal Hobart Hospital, Tasmania, Australia

Dr Steve Petrovski, Department of Physiology, Anatomy and Microbiology, School of Life Sciences, La Trobe University, Victoria, Australia

Professor Eugene H. Walters, School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Australia.

Contribution of work by co-authors for each paper:

Paper 1: Located in Chapter 1

KC R, Shukla SD[1], Walters EH[2], O'Toole RF[3]. Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease. *Microbiology*. 2017;163(4):421-30. doi: 10.1099/mic.0.000434

Author contributions:

Conceived and designed experiment: Candidate, Author 3
Wrote the manuscript: Candidate, Author 3
Reviewed and edited the manuscript: Candidate, Author 1, Author 2, Author 3

Paper 2: Located in Chapter 1

KC R, Shukla SD[1], Gautam SS[2], Hansbro PM[3], O'Toole RF[4]. The role of environmental exposure to non-cigarette smoke in lung disease. *Clinical and Translational Medicine*. 2018; 5;7(1):39. doi: 10.1186/s40169-018-0217-2.

Author contributions:

Conceived and designed experiment: Candidate, Author 1, Author 4
Wrote the manuscript: Candidate, Author 1, Author 2, Author 4
Reviewed and edited the manuscript: Candidate, Author 1, Author 3, Author 4

Paper 3: Located in chapter 2

KC R, Zosky GR[1], Shukla SD[2], O'Toole RF[3]. A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells. *Biology Methods and Protocols*. 2018;3(1): bpy010. doi: 10.1093/biomethods/bpy010

Author contributions:

Conceived and designed experiment: Candidate, Author 3
Performed the experiments: Candidate
Analysed the data: Candidate
Supervision: Author 1, Author 3
Wrote the manuscript: Candidate, Author 3
Reviewed and edited the manuscript: Candidate, Author 1, Author 2, Author 3

Paper 4: Located in chapter 3

Hyland IK[1], KC R, Munday E[2], Kilah NL[3], O'Toole RF[4], Bissember AC[5], Smith JA[6]. Synthesis and biological evaluation of 2-aminothiophenes and 2-pyrrolothiophenes: identification of novel platelet-activating factor receptor (PAFR) antagonists. (under *preparation*)

Author contributions:

Conceived and designed experiment: Candidate, Author 1, Author 4, Author 5, Author 6
Performed the experiments: Candidate (biological assays), Author 1 (chemical synthesis)
Analysed the data: Candidate, Author 1, Author 2
Wrote the manuscript: Candidate, Author 1, Author 5, Author 6
Reviewed and edited the manuscript: Candidate, Author 1, Author 2, Author 3, Author 4, Author 5, Author 6

Paper 5: Located in chapter 4

KC R, Hyland IK[1], Smith JA[2], Shukla SD[3], Hansbro PM[4], Zosky GR[5], Karupiah G[6], O'Toole RF[7]. Cow Dung Biomass Smoke Exposure Increases Adherence of Respiratory Pathogen Nontypeable *Haemophilus influenzae* to Human Bronchial Epithelial Cells. *Exposure and Health*. 2020. doi: 10.1007%2Fs12403-020-00351-y

Author contributions:

Conceived and designed experiment: Candidate, Author 6, Author 7
Performed the experiments: Candidate

Analysed the data: Candidate
Supervision: Author 5, Author 6, Author 7
Contributed reagents/materials/analysis tools: Author 1, Author 2, Author 3, Author 4
Wrote the manuscript: Candidate, Author 7
Reviewed and edited the manuscript: Candidate, Author 2, Author 3, Author 4, Author 5, Author 6, Author 7

Paper 6: A number of next-generation sequencing protocols described in Chapter 5 have been published in the *Journal of Biological Methods*.

Gautam SS[1], KC R, Leong KWC[2], Aogáin MM[3], O'Toole RF[4]. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *Journal of Biological Methods*. 2019;6(1). doi: 10.14440/jbm.2019.276

Author contributions:

Conceived and designed experiment: Candidate, Author 1, Author 2, Author 4
Performed the experiments: Candidate, Author 1, Author 2
Analysed the data: Candidate, Author 1, Author 2, Author 3
Wrote the manuscript: Candidate, Author 1, Author 2
Reviewed and edited the manuscript: Candidate, Author 1, Author 2, Author 3, Author 4

Paper 7: A draft genome sequence of a representative isolate of nontypeable *Haemophilus influenzae* collected and sequenced for this work (described in Chapter 5) has been published in the *Microbiology Resource Announcement*.

KC R, Leong KWC[1], McEwan B[2], Lachowicz J[3], Harkness NM[4], Petrovski S[5], Karupiah G[6], O'Toole RF[7]. Draft genome sequence of an isolate of nontypeable *Haemophilus influenzae* from an acute exacerbation of chronic obstructive pulmonary disease in Tasmania. *Microbiology Resource Announcement.* 2020;9(19). doi: 10.1128/MRA.00375-20

Author contributions:

Conceived and designed experiment: Candidate, Author 6, Author 7
Performed the experiments: Candidate
Analysed the data: Candidate, Author 1
Contributed sample/reagents/epidemiological data: Author 2, Author 3, Author 4, Author 5
Wrote the manuscript: Candidate, Author 7
Reviewed and edited the manuscript: Author 5, Author 6, Author 7

Paper 8: Located in Chapter 5

KC R, Leong KWC[1], Harkness NM[2], Gautam SS[3], Cooley LA[4], Lachowicz J[5], McEwan B[6], Petrovski S[7], Karupiah G[8], O'Toole RF[9]. Whole-genome analyses reveal differences in gene content between nontypeable *Haemophilus influenzae* isolates from COPD compared to other clinical illnesses *Microbial Genomics.* 2020 Jul 24. doi: 10.1099/mgen.0.000405.

Author contributions:

Conceived and designed experiment: Candidate, Author 8, Author 9
Performed the experiments: Candidate
Analysed the data: Candidate, Author 1, Author 3
Supervision: Author 3, Author 8, Author 9
Contributed sample/epidemiological data: Author 2, Author 4, Author 5, Author 6, Author 7
Wrote the manuscript: Candidate, Author 9
Reviewed and edited the manuscript: Candidate, Author 1, Author 2, Author 7, Author 8, Author 9

We, the undersigned, endorse the above stated contribution of work undertaken for each of the published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Signed:	----------	----------
	Rajendra KC	A/Pr. Gunasegaran Karupiah
	Candidate	Primary Supervisor
	Tasmanian School of Medicine	Tasmanian School of Medicine
	University of Tasmania	University of Tasmania
Date:	14 April 2020	21 April 2020

Signed:	----------	----------
	A/Pr. Ronan O'Toole	Prof. James Vickers
	Co-Supervisor	Head of School
	School of Molecular Sciences	Tasmanian School of Medicine
	La Trobe University	University of Tasmania
Date:	21 April 2020	23 April 2020

Statement of Ethical Conduct

‘The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.’

Rajendra KC

14 April 2020

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors Associate Professor Ronan O'Toole and Associate Professor Gunasegaran Karupiah for all their time and advice to make my Ph.D. experience productive and stimulating. I would like to thank them for their insightful comments, guidance and suggestions that have been invaluable in the Ph.D. pursuit and also, helped shape my career as a research scientist.

I would like to thank my co-supervisor Professor Graeme Zosky for his thoughtful and constructive feedback. I am highly grateful for the collaborations with Dr Isabel Hyland and Professor Jason Smith. They synthesized chemical compounds for my research. I appreciate the collaboration with Dr Nicholas Harkness and Dr Louise Cooley at the Royal Hobart Hospital for the bacterial genome sequencing project. I would like to acknowledge the technical assistance of Belinda McEwan and Dr Julian Lachowicz with bacterial sample and data collection at the Royal Hobart Hospital. A special thanks to Dr Steve Petrovski for providing the Illumina platform for bacterial genome sequencing conducted at La Trobe University.

I am thankful to our past and present laboratory group members: Pratikshya Pandey, Dr Sanjay Gautam and Kelvin Leong for insightful discussions and laboratory assistance. I acknowledge the technical support of Dr David Steele and James Marthick for generating smoke extracts and sequencing bacterial genomes in the laboratory. I am also grateful to Conall O'Toole for assistance with the collection of wood and cow dung samples. I would like to thank my Graduate Research Co-ordinators, Dr Stephen Richards and Dr Bruce Lyons for their time and advice.

I gratefully acknowledge the financial support by the University of Tasmania, Tasmanian Graduate Research Scholarship (TGRS), School of Medicine Conference Travel Scheme and the Graduate Research Office Conference and Research Travel Scheme.

Words cannot express how grateful I am to my family for all their love, care and support. A special thanks to my wife Pratikshya Pandey who not only inspired and supported throughout the Ph.D. journey but also provided her valuable experimental insights. Thank you for always listening to me and cheering me up.

Rajendra KC

Thesis Structure

This Thesis consists of six Chapters:

- Chapter 1 is the thesis introduction and includes the literature review, aims and hypotheses and an outline of the study approach utilised to address each specific aim. It includes a large part of content from the review articles published in the peer-reviewed journals; *Microbiology* and *Clinical and Translational Medicine*.
- Chapter 2 is a results chapter that addresses the first aim of the study; To develop a simple smoke generation system to prepare quantifiable batches of cigarette, wood and cow dung smoke extracts and test them in vitro to assess their effects in inflammatory response in human bronchial epithelial cells. This chapter is presented as published in the peer-reviewed journal *Biology Methods and Protocols*.
- Chapter 3 is a results chapter that addresses the second aim of the study; To test the activity of chemical compounds towards adherence of nontypeable *H. influenzae* to cigarette smoke exposed bronchial epithelial cells. This chapter is formatted in a manuscript format that will be submitted to a peer-reviewed journal for publication.
- Chapter 4 is a results chapter that addresses the third aim of the study; To compare *in vitro,* the effects of cigarette, wood and cow dung smoke extract exposure in attachment of nontypeable *H. influenzae* to airway epithelial cells. This chapter is presented as published in the peer-reviewed journal *Exposure and Health.*
- Chapter 5 is a results chapter that addresses the fourth aim of the study; To investigate differences in gene content between nontypeable *Haemophilus influenzae* isolates from patients with COPD and other clinical illnesses. This chapter is presented as submitted to the peer-reviewed journal *Microbial Genomics.*
- Chapter 6 is the summary of the thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	analysis of variance
BEGM	bronchial epithelial cell growth medium
BIC	Bayesian information criterion
BSA	bovine serum albumin
CC	clonal complex
CDSE	cow dung smoke extract
CEACAM	carcinoembryonic antigen-related cell adhesion molecule
cgSNPs	core genome single nucleotide polymorphism
ChoP	phosphorylcholine
COPD	chronic obstructive pulmonary disease
CSE	cigarette smoke extract
CT	computed tomography
DAPC	discriminant analysis of principal components
DAPI	4′,6-diamidino-2-phenylindole dihydrochloride
DMSO	dimethyl sulfoxide
ECM	extracellular matrix
ETS	environmental tobacco smoke
FEV1	forced expiratory volume in one second

FITC	fluorescein isothiocyanate
FVC	forced vital capacity
G-CSF	granulocyte colony stimulatory factor
GO	gene ontology
GOLD	Global Initiative for Chronic Obstructive Lung Disease
HGT	horizontal gene transfer
ICAM-1	intercellular adhesion molecule-1
IL-6	interleukin-6
IL-8	interleukin-8
LOS	lipooligosaccharide
LPG	liquefied petroleum gas
LRI	lower respiratory infection
LRTI	lower respiratory tract infection
MLST	multi-locus sequence typing
MST	minimum spanning tree
NF-κB	nuclear factor kappa B
NTHi	nontypeable Haemophilus influenzae
OR	odds ratio
PAF	platelet activating factor

PAFR	platelet activating factor receptor
PAH	polycyclic aromatic hydrocarbon
pan-GWAS	pan genome wide association studies
PBS	phosphate buffered saline
PCA	principal component analysis
PM	particulate matter
PUMA	Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America
qPCR	quantitative polymerase chain reaction
RHH	Royal Hobart Hospital
RR	relative risk
SAR	structure-activity relationship
SNP	single nucleotide polymorphism
SOD	superoxide dismutase
TNF	tumour necrosis factor
WGS	Whole genome sequencing
WHO	World Health Organization
WSE	wood smoke extract

Abstract

Chronic obstructive pulmonary disease (COPD) is a severe and progressive condition characterised by persistent respiratory symptoms and airflow limitation. Around 300 million people in the world have COPD. It has emerged as the third leading cause of mortality, claiming 3.2 million lives worldwide in 2017. An acute exacerbation of COPD, a sudden worsening of respiratory symptoms, is a major cause of morbidity and mortality in COPD patients. Various factors, including biomass smoke exposure and infection with bacteria trigger COPD exacerbations. Nearly half of the world's population uses biomass fuel for cooking and heating and is therefore at risk of exposure to noxious particles released from the combustion of biomass fuel. Among respiratory bacteria, nontypeable *Haemophilus influenzae* (NTHi) is a key pathogen implicated in colonisation and damage of airways in COPD patients.

In this thesis, I assessed the impact of biomass smoke exposure on inflammation and adherence of NTHi to human bronchial epithelial cells. Due to the lack of a standard and easily accessible procedure for the preparation of biomass smoke, I first devised a simple, cost-effective, and reproducible method for the generation of biomass smoke extracts, in particular, cow dung and wood smoke extracts. Using this method, I generated quantifiable batches of biomass smoke extracts that were utilised for the assessment of cellular responses to different types of biomass smoke. I investigated the effect of biomass smoke extracts on human airway epithelial cells with respect to expression of a known receptor of NTHi, platelet-activating factor receptor (PAFR), and the pro-inflammatory cytokines interleukin 6 (IL-6) and IL-8, using quantitative polymerase chain reaction. In addition, I examined the response of bronchial epithelial cells to adherence of NTHi using immunofluorescence microscopy. I observed an increased inflammatory response in cells exposed to biomass smoke, characterised by induction of significant levels of IL-6 and IL-8 mRNA, in comparison to mock exposed cells. I

demonstrated a dose-dependent increase in NTHi adhesion to epithelial cells following exposure to biomass smoke extracts.

I further established an association between PAFR expression and the adhesion of NTHi in biomass smoke-exposed cells. Pre-treatment with a known PAFR antagonist, WEB-2086 inhibited biomass smoke-induced adherence of NTHi in airway cells in a dose dependent manner. In addition, pre-treatment of biomass smoke-exposed airway epithelial cells with a novel WEB-2086 analogue, C17 reduced NTHi adhesion in a dose-dependent manner.

I next assessed the genomes of 568 NTHi isolates, including 40 newly sequenced clinical isolates collected from patients with different diseases, including COPD. Phylogenetic analysis based on polymorphic sites on the core genome did not provide sufficient resolution to separate COPD strains from other clinical phenotypes, suggesting a similar set of core genes are present in all clinical NTHi isolates. I applied discriminant analysis based on the presence or absence profiles of accessory genes and found a clear distinction between COPD and other disease strains. I then applied a pan genome-wide association study approach to identify the accessory genes associated with COPD, which identified a set of accessory genes that regulate metabolic functions, such as the metabolism of organic acids and oxidation-reduction reactions that regulate cellular respiration to be significantly associated with COPD strains. This result suggests that NTHi associated with COPD may exhibit genetically encoded functional variances to isolates collected from other clinical illnesses.

In conclusion, this work advances our understanding of how biomass smoke could contribute to the development and progression of COPD and highlights the potential of PAFR as a therapeutic target for reducing the impact of hazardous biomass smoke exposure on respiratory health. Further, this thesis increases our understanding of gene sets shared by NTHi strains that survive and causes disease in the COPD lung.

1 Chapter 1: Introduction

A modified version of this chapter has been accepted for publication in peer-reviewed journals. This chapter includes the literature review, aims and hypotheses and an outline of the study approach utilised to address each specific aim. It includes a large portion of text that appears in the review articles published in the journals, *Microbiology* and *Clinical and Translational Medicine.*

KC R, Shukla SD, Walters EH, O'Toole RF. Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease. Microbiology. 2017;163(4):421-30 (see Appendix). Reused by permission from **Microbiology Society** Microbiology, © 2017.

KC R, Shukla SD, Gautam SS, Hansbro PM, O'Toole RF. The role of environmental exposure to non-cigarette smoke in lung disease. *Clinical and Translational Medicine*. 2018; 5;7(1):39 (see Appendix). Reused under the terms of the **Creative Commons CC BY** license, which permits unrestricted use, distribution, and reproduction in any medium.

1.1 COPD Definition

Global Initiative for Chronic Obstructive Lung Disease (GOLD) defined Chronic Obstructive Pulmonary Disease (COPD) as “a common, preventable and treatable disease that is characterised by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases.” The chronic airflow limitation, a characteristic of COPD, is caused by a mixture of small airway disease (chronic bronchitis) and parenchymal destruction (emphysema). The relative contribution of these vary from person to person, and also depends on the type of aetiology, for example tobacco smokers are found have more emphysema whereas biomass smoke exposed individuals exhibit more lung fibrosis and bronchiolitis, and less emphysema (Rivera et al., 2008, Camp et al., 2014). However, the mechanism underlying this differential effect of smoke exposure is yet to be elucidated.

1.2 Burden of the disease

COPD is a major public health problem. According to the Global Burden of Diseases, Injuries and Risk Factors Study 2017, around 300 million people in the world have COPD with 18.5 million new cases reported in 2017 (James et al., 2018). The number of years lived with disabilities (YLDs) from COPD increased from 19.8 to 30.6 million between 1990 and 2017, representing a 54.5% increase. In terms of mortality, COPD is the third leading causes of deaths globally that claimed approximately 3.2 million lives in 2017 (Roth et al., 2018). In Australia, the number of COPD cases has increased by 70% in the last 20 years with about 1.2 million people with COPD in 2017. In 2008, the total economic impact of COPD was estimated to be $98.2 billion attributed to substantial direct and indirect costs (Harper, 2013).

However, COPD patients included in all these epidemiological studies are diagnosed based on the presence of airflow obstruction alone, which is not sufficient to diagnose COPD. Symptoms of patients are not collected and considered in the diagnosis of COPD, which is one of the major limitations of these epidemiological studies. Spirometry-defined COPD may not represent the true COPD population as many patients with airflow obstruction in epidemiological studies have asthma or cannot be necessarily diagnosed with a disease as they have no symptoms. Therefore, global epidemiological data on COPD prevalence, incidence and mortalities should be cautiously interpreted.

1.3 Diagnosis and assessment

COPD is characterised by the presence of dyspnoea, chronic cough or sputum production, and/or a history of exposure to risk factors (Pauwels et al., 2001). The presence of respiratory symptoms and exposure to noxious stimuli along with the presence of persistent airflow limitation confirms the diagnosis of COPD. The airflow limitation is measured using spirometry, which is characterised by the presence of a post-bronchodilator ratio of forced expiratory volume in one second to forced vital capacity (FEV_1/FVC) < 0.70 (GOLD, 2018). COPD assessment is used to guide therapy and can be achieved by determining the level of airflow limitation, it's impact on patient's health status and the risk of future events (such as exacerbations, hospital admissions or death). Severity of airflow limitation is classified based on post-bronchodilator FEV_1 (i.e., spirometric grade as shown in Figure 1.1) (Han et al., 2013). Symptomatic assessment can be done by using either of the two measures; a simple measure of breathlessness such as the Modified British Medical Research Council (mMRC) questionnaire (Fletcher, 1960) or a short patient-completed questionnaire, the COPD Assessment Test (CAT), assessing the impact of COPD on health status (Jones et al., 2009).

Finally, the assessment of exacerbation risk defined as an acute worsening of respiratory symptoms is best predicted by the history of earlier exacerbations, including prior hospitalisations (GOLD, 2018). Thus, COPD patients are stratified into GOLD grades 1 to 4 based on severity of airflow limitation and groups A to D based on symptom burden and risk of exacerbation for effective therapy and better management (Figure 1.1). In newly diagnosed patients, decisions on the treatment and management of COPD are initially based on the ABCD classification.

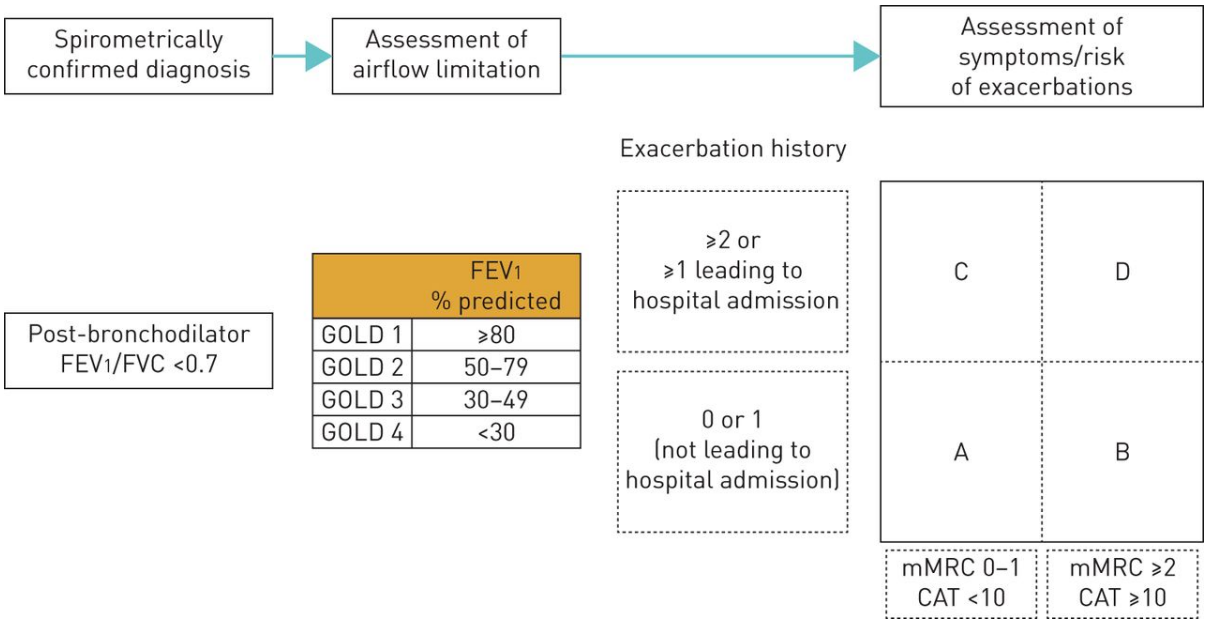


Figure 1. 1 COPD assessment tool.

FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; GOLD: Global Initiative for Chronic Obstructive Lung Disease; mMRC: modified Medical Research Council; CAT: COPD Assessment Test. Reproduced with permission from Global Initiative for Chronic Obstructive Lung Disease (GOLD, 2018), © 2020.

1.4 Aetiology of development and progression of COPD

1.4.1 Exposure to tobacco smoke

Tobacco smoke is one of the biggest public health threats, according to the World Health Organization (WHO) that kills more than 8 million people worldwide each year (WHO, 2019). More than one million of those deaths are attributed to second-hand or passive smoke exposure (WHO, 2019). Cigarette smoke contains more than 7300 chemicals, which have various toxic, mutagenic and carcinogenic effects (Rodgman and Perfetti, 2016). The major components of smoke that lead to many of the deleterious effects include nicotine, tar, ammonia, carbon monoxide, carbon dioxide, formaldehyde, acrolein, acetone, benzopyrenes, hydroxyquinone, nitrogen oxides and various pro-oxidative heavy metals (arsenic, cadmium, chromium, iron, lead, mercury, nickel and vanadium) (Rennard, 2004, Rodgman and Perfetti, 2016). Chronic exposure of the airways to this array of smoke-derived chemicals and particulates is the primary cause of pulmonary and immune dysfunction (Feldman and Anderson, 2013).

Tobacco smoke is a well-established primary risk factor for the development of COPD. The Surgeon General's Reports on Smoking and Health released in 1964 was the first to report a casual association of smoking with chronic bronchitis (Smoking, 1964). SmokeHaz, a collaborative project between the UK Centre for Tobacco and Alcohol Studies, the European Respiratory Society, and the European Lung Foundation conducted a systematic analysis that assessed the effect of active tobacco smoking on the risk of developing respiratory diseases (Jayes et al., 2016). Twenty-two studies included in this meta-analysis showed that active smoking is strongly associated with a 4.01-fold increased risk of COPD compared with non-smokers (RR, 4.01; 95% CI, 3.18-5.05). Risk of COPD was similar in between men and women (Jayes et al., 2016).

Besides active smoking, exposure to environmental tobacco smoke (ETS) has been associated with an increased risk of COPD. The 1972 Surgeon General's report showed the adverse effect of second-hand or passive smoke on respiratory health (Samet, 2014). Compared with non-smokers, exposure to environmental smoke for at least 1 h per day increases the risk of COPD by 44% (RR, 1.44; 95% CI, 1.02-2.01) (Berglund et al., 1999); however, a 13-year follow-up of the Cracow study did not find any consistent effect of passive smoking on the development of COPD in non-smoking adult population (Krzyzanowski et al., 1986). Furthermore, a prospective study conducted in the USA found that adults exposed to ETS during both childhood and adult life were 1.72 times more likely to develop COPD in their adulthood (RR, 1.72; 95% CI, 1.31-2.23); however, no increased risk was seen in those only exposed during childhood (RR, 1.09; 95% CI, 0.69-1.79) (Robbins et al., 1993).

1.4.2 Exposure to biomass smoke

The WHO estimates that nearly 4 million people die prematurely from illness attributable to household biomass smoke exposure annually, of which 55% deaths are linked to respiratory diseases, including pneumonia (lower respiratory tract infections), COPD and lung cancer (WHO, 2018c). Nearly half of the world's population rely on coal and biomass, such as wood, animal dung, and crop residues as a primary fuel source for cooking and heating purposes (WHO, 2018c). The proportion of households using clean fuels (liquid petroleum gas, biogas and electricity) varies considerably across the globe (and even on the same continent) (WHO, 2014). Access to clean fuels is especially limited to a relatively small proportion of the populations in low- and middle-income countries (Figure 1.2). Consequently, populations residing in rural or remote areas of developing countries are heavily reliant on biomass fuel for cooking or heating purposes (Figure 1.2) (Hall et al., 2013).

The particular concern with energy from biomass fuel is the use of inefficient stoves for combustion, which generates toxic gases like carbon monoxide and nitrogen oxides; suspended particulate matter containing volatile organic compounds (VOCs) such as methane, aldehydes, benzene and its derivatives; and polycyclic aromatic hydrocarbons (PAHs) like benzo[*a*]pyrene and anthracene (Leavey et al., 2017). Particulate matter (PM) with an aerodynamic diameter of <2.5 microns ($PM_{2.5}$) is light and can remain suspended in the air for longer periods (Xing et al., 2016). $PM_{2.5}$ can be inhaled deep into the lungs, and has been linked to oxidative stress and inflammation induced damage to the respiratory system (Xing et al., 2016). Moreover, in developing countries, the practice of indoor cooking in housing with poor air ventilation potentially exposes women and children to $PM_{2.5}$ up to the levels that are 1,000 times higher than the threshold recommended by WHO (25 $\mu g/m^3$) (WHO, 2006, Saksena S, 2003).

Wood is the most common biomass fuel, however, use of animal dung, such as from cows, sheep or horses, as a source of fuel is also widespread, especially in rural areas of low- and middle-income countries, including India, Nepal, and sub-Saharan Africa due to its availability in areas with limited vegetation and its lower cost (Pant, 2012). Animal dung is an inefficient biofuel in terms of heat production and burns faster as compared to wood (Pahla et al., 2017). As a result, relative to wood smoke, combustion of animal dung produces more particulate matter (23% more $PM_{2.5}$ per kilogram), toxic byproducts, such as PAHs and endotoxins, and oxidizing species such as redox active metals (copper and iron) and quinones (Mudway et al., 2005, Sussan et al., 2014).

Nonetheless, exposure to biomass smoke is not exclusively an issue in low- and middle-income countries. Use of indoor wood fires for heating purposes and for adding flavor during cooking processes, such as barbecuing and wood-smoking of meats, are becoming more popular in high-income countries, thus, increasing biomass smoke exposure (Naeher et al., 2007). In 2014,

in a survey carried out in Australia, approximately 10% of households used wood as the main source of heating (Statistics, 2014). However, in developed countries, use of biomass fuel is primarily seasonal and exposure is largely limited by better ventilation (Rogalsky et al., 2014). Besides indoor biomass smoke exposure, people in developed countries such as Australia, Canada and the USA are also exposed to outdoor biomass smoke from frequent bushfires (Reisen and Brown, 2009). Importantly, individuals exposed to biomass smoke are reported to have respiratory symptoms and are more likely to have reduced lung function (da Silva et al., 2012, Mukherjee et al., 2014, Regalado et al., 2006).

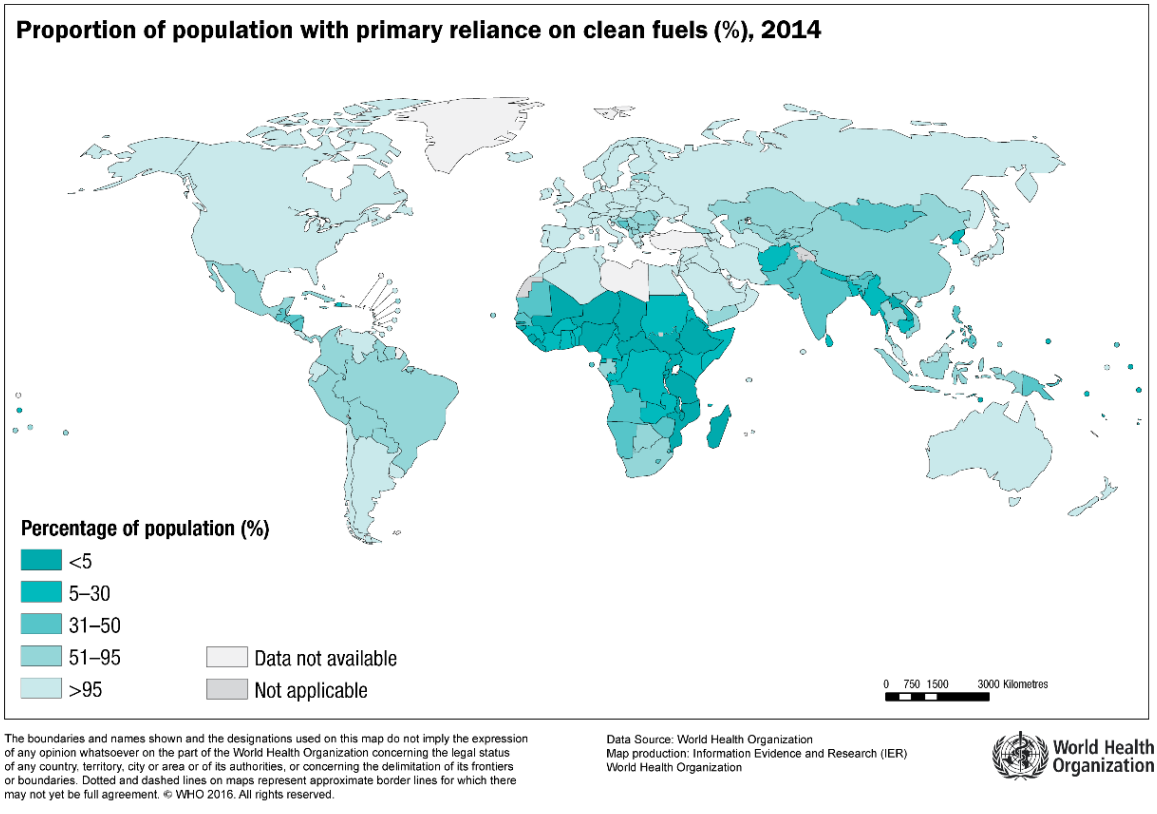


Figure 1.2 Global use of clean fuels in 2014, by the World Health Organization.

Countries with the lowest (<5%) and the highest (>95%) proportion of people using clean fuels as the primary domestic source of energy are shaded dark and light blue, respectively. Reproduced with permission from World Health Organization (WHO, 2014), Copyright (2016).

Worldwide, more than 80% of deaths caused by COPD occurred in low- and middle-income countries in 2016 (WHO, 2018b). As stated in Section 1.2 above, it is important to note that global epidemiological data on COPD burden need to be carefully interpreted due to differences in basis of diagnosis. More than 45% of COPD patients in this region are found to have never smoked (Table 1.1). The existing literature provides strong evidence that smoke from biomass fuels is an independent risk factor for the development of COPD, particularly in low- and middle-income countries where the reliance on biomass fuel is still very high. The noxious particles in the biomass smoke induce an inflammatory response through upregulation of pro-inflammatory cytokines such as interleukin 6 (IL-6), tumour necrosis factor (TNF), and granulocyte colony stimulatory factor (G-CSF), recruitment of immune cells, such as macrophages and neutrophils and upregulate gelatinases (matrix metalloproteinase 2 and 9) and epithelial-mesenchymal transition (EMT), thereby reducing the lung function and contribute to the onset/progression of COPD (Zou et al., 2014, Olloquequi and Silva, 2016, Ramos et al., 2017). Da Silva et al, in a case-control study in Brazil, evaluated the effect of exposure to biomass combustion $PM_{2.5}$ on lung function (da Silva et al., 2012). There was a significant loss in pulmonary function in non-smoker biomass users compared to non-smoker liquefied petroleum gas (LPG) users (forced expiratory volume in one second to forced vital capacity ratio (FEV1/FVC) 0.79 versus 0.85, p <0.05, respectively). In addition, the pulmonary function was negatively correlated with the duration and the concentration of $PM_{2.5}$ exposure (FEV_1/FVC: r = -0.63, p <0.05 and -0.52, p <0.05 respectively) (da Silva et al., 2012). Furthermore, women living in rural Mexico exhibited a minor reduction in FEV_1/FVC ratios in those using biomass fuel compared to clean gas users (0.80 *vs* 0.83, p = 0.03, respectively) (Regalado et al., 2006).

In a recent nationwide cross-sectional study conducted in 50,991 individuals from ten provinces of China, more than half of the COPD patients were never smokers (Wang et al., 2018). Importantly, the proportion of never smokers in female COPD patients was markedly high (91.7%). Additionally, individuals exposed to indoor biomass smoke were 62% more likely to develop COPD than the unexposed people using clean fuel sources (age and sex adjusted OR=1.6, $p = 0.003$) (Wang et al., 2018).

A meta-analysis of 11 cross-sectional and four case-control studies covering a wide range of countries identified household biomass smoke exposure as an independent risk factor for developing COPD in both men (OR 4.30, 95% CI 1.85-10.01) and women (OR 2.73, 95% CI 2.28-3.28), and in both the Asian population (OR 2.31, 95% CI 1.41-3.78) and the non-Asian population (OR 2.56, 95% CI 1.71-3.83) (Hu et al., 2010). This finding is supported by another recent cross-sectional study, the PUMA (Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America), which assessed 1740 individuals from multiple nations in South America, who were at greater risk for developing COPD (Montes de Oca et al., 2017). The PUMA study reported that individuals exposed to household biomass smoke are twice as likely to develop COPD than unexposed people (adjusted OR 2.28, 95% CI 1.18–4.41) (Montes de Oca et al., 2017). Moreover, a recent systematic review of 24 epidemiological studies revealed that household biomass smoke exposure was associated with COPD development in both urban (OR 1.6, 95% CI 1.2-2.0) and rural women (OR 2.0, 95% CI 1.5-2.8) (Sana et al., 2018).

In terms of COPD phenotypes, biomass smoke exposed individuals were found to have more lung fibrosis and bronchiolitis and less emphysema in contrast to significantly higher proportion of emphysema in tobacco smokers (radiologist CT score- a measure of the extent of emphysema based on computed tomography (CT) at inspiration and expiration) 0.7 versus

2.3, $p = 0.001$; emphysema on CT 19% versus 27%, $p = 0.046$) (Camp et al., 2014, Rivera et al., 2008). This finding is further supported by Fernandes et al., who reported significantly fewer women with emphysema in biomass exposed women than in women who were tobacco smokers (PRM^{Emph} (parametric response mapping, an image tool used to quantify emphysema based on paired CT images at inspiration and expiration) 1.84% (0.69 ± 3.72%) versus 9.85% (2.40–16.34%); $p = 0.001$) (Fernandes et al., 2017).

Table 1. 1 Distribution of never smokers with COPD and risk factors for the development of COPD in never smokers.

Study center & design	Participants	Age (years)	Proportion of never-smokers among COPD patients (%)			Risk factors for COPD in never-smokers	References
			Overall	Male	Female		
Multinational (35 centres, 16 countries; ECRHS)	17,966	20-44	17.0	13.4	21.6	Occupational exposure to vapours, gas, dust, or fumes	(Cerveri et al., 2001)
Malataya, Turkey (CS)	1,160	>18	22.5			Exposure to biomass Smoke	(Gunen et al., 2008)
China (CPH; nationwide CS)	50,991	>20	50.5	10.1	91.7	Exposure to biomass fuel smoke and $PM_{2.5}$, parental history of respiratory disease	(Wang et al., 2018)
South Africa (nationwide survey)	13,826	>18	47·6	24.8	61.0	Biomass fuel, occupational exposure, history of pulmonary tuberculosis	(Ehrlich et al., 2004)
Maswa, Tanzania (CS)	869	>35	62.1	-	-		(Magitta et al., 2018)
Västra Götaland and Norrbotten, Sweden (CS)	1,839	21-78	21	-	-	Occupational exposure to gas, dust and fumes	(Hagstad et al., 2015)
Copenhegen, Denmark (CS)	68,501	20-100	22.3	18.9	25.5		(Thomsen et al., 2013)
Multinational (12 countries; population-based survey)	73,745	>40	36 (24 USA-64 Mexico)	21	49		(Landis et al., 2014)

ECRHS, European Community Respiratory Health Survey; CS, cross-sectional study; CPH, China Pulmonary Health; COPD, chronic obstructive pulmonary disease

1.4.3 Environmental exposure to noxious particles and gases

High levels of air pollution have also been implicated as a risk factor for the development of COPD, although it is less potent than active smoking (Schikowski et al., 2014). Approximately 25–60% of COPD patients are never-smokers in developing countries (Salvi and Barnes, 2009, Wang et al., 2018, Magitta et al., 2018). Increasing population, urbanization, economic profile and pollution are several factors that contribute substantially to the COPD burden (Chan et al., 2017). The role of outdoor air pollution (including traffic-related fine particulate matter) as a causative factor for airflow limitation is gaining attention in recent times, due to increased vehicular pollution, as well as industrialization of the two most populated countries, China and India (Eisner et al., 2010).

Exposure to fine particulate matter ($PM_{2.5}$) is considered to be the most health-damaging, as they can penetrate the deep lung tissue and initiate deleterious effects on the airway, including but not limited to airway oxidative stress, pulmonary and systemic inflammation, ciliary dysfunction, amplification of infections, and increases in bronchial reactivity (Donaldson et al., 2002). The main components of PM_{10} are sulfate, nitrates, ammonia, sodium chloride, black carbon, mineral dust and water (WHO, 2018a).

Several investigators have assessed the effect of air pollution as a potential risk factor for COPD. In children and young adults, cross-sectional studies have shown a relationship between higher outdoor pollutant levels (especially traffic related pollution) and lower lung function (Holguin et al., 2007, Janssen et al., 2003). Kulkarni et al. reported a likely causative dose-dependent inverse association between the carbon content (as a biomarker of particulate matter exposure, PM_{10}) of airway macrophages with lung function in children (Kulkarni et al., 2006). Furthermore, exposure to traffic-related pollution, characterised by the residential distance

from a highway, was associated with impaired lung growth and lung function deficits at 18 years of age (Gauderman et al., 2007). Also, higher traffic density and proximity to highways was significantly associated with lower lung function (FEV_1) and FVC, but only in females (Kan et al., 2007), who are also at a greater risk of developing COPD than for those living farther away (Schikowski et al., 2005).

Long-term exposure to airborne particles and particulate matter is significantly associated with the risk of premature death and acute care hospitalizations, especially in patients with severe disease (Wordley et al., 1997, Zanobetti et al., 2008). Moreover, daily variation in exposure to outdoor air pollution (mainly the particulate matter) significantly correlates with acute exacerbations of COPD (Sunyer, 2001). The mechanisms that underlie obstruction due to air pollution are likely to be the same or/and similar to those due to cigarette smoking, but we do not yet have evidence for this, despite the detrimental effects of air pollution on lung health.

1.4.4 Occupational exposure to dusts and fumes

Several longitudinal studies have shown an association between certain dusty occupational exposures and COPD, i.e., coal mining (Attfield, 1985), gold mining (Gantt and Lincoln, 1988), work related to tunnel-construction (Ulvestad et al., 2001), low levels of concrete dust containing crystalline silica exposure (concrete production industries) (Meijer et al., 2001), exposure to cotton in textile industries (Niven et al., 1997), workers exposed to welding fumes (Bradshaw et al., 1998), grain handlers and postal workers (exposure to endotoxins) (Schwartz et al., 1995) and animal feed industry (Post et al., 1998). Moreover, chronic exposure to metallic dust (primarily, cobalt and chromium) was found to be associated with deterioration of lung function (Hamzah et al., 2016). In addition, exposures to chemical vapours, irritants

and fumes can also contribute to accelerated loss of airflow (Boschetto et al., 2006). Another study involving railroad workers reported a positive association between COPD mortality and occupational exposure to diesel exhaust (Hart et al., 2006). A population-based study found positive associations between several occupational exposure measures (mineral dusts, metal dusts and fumes, organic dusts, irritant gases or vapours, sensitizers, organic solvents, diesel exhaust, and environmental tobacco smoke) with COPD, among both ever-smokers and never-smokers (Weinmann et al., 2008).

Experimental studies in animal models have demonstrated that exposures to several agents, such as sulphur dioxide, mineral dusts, vanadium and endotoxin, are capable of inducing chronic obstructive bronchitis (Boschetto et al., 2006). Intratracheally instilled silica (quartz) produces airflow obstruction (functional change), which correlates with the presence of both emphysema and small-airway lesions (Wright et al., 1988). Inorganic dusts containing silica are also associated with neutrophil and macrophage accumulation and morphological changes in the rat lung (Wright et al., 1988). These morphological changes in small airways and lung parenchyma were similar to those in rats treated with elastase, which represents a well-established model of experimental COPD (Churg et al., 1989).

1.4.5 Genetic risk factors

Besides cigarette smoking and environmental factors, there are additional genetic factors that contribute to the development of COPD. Previous epidemiological and clinical data have demonstrated a hereditary contribution to the development of COPD, as evidenced by an aggregation of COPD in families (Tager et al., 1976, Tager et al., 1978, Higgins and Keller, 1975, Cohen et al., 1977). The most important and the first gene identified to be associated with

COPD was SERPINA1, which encodes alpha-1 antitrypsin (AAT) (Demeo et al., 2006). Since 2006 when Dewan et al. performed a GWAS and identified a variant to be significantly associated with macular degeneration. GWAS have become highly effective in identifying genes/variants associated with lung function or COPD phenotype (Dewan et al., 2006). In 2010, the Charge Consortium performed a meta-analysis of GWAS for lung function measures, FEV1 and FEV1/FVC, and identified nine loci associated with the lung function: HHIP, GPR126, ADAM19, AGER-PPT2, FAM13A, PTCH1, PID1, HTR4 and INTS12-GSTCD-NPNT (Hancock et al., 2010). In the same year, the SpiroMeta study, a meta-analysis of GWAS, discovered an additional five loci associated with FEV1 or FEV1/FVC (Repapi et al., 2010). A recent published large-scale GWAS on lung function identified 97 loci that were significantly associated with lung function, of which 30 were found to be significantly associated with COPD susceptibility (Wain et al., 2017). Many of these variants may act, at least in part, via effects on lung development, elastic fibres and epigenetic regulation pathways (Wain et al., 2017).

1.4.6 Bacterial infections

The progressive course of COPD is often aggravated by exacerbations, which is the major cause of morbidity and mortality in COPD patients (Schmidt et al., 2014). The majority of acute exacerbations of COPD are produced by microbial infections (Shimizu et al., 2015, Sethi, 2010). Respiratory bacteria, viruses and atypical bacteria are the most common pathogens implicated in exacerbations of COPD (Hewitt et al., 2016, Sethi and Murphy, 2001). Bacteria are isolated from sputum in 40-60% of exacerbations of COPD (Sethi, 1999). Of which, nontypeable *Haemophilus influenzae* (NTHi), *Moraxella catarrhalis*, and *Streptococcus pneumoniae* are the most common causes of exacerbations. These microbes are found in

abundance, colonising the oropharynx and the upper respiratory tract, where they are harmless and do not result in disease (Wilson and Hamilos, 2014). However, in chronic conditions such as COPD when innate lung defence is impaired and upon exposure to risk factors such as tobacco and biomass smokes, these pathobionts colonise and establish infection in the lower airways (Herr et al., 2009, Lange, 2009).

Exposure to cigarette smoke increases risk of respiratory infection by these pathobionts (Murphy, 2006, Nuorti et al., 2000, Fischer et al., 1997). Compared to non-smokers, smokers are found to have more pathobionts in the nasopharyngeal microflora (Brook and Gober, 2005). The same group, more recently, conducted a prospective study to determine the effect of smoking cessation on the composition of nasopharyngeal microflora and concluded the reversion of pathobiont load to normal levels after quitting of smoking (Brook and Gober, 2007). Active smoking in general doubles the risk of infection (Trosini-Desert et al., 2004). The effect of smoking is higher in patients with COPD where it increases the risk of pulmonary infection by 3.17-fold as compared to non-smoking COPD (OR, 3.17; 95% CI, 2.5-8) (Zalacain et al., 1999). Besides active smoking, exposure to ETS is an important modifiable risk factor for lower respiratory tract infection (LRTI), particularly in children (Zar and Ferkol, 2014). An updated systematic review reported smoking by both parents (OR, 1.62; 95% CI, 1.38–1.89), father (OR, 1.22; 95% CI, 1.10–1.35), or a household member (OR, 1.54; 95% CI, 1.40–1.69) significantly increased the risk of LRTI (Jones et al., 2011). In addition, antenatal maternal smoking has been found to an important risk factor associated with LRTI (incidence rate ratio, 1.62; 95% CI, 1.14–2.30) (Vanker et al., 2017).

In addition to tobacco smoke, exposure to biomass smoke is also an important risk factor for respiratory infections. A recent systematic review of 77 studies from 39 low- and middle-income countries evaluated the risk factors for mortality from acute lower respiratory infections

in children under five years of age and found that biomass smoke exposure is significantly associated with an increased risk of death from LRTI (OR 3.0, 95% CI 2.1-4.3) (Sonego et al., 2015). However, few studies have evaluated an association between biomass smoke exposure and risk of lower respiratory tract infection in the adult population. Ezzati et al. evaluated the risk of LRTI in 229 individuals between 5 and 49 years of age in Central Kenya upon exposure to biomass fuel derived PM_{10} (Ezzati and Kammen, 2001). The risk of LRTI was positively correlated to the level of PM_{10} exposure, however, the odds ratio was significant only above 2000 µg/m3 (adjusted OR 3.3, 95% CI 1.1-9.9) (Ezzati and Kammen, 2001). In older adults aged 65 and above, exposure to household biomass smoke was associated with risk of hospitalisation due to pneumonia (adjusted OR 3.3, 95% CI 1.6-6.9) (Loeb et al., 2009).

1.5 Mechanism underlying the effects of biomass smoke exposure

Despite the overwhelming burden of the disease, only a handful of studies have explored mechanisms in immunomodulatory effects of biomass smoke exposure. Sussan et al. (2014) explored the mechanisms of pulmonary responses in mice after acute (6 and 24 h) or sub-chronic (3 times a week for 8 weeks) exposure to wood or cow dung PM (Sussan et al., 2014). Acute exposure to wood smoke elicited a stronger pro-inflammatory response, as indicated by increased expression of G-CSF, keratinocyte chemoattractant (KC), C-X-C motif chemokine 10 (CXCL10), IL-6, TNF and interleukin 12 p70 subunit (IL12p70) (Sussan et al., 2014). The induction of pro-inflammatory cytokine expression was higher with cow dung than wood smoke PM (Sussan et al., 2014). On the other hand, sub-chronic exposures exhibited differences in pulmonary response, where wood smoke elicited an eosinophilic response in contrast to neutrophilic response induced by cow dung smoke (Sussan et al., 2014). The inflammatory response elicited by both wood and cow dung smoke was mediated via nuclear

factor kappa B (NF-κB) signaling (Sussan et al., 2014). In addition, in an *in vitro* study, c-Jun N-terminal kinase-activator protein-1 (JNK-AP-1) signaling, and not NF-κB pathway, was found to be involved in mediating inflammatory responses in primary human small airway epithelial cells upon exposure to dung biomass smoke (McCarthy et al., 2016a).

Moreover, in a case-control study conducted in female Indians with 142 biomass users and 126 age-matched LPG users, a significant neutrophilic inflammatory response was observed in biomass using group (Banerjee et al., 2012). Also, compared to LPG users, reactive oxygen species (ROS) generation by leukocytes (in both blood and sputum) and the systemic level of antioxidant enzyme superoxide dismutase (SOD) were higher and lower, respectively, in women using biomass fuel (Banerjee et al., 2012). A similar increase in systemic oxidative stress was reported in female rats exposed to biomass smoke, as indicated by increased plasma level of malondialdehyde and reduced level of SOD (Hu et al., 2013). Several *in vitro* studies reported depletion of antioxidants, such as ascorbate, urate and reduced glutathione in respiratory tract lining fluid when incubated with wood and animal dung smoke extracts (Mudway et al., 2005, Kurmi et al., 2013). Therefore, biomass smoke exposure induces inflammation and oxidative stress mediated lung damage potentially contributing to the development/progression of COPD.

Biomass smoke exposure is also implicated in increased susceptibility to bacterial infection through several mechanisms, including alterations in alveolar macrophage phagocytosis and/or upregulation of host surface receptors on the respiratory epithelium (Rylance et al., 2015, O'Toole et al., 2016). In an *in vitro* study, the phagosomal function of wood smoke particles exposed human alveolar macrophages was tested by uptake of fluorescently-labelled beads, *S. pneumoniae* and *Mycobacterium tuberculosis* (Rylance et al., 2015). Wood smoke exposed macrophages demonstrated reduced phagocytosis of fluorescent beads, *S. pneumoniae* and *M.*

tuberculosis with a negative linear correlation between macrophage particulate content and phagocytosis (Gordon et al., 2014). Furthermore, the oxidative stress is believed to upregulate intercellular adhesion molecule-1 (ICAM-1), and platelet activating factor receptors (PAFR) allowing attachment and invasion of respiratory bacteria, including *S. pneumoniae, H. influenzae* and *P. aeruginosa,* which are also major bacterial pathogens in COPD (Roebuck et al., 1995, Ishizuka et al., 2001, KC et al., 2017). Although airway ICAM-1 and PAFR expression were markedly upregulated in tobacco smokers and COPD patients, further studies are warranted to demonstrate the effect of biomass smoke exposure on inducing the expression of these host surface receptors (Shukla et al., 2017, Lopez-Campos et al., 2012, Shukla et al., 2016b). More mechanistic research is therefore needed to understand the cellular and molecular responses to biomass smoke, including animal dung and wood smoke. However, the lack of standardized experimental approaches for the preparation of re-usable biomass smoke extracts has hindered research in this area. Therefore, in this study, we aimed to develop a method to generate batches of biomass smoke extracts that could be preserved for longer periods and may be utilised in multiple exposure experiments, thus minimizing inter-assay variations. In our opinion, this will enable further research on mechanistic role of smoke extracts in the inflammatory response and pathogenesis of respiratory diseases including airway infections and COPD.

1.6 Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease

Bacteria utilise a wide variety of molecules on host surfaces as docking sites for tissue adhesion and host colonisation (Figure 1.1). Of particular interest, are the extracellular matrix (ECM) proteins, cell adhesion molecules (*e.g.* integrins, cadherins), and PAFR, which upon

stimulation by certain environmental and/or immunogenic insults, undergo transient upregulation (Figure 1.3). This enhances bacterial adherence and subsequent tissue invasion (Singh et al., 2012, Hauck, 2002, Shukla et al., 2015).

ECM, the acellular proteinaceous part of animal connective tissue, constitutes an anchoring platform for epithelia, designated the basement membrane (BM), and also surrounds blood capillaries and neurons (Chagnot et al., 2012). It consists of collagen, elastin, fibrillin, laminin (Ln), fibronectin (Fn), vitronectin, thrombospondin, proteoglycans and hyaluronic acid. Besides its ubiquitous distribution, ECM biosynthesis is significantly enhanced following viral infections (*e.g.* influenza A virus) and traumatic injury (*e.g.* ligament rupture) as a natural response to tissue repair, and is therefore, an attractive target for adherence and invasion by several bacterial pathogens, such as *N. meningitidis*, *S. pneumoniae*, and NTHi (Singh et al., 2012, Su et al., 2016, Eberhard et al., 1998, Li et al., 2015, Pracht et al., 2005, Neurath, 1993).

In addition to ECM components, cell adhesion molecules, including integrins, cadherins, selectins, and members of the immunoglobulin superfamily of cell adhesion molecules (IgCAMs), are also involved in bacterial adhesion (Hauck, 2002, Grigg et al., 2012). Integrins are heterodimeric (composed of two subunits, α and β) transmembrane glycoproteins that attach cells to extracellular matrix proteins of the basement membrane or to ligands on other cells (Barczyk et al., 2010, Plow et al., 2000). Several bacteria bind to integrins directly whereas others engage them via ECM proteins, such as fibronectin and collagen. Bacterial-integrin binding can trigger host intracellular signalling leading to actin cytoskeleton remodelling and subsequent bacterial invasion (Hauck, 2002).

IgCAMs including, carcinoembryonic antigen-related cell adhesion molecule (CEACAM) and ICAM-1, constitute the other major class of host cell receptors utilized by bacterial adhesion systems (Hauck, 2002). The CEACAM family is a group of highly glycosylated intercellular

adhesion molecules involved in signalling events that mediate key cellular processes that include cell adhesion, proliferation, differentiation and tumour suppression (Kuespert et al., 2006). They comprise an N-terminal Ig variable (IgV)-like domain followed by up to six Ig(C) domains. Twelve different CEACAM proteins have been identified in humans to date with CEACAM-1, CEACAM-5 and CEACAM-6 found in epithelial cells, and CEACAM-3 present exclusively in granulocytes (Beauchemin et al., 1999).

ICAM-1 (CD54) is a cell surface glycoprotein that serves as a counter-receptor for leucocyte β2 integrins, lymphocyte function associated antigen (LFA-1) (CD11a/CD18) and macrophage adhesion ligand 1 (Mac-1) (CD11b/CD18) (Staunton et al., 1988). It is constitutively expressed in low levels on endothelium, fibroblasts and various epithelia (*e.g.* bronchial, intestinal, and urinary tract), however, its expression is markedly upregulated at sites of inflammation (Tosi et al., 1992, Frick et al., 2000, Huang et al., 1996, Chan et al., 1995, Shukla et al., 2017). Interactions between ICAM-1 and β2 integrins are known to have a central role in mediating leukocyte recruitment in the inflammatory response. This may lead to partial protection from invading pathogens but may also result in neutrophil-induced chronic epithelial injury (Albelda et al., 1994, Sumagin et al., 2014). A sustained inflammatory process may further upregulate adhesion receptors.

Finally, the other class of host cell receptor, PAFR is a G-protein-coupled 7-transmembrane domain receptor, physiologically recognized by a phospholipid, platelet activating factor (PAF) (Ishii et al., 2002, Honda et al., 1991). PAFR plays a role in a wide range of biological processes such as vasodilation, cell proliferation, angiogenesis, and regulation of the inflammatory response (Ishii et al., 2002). Also, over the last decade, there has been increasing evidence emerging that PAFR is a major epithelial receptor used by specific respiratory and intestinal bacteria for adhesion to and also invasion of host epithelium (Shukla et al., 2015, Keely et al.,

2010). Moreover, PAFR expression is inducible and is directly linked to increased susceptibility to infection by both Gram-positive and Gram-negative bacteria (Keely et al., 2010, Grigg, 2012).

S. pneumoniae and NTHi, along with some strains of *Pseudomonas aeruginosa*, a major bacterial pathogen in cystic fibrosis, share a common adhesin, known as phosphorylcholine (ChoP), in their cell wall (Swords et al., 2000, Grigg et al., 2012, Barbier et al., 2008). ChoP mimics PAF, which is the natural arachidonic acid derived ligand for PAFR expressed on bronchial and alveolar epithelial cells. PAFR has been shown to be upregulated in the airway epithelium of both the large and small airways in smokers and COPD patients (Shukla et al., 2016b, Shukla et al., 2014). Furthermore, elevated PAFR expression correlated with higher levels of adhesion to human bronchial epithelial cells by NTHi and *S. pneumoniae*, the major causative agents of acute exacerbations of COPD (Shukla et al., 2016a).

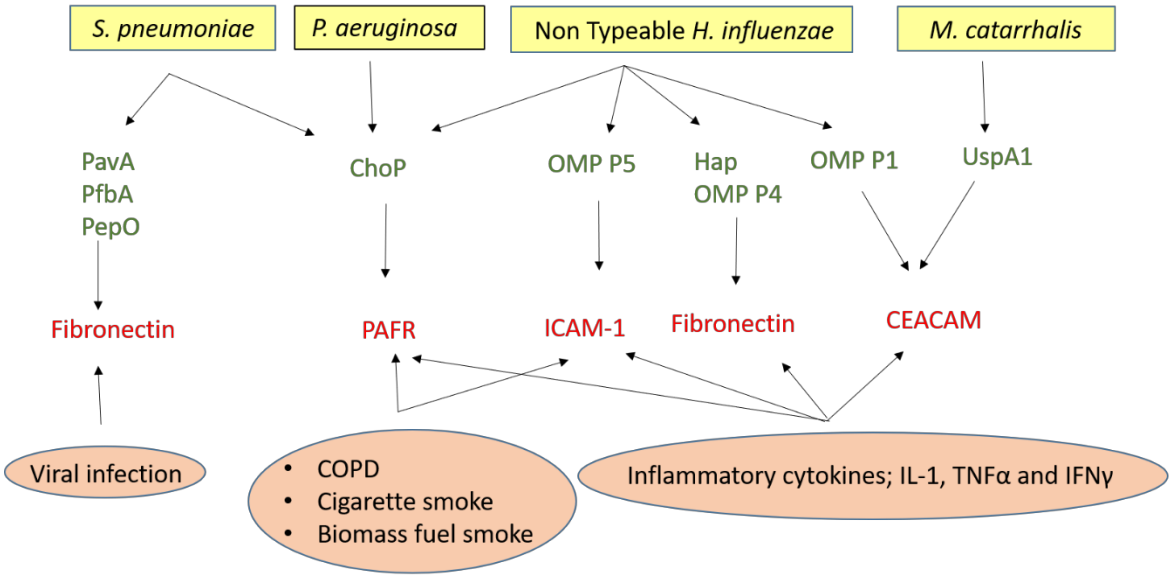


Figure 1.3 Temporal upregulation of host surface receptors in the respiratory system.

Host surface receptors in the respiratory tract are upregulated in response to viral infection, exposure to cigarette and biomass fuel smoke, as well as inflammatory cytokines. Bacterial pathogens including non-typeable *Haemophilus influenzae*, *Streptococcus pneumoniae*,

Pseudomonas aeruginosa and *Moraxella catarrhalis* exploit the upregulated receptors for attachment via their cognate adhesins. PavA, pneumococcal adhesion and virulence A; PfbA, plasmin-fibronectin binding protein A; PepO, pneumococcal endopeptidase O; TGFβR, transforming growth factor-beta receptor; COPD, chronic obstructive pulmonary disease; PAFR, platelet activating factor receptor; Hap, haemophilus adhesion and penetration; OMP, outer membrane protein; CEACAM, carcinoembryonic antigen cell adhesion molecule; UspA1, ubiquitous surface protein A1; IL-1, interleukin 1; TNFα, tumour necrosis factor-alpha; IFNγ, interferon-gamma. Yellow, blue and orange coloured text boxes represent bacteria, inhibitors and factors affecting expression of host cell receptors, respectively. Green and red coloured texts represent bacterial adhesins and their cognate host cell receptors, respectively.

1.7 Targeting platelet activating factor receptor (PAFR) in respiratory infections and inflammatory conditions

As discussed above, PAFR is utilised for adhesion to human bronchial epithelial cells by NTHi and *S. pneumoniae* which are major bacterial pathogens in COPD exacerbations. PAFR antagonists, such as Ginkgolide-B (BN52021), CV-3988, PCA-4248, CAS-99103-16-9, and WEB-2086 (Apafant), have been reported to block the attachment of bacterial pathogens to respiratory epithelium (Hergott et al., 2015, Grigg et al., 2012, Negro Alvarez et al., 1997, Barbier et al., 2008). Of these, WEB-2086 has recently been demonstrated to significantly inhibit both NTHi and *S. pneumoniae* adherence to cigarette smoke exposed human bronchial epithelial cells (Shukla et al., 2016a). This underscores the possibility that PAFR, that are upregulated in smokers and COPD patients, could be targeted to specifically block chronic bacterial infections of the lower respiratory tract.

In COPD patients, the upregulation of PAFR expression, in addition to providing a gateway to respiratory infections, promotes disease progression independent of bacterial colonisation.

PAFR expressed on neutrophils, the most important effector cells in COPD development, upon activation induces the autophagic death of neutrophils (Lv et al., 2017). This releases neutrophil elastase, which results in extensive lung tissue damage and hypersecretion of mucus into the airways, causing further airflow obstruction (Nadel, 2000). This provides a basis for the investigation of PAFR as a potential therapeutic target for treating COPD exacerbations.

Besides, PAFR has been extensively studied as a promising target for a number of inflammatory diseases, including cancer and asthma (Chung and Barnes, 1991, Tsoupras et al., 2009). PAF and a series of oxidised phospholipids, well-characterised natural agonists of PAFR have been implicated in theses inflammatory conditions (Subbanagounder et al., 1999). PAF and PAF-like phospholipids bind to PAFR and stimulate multiple signal transduction pathways including phospholipase C, D, A2, mitogen-activated protein kinases (MAPKs) and the phosphatidylinositol-calcium second messenger system, leading to pro-inflammatory cascades (Shukla, 1992, Iovino et al., 2013, Iovino et al., 2016).

PAFR is overexpressed in cancers, such as non-small cell lung cancer (NSCLC) as well as in breast, colorectal, gastric, ovarian and prostate carcinomas (Ji et al., 2016, Chen et al., 2015b, Chen et al., 2015a, Aponte et al., 2008). Overexpression of PAFR correlates with clinical stages, survival time, and distant metastasis. Inhibition of PAFR activity or knockdown of the expression of PAFR has been demonstrated to effectively block the invasion of breast and ovarian tumour cells (Yu et al., 2014a, Cellai et al., 2006). Moreover, PAFR antagonists have been implicated in improving the efficacy of chemo-and radiotherapy in breast and ovarian cancers (da Silva Junior et al., 2018, Yu et al., 2014b).

Additionally, PAF is considered to have a potential role in the pathogenesis of asthma. Asthmatic patients have increased levels of PAFR expression in their lungs, which contributes to complex inflammatory processes in the airways (Shirasaki et al., 1994). PAFR, in a mouse

model, has been demonstrated to be critical for the development of airway hyperresponsiveness, a cardinal feature of asthma (Ishii et al., 2004). PAFRs are also implicated as important mediators of house dust mite allergy development (Patel and Kearney, 2017). Moreover, PAFR antagonists appear to be effective in inhibiting systemic, cellular and pulmonary effects induced by PAF in patients with mild bronchial asthma (Hsieh, 1991, Hozawa et al., 1995, Gomez et al., 1998). However, in asthmatic patients challenged with allergen, protective effect of PAFR antagonist has been found to be modest against late asthmatic response and limited against the early response and airway hyperresponsiveness (Evans et al., 1997). This indicates the possibility for PAFR independent allergic inflammation in asthmatic patients (Ishii et al., 2004).

1.8 Platelet activating factor receptor antagonists

Ginkgolides are the natural PAF antagonists, of which ginkgolide B (BN-52021) is the most widely examined PAF antagonist (Chung et al., 1987). Ginkgolides exhibit off target biological effects due to serine proteinase inhibitory activity (Braquet, 1997). CV-3988, a thiazolium derivative was the first synthetic substance with selective activity against PAFR (Terashita et al., 1983). Although CV-3988 was tolerable, orally was not active, which limited its therapeutic utility (Arnout et al., 1988). Later, hetrazepine class of antagonists were developed, particularly WEB-2086, which is orally active, safe and well tolerated in human (Adamus et al., 1989). WEB-2086 is a derivative of brotizolam, which is a CNS active agent with sedative, hypnogenic and anxiolytic effects (Casals-Stenzel, 1991). Thus, both WEB-2086 and brotizolam have the core diazepine structure that exhibits dual biological functions: provides potency and activity for PAFR antagonist and mediates neurogenic effect *via* benzodiazepine receptor (Casals-Stenzel, 1987, Hirouchi et al., 1992). A clinical study in asthmatic patients

reported an ineffectiveness of WEB-2086 when administered at a lower dose of 40 mg whereas it showed a protective effect when used at the higher dose of 80 mg (Tamura et al., 1996, Spence et al., 1994). However, the off-target activity of WEB-2086 towards benzodiazepine receptor may limit the therapeutic use of WEB-2086 at a higher dose. Furthermore, the requisite fused diazepine ring in WEB-2086 is synthetically less accessible that has complicated the large-scale synthesis of this compound (Hyland et al., 2018). Thus, there is a need for developing a novel PAFR antagonist that is synthetically more accessible and has less off target effects, i.e., highly selective towards PAFR.

1.9 Nontypeable Haemophilus influenzae

NTHi is a major bacterial pathogen associated with COPD and its exacerbations (Murphy et al., 2004). Their primary ecological niche is the human upper respiratory tract, where they colonise as part of the normal flora without causing any disease. As discussed above, increased expression of PAFR on the lower airway epithelium in COPD patients and in individuals exposed to other risk factors, such as cigarette and biomass smoke provides a gateway for colonisation and establishment of infections in the lower airways. Besides host factors, NTHi utilises its various virulence factors, including endotoxins (lipopolysaccharide), adherence factors, iron uptake proteins and immune evasion factors that contribute to pathogenesis in diseases, including COPD (Duell et al., 2016). Unlike, encapsulated *H. influenzae*, which possesses carbohydrate capsule as a major virulence factor, for NTHi there is no single feature that is characteristic of all disease-associated strains (Moxon and Vaughn, 1981). It is possible that NTHi strains with similar combinations of virulence-related genes might be associated with different diseases and those with the potential to cause similar disease might have different combinations of virulence factors (Erwin and Smith, 2007). It is not yet clear whether NTHi

strains associated with different clinical phenotypes differ in virulence factors (Erwin and Smith, 2007). Moreover, the extent to which disease isolates of NTHi are phenotypically or genotypically similar has not been studied extensively.

1.9.1 Phenotypic diversity of NTHi isolates

NTHi infection leads to a broad range of disease symptoms that result from complex host-pathogen interactions (Hardy et al., 2003, Craig et al., 2001, Williams et al., 2001). Genomic heterogeneity of NTHi further complicates this interaction and diversifies the clinical phenotypes (Shen et al., 2005, Erwin et al., 2005). In different ecological niches, such as middle ear, sinuses, blood, and the respiratory tract, NTHi are exposed to different microenvironments defined by their unique nutrient availability, pH, oxidising potential and immune response (Huffnagle et al., 2017). NTHi adapt to these diverse niches by developing metabolic capacity that enable them to utilise the available resources efficiently and withstand the challenges. Classically, *H. influenzae* strains were phenotypically grouped into “biotypes” based primarily on the three biochemical tests: ornithine decarboxylation, indole production, and urea hydrolysis (Kilian, 1976). Unlike capsulated strains, NTHi isolates are found to be more diverse in biotype (Kilian, 1976). Further, NTHi strains associated with a particular niche and disease exhibit diversity in their phenotype (based on biotype). NTHi strains isolated from middle ear of patients with chronic otitis media belonged to five different biotypes I, II, III, V and VI (DeMaria et al., 1984). Biotypes II, III and V are found to be associated with both invasive strains and commensals of the upper respiratory tract diseases (Long et al., 1983, Wallace Jr et al., 1981). This suggest intra-niche and intra-disease specific diversity in NTHi based on their biotype (phenotype).

Besides metabolic ability, presence and absence of surface molecules and structure of lipooligosaccharides (LOS) add diversity to NTHi phenotype. Heterogeneity of outer membrane proteins and LOS structure may allow NTHi to manipulate interaction with the host and to evade host immune defences (Swords et al., 2003, Osman et al., 2018). The outer membrane protein, fimbriae are found to be more prevalent among NTHi strains isolated from throat than the middle ear strains, while high-molecular-weight adhesins are more prevalent among middle ear than throat isolates of NTHi (Ecevit et al., 2004). In addition, specific LOS modifications, such as terminal sialylation of the LOS that provide protection to NTHi against the host bactericidal IgM and complement-mediated serum killing add diversity to NTHi strains (Jackson et al., 2019). Furthermore, acetylation of LOS has been found to be associated with invasive NTHi strains as compared to non-invasive isolates, which imparts them a serum resistance phenotype (Phillips et al., 2019).

1.9.2 Genomic diversity of NTHi strains

NTHi diversity studies have shown uniqueness in genomic content of each strain (Erwin et al., 2005, Shen et al., 2005, FARJO et al., 2004). NTHi are naturally transformable and comparative genomic studies based on whole genome sequencing have demonstrated the existence of substantial genomic plasticity among NTHi strains (Shen et al., 2005, Hogg et al., 2007). This results mainly from horizontal gene transfer (HGT) processes, which cause nonuniform genetic exchanges among bacteria (Dutta and Pan, 2002). The high rate of HGT in NTHi has been demonstrated by the presence of multiple unique strains co-localised within the biofilms and by the simultaneous existence of polyclonal NTHi populations during chronic infections, such as COPD and cystic fibrosis (Murphy et al., 1999, Molin and Tolker-Nielsen, 2003). This diversity within a population is important to ensure survival of the population.

Environmental challenge may selectively enrich the resilient strains and decrease the susceptible strains, while maintaining survival of the population (Lozupone et al., 2012). Thus, evolutionary pressures have selected beneficial mechanisms that generate diversity across the NTHi population.

Shen et al. (2005) showed that approximately 10% of the genes present in clinical strains of NTHi are novel with respect to the reference genome of Rd KW20 strain (Shen et al., 2005). Moreover, they showed that no two clinical strains have the same set of unique genes, suggesting a unique distribution of genes among NTHi strains from the population supra genome (Shen et al., 2005). In 2007 Hogg et al. conducted comparative genomic analysis of 13 NTHi isolates, including 12 clinical strains and for the first time characterised the *H. influenzae* supragenome and studied the distribution of genetic diversity among the strains (Hogg et al., 2007). They reported an existence of an NTHi supragenome containing between 4,425 and 6,052 genes, which was much larger than the genome of any single bacterium (Hogg et al., 2007). Each typical pair of strains was found to differ by more than 300 genes, suggesting large heterogeneity among NTHi strains with respect to their clinical sources. This study also suggested the possibility of association between certain set of genes and the discrete disease phenotypes of the clinical NTHi strain (Hogg et al., 2007).

De Chiara et al. (2014) examined the association between phylogenetic analysis and the clinical source of NTHi strains using large set of genome sequences, comprising of 97 NTHi isolates (De Chiara et al., 2014). They reported the separation of NTHi strains into discrete clades based on their core genes sequence and also based on the presence and absence profile of certain virulence factors. However, this separation did not correlate with the clinical source (De Chiara et al., 2014). In other words, closely related strains forming a discrete clade were from different clinical phenotypes. This suggests that there is no association between set of core genes or

known virulence genes and the discrete disease phenotypes. However, they did not take into consideration the role of accessory genes in defining the NTHi patho-phenotypes.

1.9.3 Molecular tools for differentiation between NTHi strains

NTHi strains are highly diverse, and various typing methods have been proposed for epidemiological studies and pathogenesis investigations. Traditional biotyping and phenotyping methods such as outer membrane protein and LOS analysis (Barenkamp et al., 1982, Murphy et al., 1987) have gradually been replaced by genotypic techniques. Over a long period of time, genetic diversity of NTHi has been studied using ribotyping, restriction fragment length polymorphism (RFLP), multilocus enzyme electrophoresis, and multilocus sequence typing (MLST) (Bruce and Jordens, 1991, van Alphen et al., 1997, Meats et al., 2003). However, these genotyping techniques provide only a moderate typing resolution. Subtyping of genome pattern by pulsed-field gel electrophoresis (PAGE) typing provides a much stronger discriminatory power than all of the above mentioned typing methods (Saito et al., 1999). It involves the comparison of patterns of genomic DNA digested with a rare cutting restriction enzyme, such as *SmaI* (Lee et al., 2006). Whole genome sequencing (WGS) allows the examination of the entire region of a genome and hence provides the highest level of resolution to discriminate between the strains, particularly suitable to unravel differences between closely related strains that are associated with different disease phenotypes (Salipante et al., 2015). With the recent development in sequencing technologies, next-generation sequencing (NGS) is now available that enables high-throughput sequencing of entire bacterial genomes at an affordable cost and with short turnaround time (Goodwin et al., 2016). This can be coupled with appropriate bioinformatics tools for robust handling and analysis of big

sequence data. Thus, whole genome sequence analysis can be used to detect small genetic differences with a potential to distinguish between very closely related isolates.

1.9.4 Distinction between clinical strains of NTHi

Some studies support the idea that differences between strains in virulence-related phenotypes correlate with their clinical sources. Bresser and colleagues investigated how persisting and nonpersisting strains from chronic bronchitis differ in the extent to which they activate epithelial cells to induce inflammation (Bresser et al., 1997). Persisting strains were found to induce a weak inflammatory response, producing less IL-6 and IL-8 in H292 lung epithelial cells than nonpersisting strains (Bresser et al., 1997). This difference in their ability to affect airway function was not explained by differences in adherence to H292 cells (Bresser et al., 1997). Later in 2005, Chin et al reported higher adherence of NTHi to primary human tracheobronchial epithelial cells for strains isolated from COPD patients having exacerbations than those with stable conditions (Chin et al., 2005). This correlated with their ability to induce inflammatory response: NTHi strains from COPD exacerbations produced more IL-8 in airway epithelial cells. Moreover, they compared these strains using an *in vivo* mouse model of airway infection and demonstrated that the strains associated with exacerbations caused more airway neutrophil recruitment than the colonisation strains (Chin et al., 2005). Since, COPD exacerbations are reported to be associated with acquisition of new strains of bacteria (Sethi et al., 2002) this study provides evidence of a strong link between infection with a new NTHi strain and induction of high inflammatory response. Although both studies suggested differences in bacterial factors to be key driver of pathogenesis, neither of these studies investigated the genetic basis of differences in these virulence-related phenotypes.

Recently, Pettigrew and colleagues conducted a 15-year prospective study, in which they analysed whole genomes of 101 longitudinal pairs of NTHi from COPD patients and assessed genomic changes that occurred during persistence in the human airways (Pettigrew et al., 2018). They reported changes in simple sequence repeats in multiple genes that regulate vital virulence functions, such as adherence, nutrient uptake, and modification of LOS. Moreover, analysis of polymorphic sites in 12 different surface antigens showed amino acid sequence changes in eight antigens during persistence of NTHi in COPD airways (Pettigrew et al., 2018). These results indicate that persisting NTHi isolates undergo genomic changes in simple sequence repeats of virulence associated genes and polymorphism in vaccine candidate genes, which are the major mechanisms for adaptation of NTHi in the respiratory tract of COPD patients.

More recently, Moleres et al (2018) examined 92 NTHi isolates from 13 COPD patients over 1 to 9 years to study genetic variation association with bacterial adaptation to the COPD lung (Moleres et al., 2018). They identified recurrent loss-of-function mutations in the *ompP1* (*fadL)* gene, in addition to genetic changes in phase-variable genes as reported previously by Pettigrew et al. This loss of OmpP1 function reduces NTHi's ability to adhere to and infect airway epithelial cells but enhances its resistance to bactericidal fatty acids, which are inflammatory mediators enriched in the COPD lung (Moleres et al., 2018). Extending this analysis to large number of publicly available NTHi genomes, they found truncated *ompP1* alleles in majority of the isolates from the lower airway tracts as compared to those collected from other body parts. This suggests COPD airways are colonised predominantly by specific NTHi strains harbouring loss-of-function mutations in the *ompP1* gene.

1.10 Concluding remarks for the literature review and questions for the present thesis

COPD is a major public health issue with a global increase in prevalence, particularly among non-smokers. Biomass smoke exposure has been recognised as an important risk factor contributing to the development and progression of COPD. The burden of biomass smoke exposure is alarming as nearly half of the world's population are at risk of exposure to noxious particles released by combustion of biomass fuel. However, the underlying mechanism how biomass smoke impacts the human respiratory tracts and contributes to COPD progression is still unclear. This is in part due to the lack of a standard and easily accessible procedure for the preparation of biomass smoke. Thus, there is a need for standardized experimental approaches for the preparation of re-usable biomass smoke extract that could be used for the assessment of cellular responses to different types of biomass smoke.

Infection with respiratory bacteria and viruses is another important factor implicated in majority of exacerbations during COPD. NTHi is a key pathogen associated with comorbidity in COPD patients. It utilises PAFR, expressed on the cell surface of human airway epithelial cells for binding to and colonising in the lower airways. Application of PAFR antagonist could therefore be a potential therapeutic strategy for prevention and treatment of COPD exacerbations. Available PAFR antagonists are either ineffective at low dose, orally inactive, or are synthetically inaccessible, so newer drugs have to be developed that is nontoxic and amenable to large-scale production. In addition to PAFR mediated adherence, NTHi utilises a wide array of factors to survive and colonise in the COPD airways. It is still not clear what makes NTHi competent to establish infection in lower airways of COPD patients and how are these COPD strains different from NTHi isolates collected from other clinical diseases.

1.11 Hypotheses and Aims

1. **Hypothesis:** Generation of quantifiable and preservable smoke extracts could be used in multiple exposure experiments for comparative assessment of cellular responses to different types of smoke.

   Aim: Develop a simple smoke generation system to prepare quantifiable batches of cigarette, wood and cow dung smoke extracts and test them *in vitro* to assess their effects in inflammatory response in human bronchial epithelial cells.

2. **Hypothesis:** Compared to WEB-2086 its analogue without the core diazepine structure is synthetically more accessible and exhibits potency towards inhibiting adherence of NTHi to bronchial epithelial cells.

   Aim: Test the activity of chemical compounds (analogues of WEB-2086) towards adherence of nontypeable *Haemophilus influenzae* (NTHi) to cigarette smoke exposed bronchial epithelial cells.

3. **Hypothesis:** Exposure of bronchial epithelial cells to biomass smoke extracts increases adherence of NTHi.

   Aim: Compare, *in vitro*, the effects of cigarette, wood and cow dung smoke extracts exposure in attachment of NTHi to airway epithelial cells.

4. **Hypothesis:** NTHi associated with COPD may exhibit genetically encoded functional variances compared to NTHi isolates from non-COPD clinical illnesses.

   Aim: Investigate differences in gene content between NTHi isolates from patients with COPD and other clinical illnesses.

2 Chapter 2: A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells

This chapter that addresses the first aim of the study; To develop a simple smoke generation system to prepare quantifiable batches of cigarette, wood and cow dung smoke extracts and test them *in vitro* to assess their effects in inflammatory response in human bronchial epithelial cells.

This Chapter is presented as a Word document unchanged from the final accepted version published in a peer-reviewed journal as

KC R, Zosky GR, Shukla SD, O'Toole RF. A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells. *Biology Methods and Protocols*. 2018;3(1): bpy010 (see appendix).

2.1 Abstract

Nearly half of the world's population uses biomass fuel for the purposes of cooking and heating. Smoke derived from biomass increases the risk of the development of lung diseases, including pneumonia, chronic obstructive pulmonary disease, airway tract infections, and lung cancer. Despite the evidence linking biomass smoke exposure to pulmonary disease, only a small number of experimental studies have been conducted on the impact of biomass smoke on airway epithelial cells. This is in part due to the lack of a standard and easily-accessible procedure for the preparation of biomass smoke. Here, we describe a cost-effective and reproducible method for the generation of different smoke extracts, in particular, cow dung and wood smoke extracts for use in a range of biological applications. We examined the effect of the biomass smoke extracts on the expression of a bronchial epithelial cell surface receptor, the platelet activating factor receptor (PAFR), a known responder to cigarette smoke exposure. Similar to cigarette smoke extract treatment, we observed a dose-dependent increase in PAFR expression on human airway epithelial cells that were subjected to cow dung and wood smoke extracts. This method provides biomass smoke in a re-usable form for cell and molecular bioscience studies on the pathogenesis of chronic lung disease.

2.2 Introduction

It is estimated that nearly 3 billion people worldwide are exposed to biomass smoke, generated from burning wood, crop residues, or animal dung for household cooking and heating (WHO, 2016). Biomass smoke is the leading environmental cause of death and disability, causing over 4 million deaths each year (Lim et al., 2012). Several epidemiological studies have associated biomass smoke exposure with lung diseases, including chronic obstructive pulmonary disease

(COPD), airway infections, and lung cancer (Assad et al., 2015, Assad et al., 2016, Gordon et al., 2014, Bruce et al., 2015, Capistrano et al., 2017). Similarly, *in vitro* studies have found that human lung cells exhibit impaired inflammatory and immune responses following exposure to biomass smoke (Mehra et al., 2012b, McCarthy et al., 2016b). Inhalation of animal dung biomass smoke is of particular concern to human health as it has the highest polluting potential per unit energy released compared to wood smoke (Mudway et al., 2005). Airway epithelial cells are the primary target of inhaled smoke; therefore, the responses of epithelial cells to different types of biomass smoke are of considerable interest. Although, extensive *in vitro* studies have been performed on the effects of tobacco smoke on the expression of host receptors on respiratory epithelial cells and on susceptibility to bacterial infection (Grigg et al., 2012, Shukla et al., 2016a, O'Toole et al., 2016), only a small number of comparable studies have been performed using biomass smoke. More mechanistic research is therefore needed to understand the cellular and molecular responses to biomass smoke, including animal dung and wood smoke. Currently, we do not have standardized experimental approaches for the preparation of re-usable biomass smoke extract and for the assessment of cellular responses to different types of biomass smoke. Here, we devised a low-cost and reproducible biomass smoke generation system and tested the extracts produced on the expression profile of an established marker of cigarette-smoke exposure, the platelet activating factor receptor (PAFR), on human bronchial epithelial cells (O'Toole et al., 2016, Shukla et al., 2015, KC et al., 2017).

2.3 Materials and Methods

2.3.1 Preparation of cigarette smoke extract (CSE)

Cigarette smoke extract was prepared at the College of Health and Medicine, University of Tasmania, Australia as previously described (Granstrom et al., 2006). Briefly, the filter from a

Marlboro cigarette butt was replaced with a sterile cotton wool filter and was smoked using a water aspirator (Grigg et al., 2012). The water aspirator consisted of a tee with hose barbs on three sides fitted with hoses. The hose-fitted tee was clamped in a stand as shown in Figure 2.1A. One of the hoses from the tee fitting was connected to a tap, the second on the opposite side drained water to a sink, and the third hose at a right angle held the cigarette roll. When water was passed through the tube, a vacuum was generated by the Venturi effect, drawing smoke from a burning cigarette (Baylar and Ozkan, 2006). Here, the flow of water was maintained at the constant rate of 110 mL/s. The flow of water ensured the continual smoking of the cigarette, thereby collecting the cigarette smoke material in the cotton-wool filter (Grigg et al., 2012, Granstrom et al., 2006). After complete combustion of a cigarette, the cotton-wool filter was removed and then placed into another cigarette from which the filter had been removed. This way, the same cotton filter was used in the smoking of three Marlboro cigarettes.

The cigarette smoke material retained in the cotton filter was quantified by measuring the weight of cotton filter before and after the combustion of the three cigarettes. The cotton filter was then vortexed in 1 mL dimethyl sulfoxide (DMSO). The solubilized smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The cigarette smoke extract (CSE) was then filter-sterilized through a 0.22 μm nylon membrane filter (ThermoFisher Scientific, USA), which is resistant to DMSO. The filtrate was then re-quantified by weight measurement.

2.3.2 Preparation of cow dung smoke extract (CDSE)

Cow dung, that was collected from a local farm near Hobart, Tasmania, was sun-dried for approximately 5 days and was crushed into fine particles using a mortar and pestle. Cow dung

powder were rolled in a paper with a sterile cotton wool filter at one of the ends, similar to a filtered-cigarette (Figure 2.1B). Four such rolls were prepared from a total of 7.5 g of cow dung fine particles, such that each roll contained 1.875 g of cow dung powder. The dung roll was then burned and smoked using the water aspirator, as described for cigarette smoking. After complete combustion of a dung roll, the cotton filter was removed and placed in another cow dung roll. In this way, four such cow dung rolls were smoked using the same cotton-wool filter. Finally, the cow dung smoke extract was prepared by vortexing the cotton wool filter in 1 mL DMSO. The solubilized smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The cow dung smoke extract (CSE) was then filter-sterilized through a 0.22 μm nylon membrane filter and the filtrate was re-quantified by weight measurement.

2.3.3 Preparation of wood smoke extract (WSE)

Firewood was collected from a local supplier near Hobart, Tasmania. For wood smoke generation, 1.875 g of wood shavings were rolled in paper with a sterile cotton wool filter at one of the ends, similar to a filtered-cigarette (Figure 2.1C). Four such rolls were prepared from a total of 7.5 g of wood shavings. Each wood shaving roll was then burned and smoked using the water aspirator, as described for cigarette smoking. After complete combustion of a wood shaving roll, the cotton filter was removed and placed into another wood shaving roll. Finally, the wood smoke extract (WSE) was prepared by vortexing the cotton-wool filter in 1 mL DMSO. The solubilized smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The wood smoke extract (WSE) was then filter-sterilized through a 0.22 μm nylon membrane filter and the filtrate was re-quantified by weight measurement.

2.3.4 Normalization of smoke extracts

To compare the effects of different smoke extracts, the prepared extracts were normalized to the same concentration and were stored at -20^{o}C in aliquots of 100 µL until use. The normalized smoke extracts were diluted in bronchial epithelial cell growth medium (BEGM) for use in the smoke extract exposure experiments.

2.3.5 *In vitro* BEAS-2B cell culture

As airway epithelial cells are the primary cells to respond to smoke, an immortalized cell line of human bronchial epithelial cells, BEAS-2B (Catalogue no 95102433, Sigma-Aldrich), was selected for this study. The BEAS-2B cells were maintained at 37°C, 5% CO_2 in bronchial epithelial cell growth medium (BEGM) (Lonza, Basel, Switzerland) supplemented with the BulletKit (Lonza). The BEAS-2B cells were sub-cultured in T75 flasks (Corning Inc., Corning, NY, USA), and were used for the experiments at passage numbers ≤ 15 passages. Sterile 8-well chambered glass slides (Millipore, Billerica, MA, USA) were pre-coated by incubating overnight at 4^{o}C with 200 µL of 5% (v/v) bovine collagen I (ThermoFisher Scientific, USA), prepared in 20 mM acetic acid. The wells were rinsed twice with pre-warmed phosphate buffered saline (PBS) followed by seeding of the BEAS-2B cells at a cell density of 30,000 cells per well in 200 µL BEGM and incubated overnight at 37°C, 5% CO_2. On the following day, the culture media was replaced with fresh BEGM and incubated at 37°C, 5% CO_2 for 24 h before the *in vitro* smoke extract exposure experiments.

2.3.6 Exposure of BEAS-2B cells to smoke extracts

Approximately 50,000 to 60,000 BEAS-2B cells in each well were exposed to 200 µL of BEGM containing five different concentrations of CDSE and WSE, ranging from 8.75 ng/mL to 8.75 µg/mL, for 4 h at 37°C and 5% CO_2. Parallel exposures of BEAS-2B cells to CSE in the concentration range of 8.75 ng/mL to 8.75 µg/mL were also performed for comparison.

2.3.7 Immunofluorescence

After 4 h of exposure to the smoke extracts, the media was discarded, and the cells were washed twice with 200 µL PBS pre-warmed at 37°C. The cells were then fixed with 200 µL of 4% (w/v) paraformaldehyde (Sigma-Aldrich) for 20 minutes at room temperature. The cells were rinsed twice with 200 µL of PBS and permeabilized with 100 µL of chilled (-20°C) acetone for 10 minutes at room temperature. After washing the cells again twice with 200 µL of PBS, the non-specific binding sites were blocked with 200 µL of 1% (w/v) bovine serum albumin (Sigma-Adrich), prepared in PBS containing 0.1% (v/v) Tween-20 (Sigma-Aldrich), for 1 h at room temperature. Cells were then incubated overnight with 100 µL of 2.5 µg/mL monoclonal antibody (mAb) against the human platelet activating factor receptor (PAFR) protein (11A4, Clone 21, Cayman Chemical Company, USA) at 4°C in the dark. The cells were then rinsed three times with 200 µL of 0.1% (w/v) bovine serum albumin (BSA) in PBS and incubated for 1 h with 100 µL of 1:100 dilutions of Alexa Fluor 594 conjugated goat anti-mouse IgG (H+L) secondary antibody (ThermoFisher Scientific, USA) at room temperature. After rinsing three times with 200 µL of 0.1% (w/v) BSA in PBS, the cells were stained with 200 µL of 1 µg/mL 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI) (ThermoFisher Scientific, USA) for 15

min at room temperature. Finally, the cells were washed three times with 200 μL of PBS, air-dried and the slides were mounted with Dako fluorescence mounting media (Agilent, USA).

2.3.8 Microscopy and image analysis

Cell preparations were examined under 400× magnification using an Olympus BX50 epifluorescence microscope with NIS elements software (Nikon; Tokyo, Japan) and Cool Snap Hq2 CCD camera (Photometrics, Tucson, AZ, USA). Five images were taken per well from different points using multi-fluorescence channels designed for simultaneous detection of emission from the fluorochromes DAPI (violet excitation and blue emission, 200 ms exposure), and Alexa Fluor 594 (green excitation and red emission, 300 ms exposure). The level of cellular PAFR protein expression was quantified as a measure of total cell fluorescence intensity using the software ImageJ (NIH, USA) (Schneider et al., 2012). The cellular fluorescence was corrected against the background fluorescence using the following formula:

Total cell fluorescence (TCF) = integrated density – (area of selected cell × mean fluorescence of the background).

2.3.9 Quantitative real-time polymerase chain reaction (qPCR) analysis

The expression of PAFR was also determined at the transcriptional level using qPCR. The BEAS-2B cells were seeded into sterile clear-flat bottom 12 well plates (Corning Inc) at a density of 2×10^5 cells per well and incubated overnight at 37^{o}C and 5% CO_2. The next day, cells were exposed to different concentrations of CSE, CDSE and WSE at 37^{o}C and 5% CO_2. After 3 h, total RNA was extracted with Tri-reagent (Sigma-Aldrich), according to the

manufacturer's instructions. It was then treated with DNase (Promega). 490 ng of RNA was converted into first-stranded cDNA using a SensiFAST cDNA synthesis kit (Bioline). The cDNA generated was amplified on a LightCycler 480 System (Roche) with the SensiFAST Probe No-ROX kit (Bioline) in a total volume of 20 µL. The relative fold change of mRNA expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

2.3.10 BEAS-2B cell viability assay

The BEAS-2B cells were seeded into a sterile clear-flat bottom 96 well plate (Sigma-Aldrich) at a density of 5,000 cells per well and incubated overnight at 37^{o}C and 5% CO_2. The cells were exposed to different concentrations of CSE, CDSE and WSE. Alamar Blue (Life Technologies) was then added to each well at a final concentration of 10% (v/v). The absorbance readings were taken at 570 nm and 600 nm at 2 and 4 h post-exposure to smoke extracts at 8.75 and 87.5 µg/mL concentrations using a Spectromax Spectrophotometer Microplate Reader (Molecular Devices, USA). The percent reduction of Alamar Blue was calculated using the following formula:

$$\% \text{ Reduction of Alamar Blue Reagent} = \frac{(Eoxi600 \times A570) - (Eoxi570 \times A600)}{(Ered570 \times C600) - (Ered600 \times C570)} \times 100$$

Molar extinction coefficient of oxidized Alamar Blue at 570 nm (Eoxi570) = 80586; at 600 nm (Eoxi600) = 117216

Absorbance of test wells at 570 nm (A570); at 600 nm (A600)

Molar extinction coefficient of reduced Alamar Blue at 570 nm (Ered570) = 155677; at 600 nm (Ered600) = 14652

Absorbance of negative control well (no cells) at 570nm (C570); at 600 nm (C600)

2.3.11 Statistical analysis

Data were expressed as mean ± standard error, median and interquartile range using the Microsoft Excel Statistics package (Microsoft Corporation, Redmond, WA, USA) and analyzed using GraphPad Prism version 5.0 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com. Comparisons between groups were performed using unpaired two-tailed t-tests with Welch's correction and one-way analysis of variance (ANOVA) with Dunnett's multiple comparison analysis.

2.3.12 Ethical approval

This work was conducted in accordance with Ethics Approval #H0016214 from the Tasmanian Health and Medical Human Research Ethics Committee.

2.4 Results

2.4.1 Cigarette, cow dung and wood smoke extracts preparation

66.52 mg, 124.7 mg and 131.2 mg of smoke particles were retained in cotton filter after the combustion of 3 cigarettes and 7.5 grams of cow dung powder and wood shavings. The retained smoke particles were then solubilized in DMSO. The concentration of DMSO dissolved

cigarette, cow dung, and wood smoke particles were 10.85 mg/mL, 43.7 mg/mL, and 36.0 mg/mL, respectively. After filter sterilization, the final concentration of cigarette, cow dung, and wood smoke material were 8.75 mg/mL, 24.4 mg/mL and 31.64 mg/mL, respectively. The CSE, CDSE and WSE concentrations were all normalized to same concentration of 8.75 mg/mL and were used in subsequent exposure experiments over the range from 8.75 ng/mL to 87.5 µg/mL.

2.4.2 CSE exposure increases PAFR expression on bronchial epithelial cells

PAFR expression was measured based on fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. Previous studies have reported that PAFR expression is upregulated in BEAS-2B cells exposed to CSE and that maximal induction occurred at 4 h of CSE exposure (Shukla et al., 2016a, Grigg et al., 2012). Here, cigarette smoke extract exposure for 4 h significantly increased the expression of PAFR on the bronchial epithelial cells (Figure 2.2A-C). We observed a dose-dependent increase in PAFR expression upon stimulation with CSE at a concentration range of 8.75 ng/mL to 87.5 µg/mL. In comparison to the mock (1% DMSO) treated control BEAS-2B cells, the mean PAFR expression was approximately 1.18, 1.39, 1.56, 2.06 and 6.10 times higher in the 8.75 ng/mL, 87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL CSE treated cells, respectively (Figure 2.2C). In addition, to demonstrate that the smoke content does not decay significantly over time, we compared the level of activity of the CSE over a 3-month period. There has been no significant loss in PAFR induction following -20^{o}C freezer storage of the CSE (Figure 2.5). The viability of the BEAS-2B cells exposed to 8.75 and 87.5 µg/mL of CSE for 2 and 4 h was assessed relative to mock (1% DMSO) treated control cells using the Alamar Blue assay. The mean relative viability of BEAS-2B cells treated with 8.75 µg/mL CSE for 2 and 4 h was at

106.2 ± 3.1 % (SEM (standard error of the mean)) and 95.8 ± 11.4 %, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/mL CSE for 2 and 4 h, the mean relative viability was at 100.7 ± 5.0 % and 95.8 ± 11.5 %, respectively.

2.4.3 CDSE treatment induces PAFR expression on bronchial epithelial cells

The expression of PAFR on BEAS-2B cells was significantly upregulated by exposure to cow dung smoke extract (Figure 2.3A-C). We observed a dose dependent increase in PAFR expression upon exposure to CDSE at a concentration range of 8.75 ng/mL to 87.5 µg/mL. The mean cellular PAFR expression was approximately 1.17, 1.24, 1.51, 1.84 and 4.67 times higher than the mock (1% DMSO) treated control BEAS-2B cells in the 8.75 ng/mL,87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL CDSE stimulated cells, respectively (Figure 2.3C). Induction of PAFR expression upon CDSE exposure was comparable over the concentration range of 8.75 ng/mL to 87.5 µg/mL in experiments conducted 3 months apart (Figure 2.5). From the Alamar Blue assay, the mean relative viability of BEAS-2B cells exposed to 8.75 µg/mL CDSE for 2 and 4 h was at 97.4 ± 6.1 % and 97.2 ± 13.9 %, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/mL CDSE for 2 and 4 h, the mean relative viability was at 102.2 ± 7.9 % and 90.8 ± 10.7 %, respectively.

2.4.4 WSE exposure upregulates PAFR expression on bronchial epithelial cells

Wood smoke extract exposure was also associated with an induction of PAFR expression on the bronchial epithelial cells (Figure 2.4A-C). Treatment with WSE for 4 h resulted in a concentration dependent increase in the expression of PAFR on BEAS-2B cells. Compared to the 1% DMSO treated control cells, the mean cellular PAFR expression was approximately

1.28, 1.27, 1.50, 1.99 and 4.34 times higher in 8.75 ng/mL, 87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL WSE exposed BEAS-2B cells, respectively (Figure 2.4C). The PAFR inducing activity of WSE was similar over the concentration range of 8.75 ng/mL to 87.5 µg/mL in experiments conducted 3 months apart (Figure 2.5). From the Alamar Blue assay, the mean relative viability of BEAS-2B cells exposed to 8.75 µg/mL WSE for 2 and 4 h was at 101.5 ± 9.2 % and 83.8 ± 8.2 %, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/mL WSE for 2 and 4 h, the mean relative viability was at 93.8 ± 7.4 % and 79.2 ± 8.1 %, respectively.

2.4.5 Transcriptional response of BEAS-2B to CSE, CDSE, and WSE

To examine the expression of PAFR at the transcriptional level, the relative PAFR mRNA expression, normalized to glyceraldehyde-3-phosphate (GAPDH), was measured post-exposure to the smoke extracts at 8.75 µg/mL and 87.5 µg/mL concentrations. Compared to mock (1% DMSO) treated controls, the transcriptional level of PAFR was increased 2.45, 3.37 and 2.65-fold after exposure to CSE, CDSE, and WSE, respectively, at 8.75 µg/mL (Table 2.1). The mRNA levels of PAFR was 3.19, 4.17, and 3.38-fold higher in BEAS-2B cells exposed to 87.5 µg/mL concentrations of CSE, CDSE, and WSE, respectively.

Previous studies have reported a respiratory inflammatory response to cigarette, animal dung, and wood smoke exposure (Mehra et al., 2012a, McCarthy et al., 2016b, Kocbach et al., 2008, Swanson et al., 2009, Shao et al., 2018). To investigate the inflammatory response *in vitro*, the BEAS-2B cells were exposed to 8.75 and 87.5 µg/mL concentrations of CSE, CDSE, and WSE for 3 h and mRNA expression was measured. The mRNA levels for pro-inflammatory cytokines, interleukin-1 beta (IL-1β), IL-6, and IL-8 were increased by 5.56-, 7.15-, and 10.86-

fold, respectively, post-exposure to 8.75 µg/mL CSE (Table 2.1). Exposure of BEAS-2B cells to 87.5 µg/mL CSE resulted in an 8.1-, 13.8-, and 11.4-fold increase in mRNA levels of the inflammatory mediators, IL-1β, IL-6, and IL-8. A similar increase in inflammatory mediators IL-1β, IL-6, and IL-8 was observed upon exposure of BEAS-2B cells to 8.75 and 87.5 µg/mL concentrations of CDSE and WSE (Table 2.1).

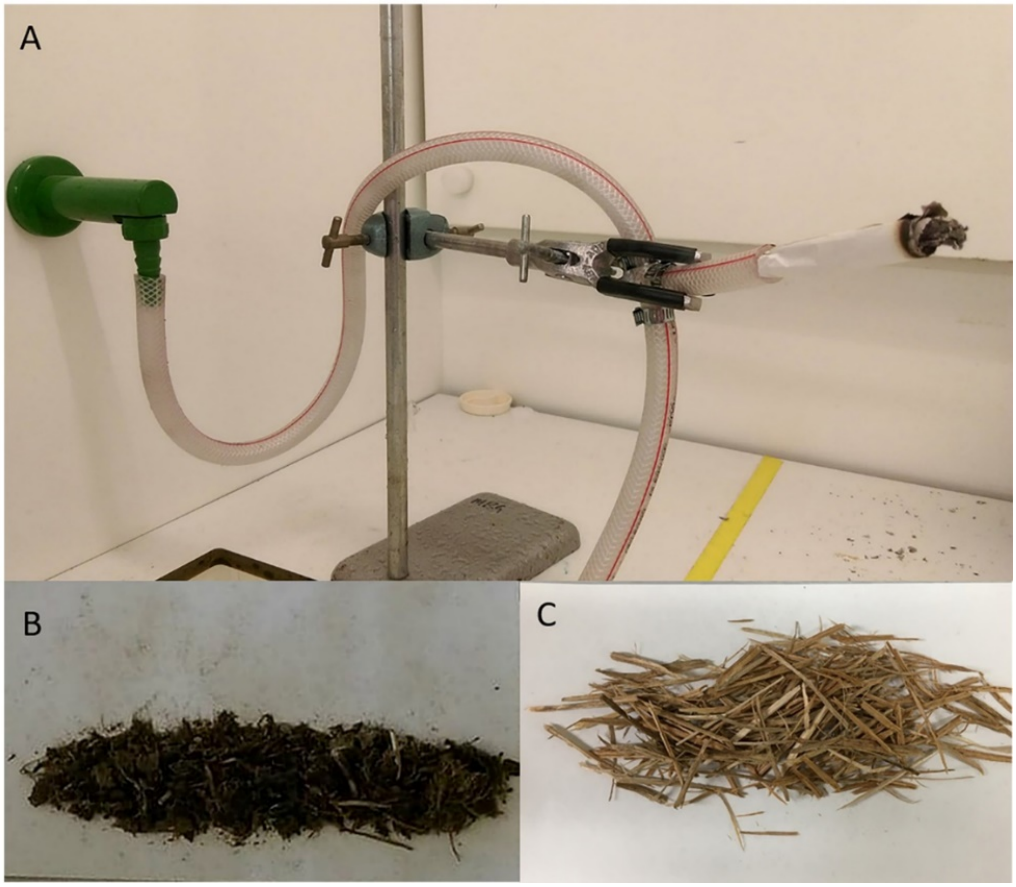


Figure 2.1 Generation of cow dung and wood smoke extract.

(A) A water aspirator was set up to draw smoke from a burning cow dung/wood roll using the vacuum created by the flow of water. (B) Cow dung was sun-dried, crushed into fine particles and rolled in paper. (C) Wood was cut into small chips and rolled in paper.

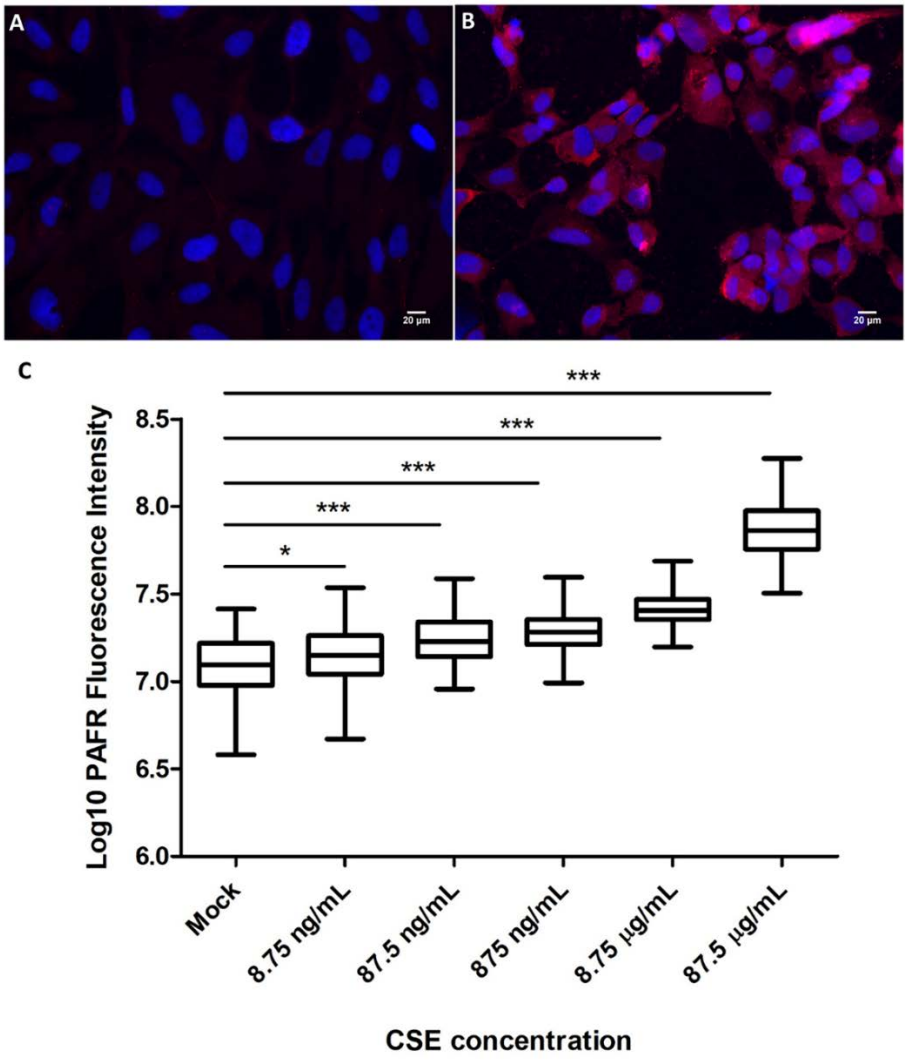


Figure 2.2 Cigarette smoke extract (CSE) exposure and PAFR expression on bronchial epithelial cells.

(A) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (B) BEAS-2B cells exposed to 87.5 µg/mL CSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 µg/mL, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 µg/mL, blue). Magnification = 400×. (C) Response to different concentrations of CSE. PAFR expression corresponds to log10 of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. PAFR expression was significantly increased in 8.75 ng/mL, 87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL CSE exposed BEAS-2B cells. Data are representative of two independent experiments (*$p < 0.05$, ***$p < 0.0001$, One-way ANOVA with Dunnett's multiple comparison test).

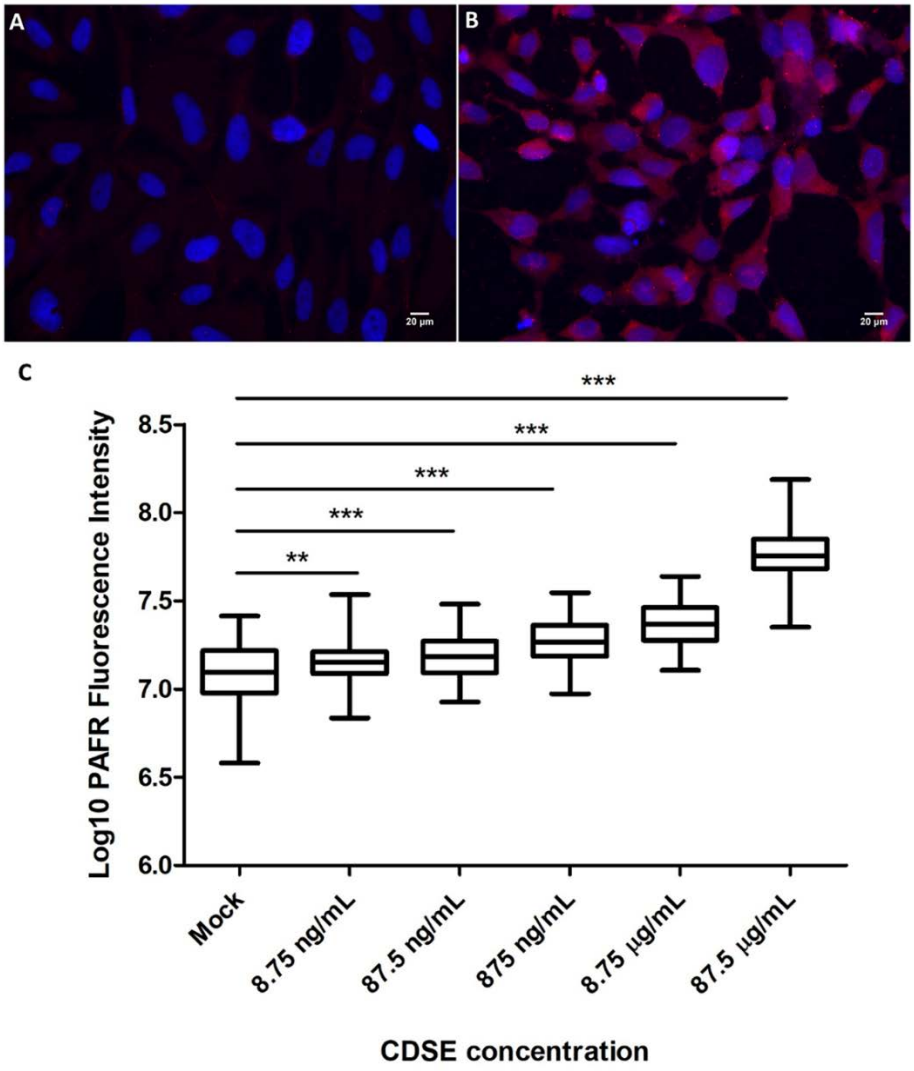


Figure 2.3 Cow dung smoke extract (CDSE) exposure and PAFR expression on bronchial epithelial cells.

(A) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (B) BEAS-2B cells exposed to 87.5 µg/mL CDSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 µg/mL, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 µg/mL, blue). Magnification = 400×. (C) Response to different concentrations of CSE. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. The PAFR expression was significantly increased in 8.75 ng/mL, 87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL CDSE treated BEAS-2B cells. Data are representative of two independent experiments (**$p < 0.001$, ***$p < 0.0001$, One-way ANOVA with Dunnett's multiple comparison test).

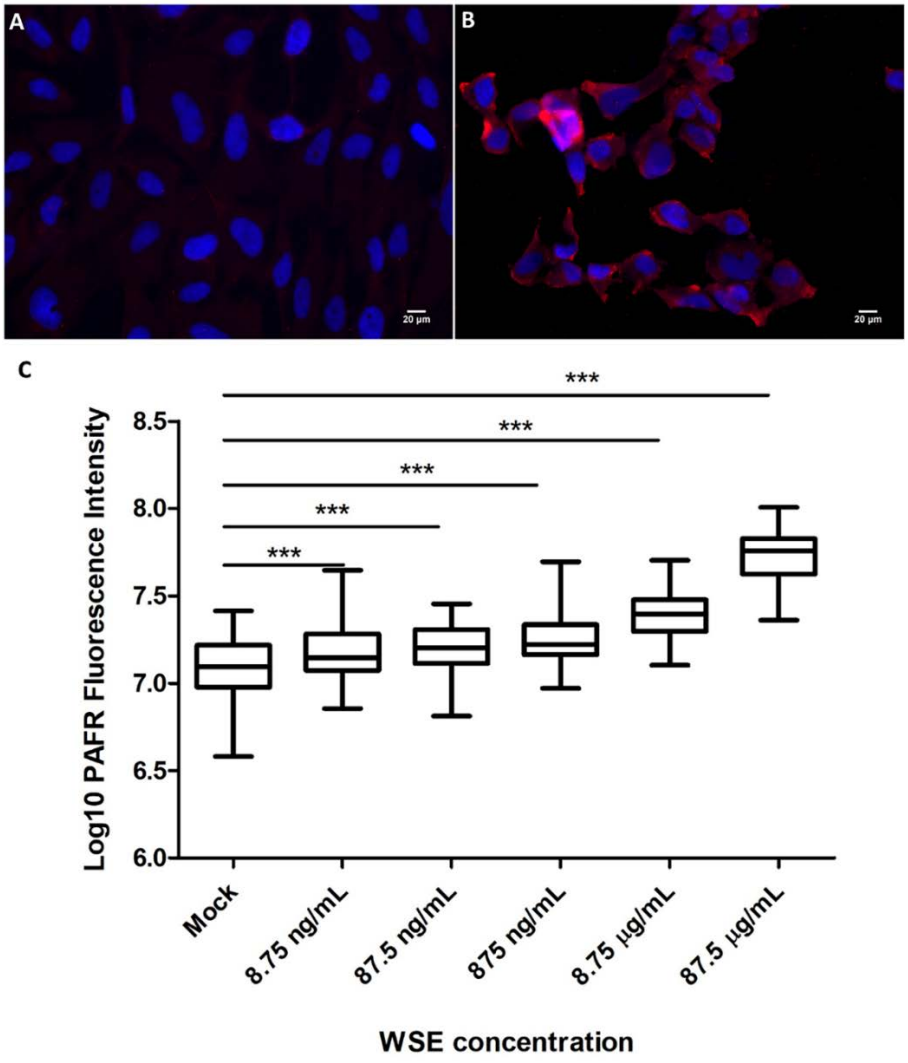


Figure 2.4 Wood smoke extract (WSE) exposure and PAFR expression on bronchial epithelial cells.

(A) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (B) BEAS-2B cells exposed to 87.5 µg/mL WSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 µg/mL, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 µg/mL, blue). Magnification = 400×. (C) Response to different concentrations of CSE. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. The PAFR expression was significantly increased in 8.75 ng/mL, 87.5 ng/mL, 875 ng/mL, 8.75 µg/mL and 87.5 µg/mL WSE exposed BEAS-2B cells. Data are representative of two independent experiments (***$p < 0.0001$, One-way ANOVA with Dunnett's multiple comparison test).

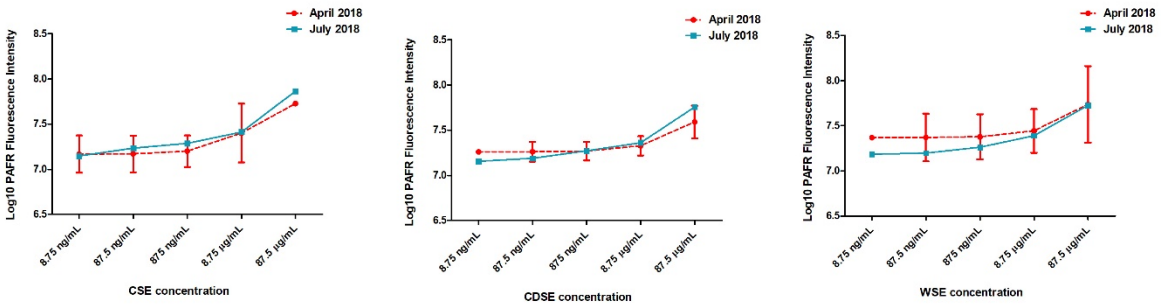


Figure 2.5 Comparison of effects of smoke extracts on PAFR expression in experiments conducted 3 months apart.

All of the smoke extracts, CSE, CDSE and WSE, were run in the concentration range from 8.75 ng/mL to 87.5 µg/mL in the July 2018 experiment. The CSE extract was run in the same concentration range from 8.75 ng/mL to 87.5 µg/mL in the April 2018 experiment. For the April 2018 experiments, the CDSE and WSE data were interpolated to the 8.75 ng/mL to 87.5 µg/mL concentration range using non-linear least squares regression. The levels of PAFR expression in the July 2018 experiments with the CSE, CDSE and WSE samples were within 95% CI of the levels obtained for the April 2018 experiments, and therefore, there was no significant decay in the PAFR inducing activity of the smoke extracts. Data are representative of two independent experiments. The error bars represent 95% CI of the predicted $\log_{10}$ PAFR fluorescent intensity for the April 2018 data.

Table 2.1 The smoke extracts CSE, CDSE and WSE induce PAFR, IL-1β, IL-6, and IL-8 mRNA expression by the BEAS-2B cells.

Marker	**CSE**		**CDSE**		**WSE**	
	A	B	A	B	A	B
PAFR	2.45 ± 0.08***	3.19 ± 0.69	3.37 ± 0.18***	4.17 ± 1.13	2.65 ± 0.25**	3.38 ± 0.5*
IL-1β	5.56 ± 0.65**	8.06 ± 2.95	8.26 ± 0.27***	14.32 ± 4.73	9.6 ± 1.24**	6.69 ± 0.45**
IL-6	7.15 ± 0.58**	13.83 ± 4.95	12.39 ± 0.4***	12.02 ± 1.57**	10.93 ± 1.05**	13.42 ± 5.74
IL-8	10.86 ± 1.61**	11.44 ± 4.42	11.64 ± 1.0**	20.25 ± 2.43**	10.62 ± 0.6***	8.67 ± 1.81*

Relative fold change in mRNA of PAFR and inflammatory cytokines normalized with GAPDH among BEAS-2B cells exposed to CSE, CDSE and WSE for 3 h at concentrations (A) 8.75 µg/mL and (B) 87.5 µg/mL (*p < 0.05, **p <0.001, ***p < 0.0001, unpaired two-tailed ‘t’ test with Welch’s correction. The data is presented as the mean of the observed fold change ± SEM (standard error of the mean), $n = 4$ per group.

2.5 Discussion

Nearly 4.3 million people die every year from illnesses attributable to the inhalation of biomass smoke (WHO, 2016). Among these deaths, 22% are due to COPD and 12% due to pneumonia (WHO, 2016). Biomass fuels (wood, animal dung and crop residues) are the major source of domestic energy for cooking and household heating, especially in developing countries. Emissions from biomass contain a multitude of pollutants that adversely affect human health, such as suspended particulate matter, methane, free radicals, aldehydes, toxic gases like carbon monoxide and nitrogen oxides, and polycyclic aromatic hydrocarbons like benzo[*a*]pyrene and anthracene (Larson and Koenig, 1994). Furthermore, animal dung combustion produces more

toxic byproducts, including particulates (23% more $PM_{2.5}$ per kg of sample), reactive oxygen species, and microbial products, compared to wood smoke (Sussan et al., 2014, Mudway et al., 2005). Several epidemiological studies have correlated biomass smoke exposure with the risk of development of lung diseases, including chronic obstructive pulmonary disease (COPD), lung cancer and airway infections (Assad et al., 2015, Assad et al., 2016, Gordon et al., 2014, Bruce et al., 2015, Capistrano et al., 2017). However, there are only a limited number of studies that have explored mechanisms in biomass smoke induced-pulmonary inflammation and susceptibility to respiratory infections (Lee et al., 2015). This knowledge gap is in part due to the lack of a standardized low-cost technique for the generation of biomass smoke in the laboratory.

In work by McCarthy and colleagues, biomass smoke from the combustion of horse dung was pumped into a chamber in which human small airway epithelial cells were exposed (McCarthy et al., 2016b). While this method delivered smoke to the epithelial cells, it involved immediate use of the smoke generated and did not allow for storage of batches of biomass smoke for subsequent re-use (McCarthy et al., 2016b). It also required the employment of a cigarette smoking machine (Baumgartner-Jaeger CSM2072i) to generate the smoke. In a study by Li and co-workers, biomass smoke from the combustion of rice chaff was bubbled through the cell culture growth medium, Dulbecco's modified Eagle's medium (Li et al., 2016). Again, this method generated biomass smoke, but the extracts could not be quantified in terms of mass per volume due to the presence of multiple nutrient elements in the growth medium. Furthermore, many components of biomass smoke are not directly soluble in aqueous solutions such as growth medium, and as such will not be retained when the smoke is bubbled through the medium (Li et al., 2016). In work by Huang and co-investigators, wood smoke from the burning of Chinese fir was collected directly onto a glass filter with a 1.6 μm pore size (Huang et al., 2017). This method collected wood smoke particles as intended but it is likely that many of the

smaller components of the smoke, such as volatile organic compounds, from the combustion of the wood would have passed through rather than have been captured on the glass filter.

In our method, the smoke material that was collected in the cotton wool was first incubated in the solvent DMSO, which dissolves both polar and nonpolar compounds, overnight before the filtration step to maximize solubilisation of the components. Furthermore, most of the smoke material in the DMSO was retained after the filter sterilization step based on both the weight and absorbance measurements. By including quantification steps at several of the preparation steps, we were able to determine the concentration of smoke-derived material in mg/mL in each of the smoke extracts. In addition, we were able to generate batches of smoke extracts that could be preserved indefinitely and used in multiple exposure experiments, minimizing inter-assay variation. And importantly, our protocol does not require the purchase of expensive equipment and therefore, is suitable for use in resource-limited situations.

To test our biomass smoke extracts, we compared their effect on the expression of the platelet activating factor receptor (PAFR) on the human bronchial epithelial cells. PAFR is a G-protein coupled seven transmembrane domain receptor, involved in various leucocyte functions, platelet aggregation and inflammation (Ishii et al., 2002). Previous studies have shown that PAFR expression is upregulated in response to a variety of insults including cigarette smoke, e-cigarette vapor, urban particulate matter, and welding fumes (Suri et al., 2016, Grigg et al., 2017, Miyashita et al., 2018, Mushtaq et al., 2011). In terms of infection, PAFR is commonly utilized by the major respiratory bacteria, non-typeable *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Pseudomonas aeruginosa* as a surface receptor for adhesion of airway epithelial cells. These species express a common adhesin, known as phosphorylcholine (ChoP), in their cell wall that recognizes and binds host cell PAFR enabling establishment of infection of the respiratory tract (Grigg et al., 2012, Swords et al., 2000, Barbier et al., 2008).

We determined that PAFR expression is increased in bronchial epithelial cells following exposure to cow dung and wood smoke extracts in a dose-dependent manner at both the protein and mRNA levels (Figures 2.2-2.4, Table 2.1). Moreover, the induction of PAFR expression upon exposure to smoke extracts was comparable over the period of 3 months, which indicates that the smoke content was preserved in DMSO at least for 3 months. The upregulation of PAFR may, therefore, represent a molecular mechanism through which these biomass smoke types could increase susceptibility to lung diseases including airway infections. Furthermore, we detected increased expression of pro-inflammatory mediators IL-1β, IL-6, and IL-8 following exposure of BEAS-2B cells to our CSE, CDSE and WSE preparations in accordance with earlier studies on the effect of smoke on respiratory cells (Table 2.1). Therefore, the ability to produce CDSE and WSE in a usable form, by applying a simple and cost-effective water aspirator-based method, will enable further research on their mechanistic role in the inflammatory response and pathogenesis of respiratory disease including COPD. Furthermore, it will facilitate the discovery of novel therapeutic compounds that reduce the effects of biomass smoke on host cells and tissues of the respiratory system.

3 Chapter 3: Synthesis and biological evaluation of 2-aminothiophenes and 2-pyrrolothiophenes: identification of novel platelet-activating factor receptor (PAFR) antagonists

In the previous chapter (Chapter 2), we developed a technique for generation of quantifiable batches of cigarette smoke extracts. This chapter aimed to utilise the generated cigarette smoke extracts for inducing the adherence of nontypeable *H. influenzae* to bronchial epithelial cells and test the activity of 2-aminothiphenes and 2-pyrrolothiophenes (analogues of WEB-2086) towards inhibiting this adhesion.

This Chapter is presented as a Word document unchanged from the manuscript version that will be submitted to a peer-reviewed journal for publication as

Hyland IK, KC R, Munday E, Kilah NL, O'Toole RF, Bissember AC, Smith JA. Synthesis and biological evaluation of 2-aminothiophenes and 2-pyrrolothiophenes: identification of novel platelet-activating factor receptor (PAFR) antagonists.

3.1 Summary

A small library of novel 2-aminothiophenes and 2-pyrrolothiophenes was constructed via a direct and efficient synthetic strategy. These heterocycles, featuring structurally-distinct core scaffolds, were evaluated for their capacity to antagonise the platelet-activating factor receptor (PAFR) and these results were benchmarked against one of the most well-studied PAFR antagonists WEB-2086. In this way, we determined that one of the new 2-pyrrolothiophenes prepared in this study represents a more potent PAFR antagonist than WEB-2086 in a chronic obstructive pulmonary disease (COPD) model.

3.2 Introduction

More than 40 diseases and disorders are associated with the inflammatory actions of platelet-activating factor (PAF) (**1**) and its receptor (PAFR) (Figure 3.1) (Hyland et al., 2018). This includes cancers, in addition to respiratory and neurodegenerative diseases. Although the structure of the PAF receptor is unknown it is thought to contain a large lipophilic binding pocket that can accommodate considerable steric bulk, a hydrogen-bond donor interacting with either a carbonyl or oxygen atom, and a functional group capable of interactions with pyridine-like moieties (Chung et al., 1995). The PAF receptor has been shown to be upregulated in smokers and chronic obstructive pulmonary disease (COPD) patients (Shukla et al., 2016b).

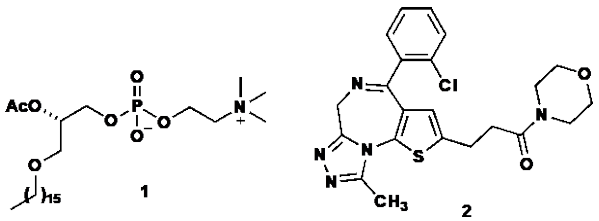


Figure 3.1 PAF (**1**) and WEB-2086 (**2**).

It is estimated that COPD kills over three million people per year worldwide. Heart disease and stroke are thought to be the only other greater causes of death globally, and it is anticipated that without minimising risk factors, the mortality rate due to COPD will increase in the coming decades (Farr et al., 1980, Choi and Koh, 2018). Exacerbations from microorganisms and environmental pollutants are a major cause of morbidity and mortality in COPD (Viniol and Vogelmeier, 2018). *Streptococcus*, *Haemophilus*, and *Pseudomonas* species, in particular, are known to be responsible, as they have the capacity to directly interact with the PAF receptor (Grigg et al., 2012, Shukla et al., 2016a, Barbier et al., 2008, Beasley et al., 2012, Clementi and Murphy, 2011, Iovino et al., 2013).

When pathogens such as non-typeable *Haemophilus influenzae* (NTHi) and *Streptococcus pneumoniae* bind to PAF receptors in the respiratory tract, the normal immune responses can be bypassed, which enables them to colonise and invade the lung epithelium and deeper tissues, causing chronic infection (Beasley et al., 2012, Cundell et al., 1995, Swords et al., 2000).

Furthermore, lung infections can initiate biochemical cascade reactions that result in an altered immune response from the host, which causes a significant increase in pro-inflammatory cytokines (Soler et al., 1999, Sethi et al., 2006, Hill et al., 2000, Bresser et al., 2000). The latter response enhances the incidence of bacterial adhesion to lung epithelium cells by inducing the upregulation of several receptors in the lower respiratory tract, including the PAF receptor (Beasley et al., 2012). Consequently, the upregulation of PAFR expression in airway epithelial and alveolar cells in smokers and COPD patients could provide a gateway to respiratory infections (Shukla et al., 2016b, KC et al., 2017). PAFR antagonists can be used to control these inflammatory responses.

A range of structurally-diverse molecules exhibit PAFR antagonistic activity, however, no pharmaceutical drugs are currently available (Hyland et al., 2018). The diazepine WEB-2086

(**2**) is one of the most well-studied PAFR antagonists and is extremely selective. Its utility (and the ability to prepare novel derivatives) is compromised by the challenges associated with the lengthy and inefficient synthetic route leading to this target (<10% yield over 5 steps) (Weber and Heuer, 1989). Previous studies evaluating the activity of WEB-2086 analogues have explored the effects of varying side-chains and aromatic groups on PAFR antagonism with mixed success (Weber and Heuer, 1989, Walser et al., 1991). Furthermore, given the array of potential off-target therapeutic effects that may be elicited by diazepine-based molecules, we were interested in developing PAFR antagonists that did not comprise this motif (Walser et al., 1991).

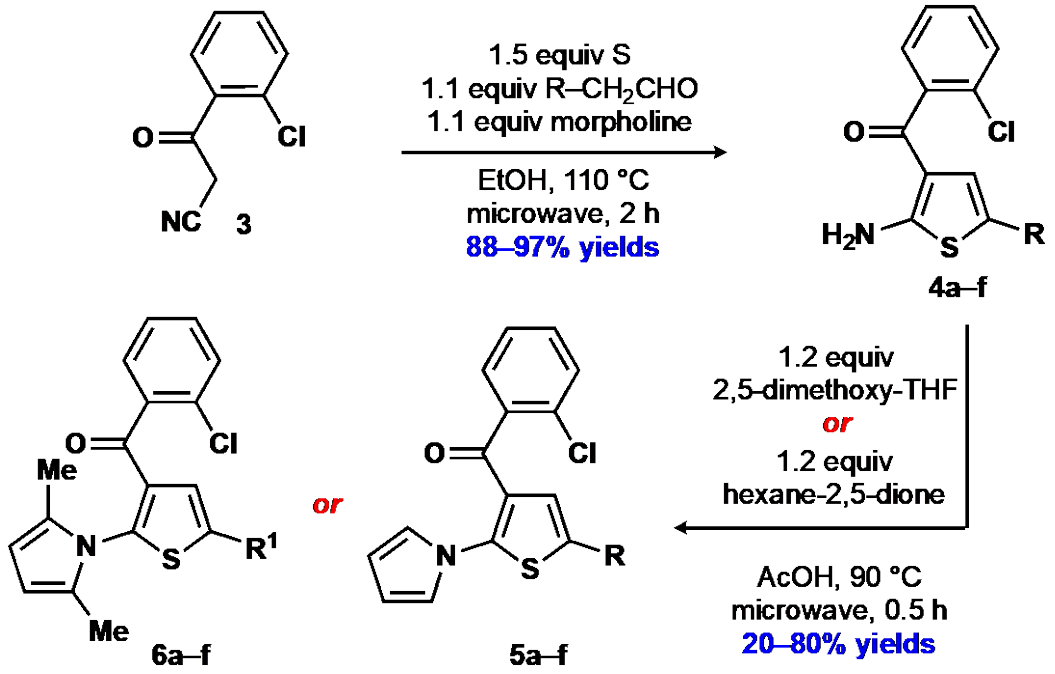
With these issues in mind, the aim of this study was to identify simpler, more readily accessible analogues of WEB-2086 that offer comparable or enhanced PAFR antagonism. To this end, a small library of novel 2-aminothiophenes and 2-pyrrolothiophenes featuring structurally-distinct and conformationally-unrestrained core structures were synthesised and screened in an assay that evaluated PAFR-dependent bacterial adhesion in a human bronchial epithelial cell model of cigarette smoke exposure. These data were benchmarked against results employing WEB-2086 and our findings are reported herein.

3.3 Results and discussion

We were driven to prepare new non-diazepine-containing molecules that retained many of the core elements and functional groups present in WEB-2086 (**2**), but which could be prepared quickly and via a more modular approach. Accordingly, a direct and efficient synthesis was developed to access heterocycles 4–6 (Scheme 1). In the first step, commercially available nitrile **3** was reacted with an appropriate aldehyde or ketone exploiting a multi-component

Gewald-type approach to efficiently deliver 2-aminothiophenes **4** (Figure 3.2) (Aurelio et al., 2010). Compounds **4b–e** featured n-alkyl substituents at the 5-positions, which derived from the aldehyde reactant employed in the condensation reaction. Heterocycle **4f**, synthesised by reaction with cyclohexanone, featured a six-membered ring annulated at the 4- and 5-positions. Amines **4** were then subjected to either Clauson–Kaas or Paal–Knorr reactions to furnish pyrrole rings at the 2-positions of these thiophenes. This afforded rapid access to analogues featuring either unsubstituted pyrrole moieties (**5a–f**) or 2',5'-dimethyl-substituted pyrroles (**6a–f**).

Scaffolds **4a–f**, **5a–f** and **6a–f** differ markedly from the fused, conformationally-restrained core structure of WEB-2086 (**2**). Unlike compound **2**, which is constrained by the predominantly planar arrangement of its ring-fused tricyclic core, all of these molecules allow free rotation about the C–N bond at the 2-position of the thiophene. It is anticipated that this design feature will cause the molecules **4–6** to favour non-planar conformations. Indeed, this is supported by the structures of heterocycles **5a** and **6a** in the solid state, which were confirmed via single-crystal X-ray crystallography, where the angle between the plane of the pyrrole and thiophene rings is 52.19(7)^{o} and 108.64(16)^{o} (major component), respectively (Figure 3.3). The substituents at the 5-positions of thiophenes **4b–f**, **5b–f** and **6b–f** were incorporated to investigate the effects of sterically-varied, non-polar substituents and were guided by germane research focused on the preparation of WEB-2086 derivatives (Walser et al., 1991).



Scheme 1 Synthesis of heterocycles **4a-f**, **5a-f** and **6a-f**.

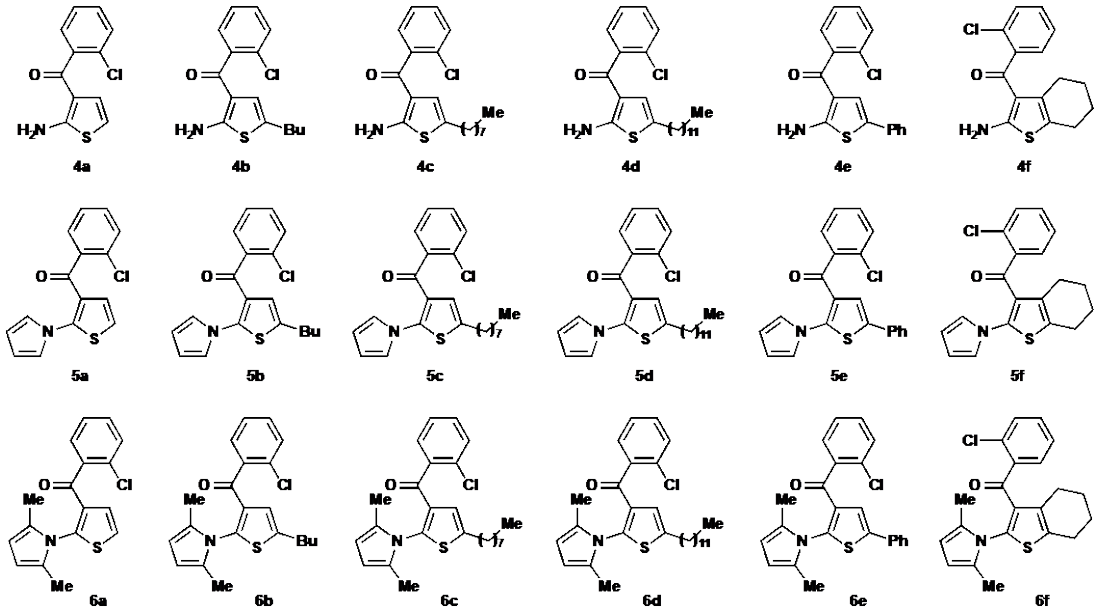


Figure 3.2 2-Aminothiophenes and 2-pyrrolothiophenes synthesised in this study.

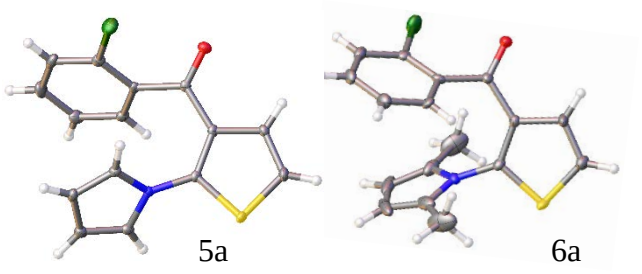


Figure 3.3 Molecular representations of **5a** and **6a**.

Ellipsoids are shown at 50% probability. Angle between the plane of the pyrrole and thiophene rings is 52.19(7)° and 108.64(16)° (major component), respectively.

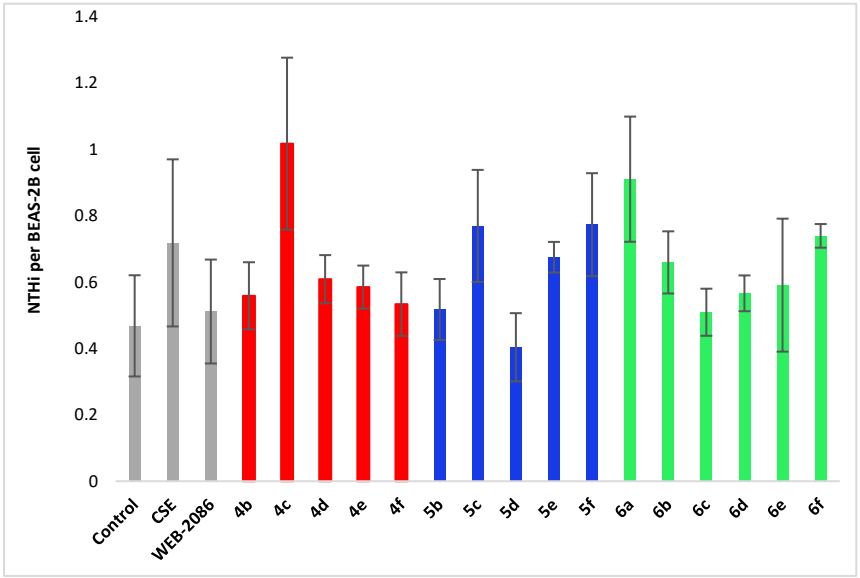


Figure 3.4 Biological evaluation of novel heterocycles **4a–f**, **5a–f** and **6a–f** for PAFR antagonism in lung epithelial cells.

The BEAS-2B cells were exposed to CSE for 4 h followed by treatment with 10 μM of compounds for 1 h and subsequent challenge with NTHi for an additional hour. The data represent the mean ± standard error for triplicates.

The assay used to test these compounds involved treating lung epithelial cells (BEAS-2B) with cigarette smoke extract (CSE) to cause an upregulation in the PAFR levels. Cells were then inoculated with bacteria (NTHi) and the levels of bacterial adhesion in the presence of PAFR antagonists was measured. When cells are treated with WEB-2086, there is a decrease in the level of bacteria binding. The results in Figure 3.4 show that compounds **4b–f** decreased bacterial adherence to levels comparable to that of WEB-2086, with the exception of **4c**, which did not appear to reduce bacterial adhesion with respect to the CSE treated cells (within error).

Compounds **5b–f** with the pyrrole substituent showed mixed results, with one compound in particular, **5d** exhibiting the most reduction in bacterial adhesion. Further dose-response testing showed that compound **5d** was at least as potent in reducing bacterial adhesion as WEB-2086 at 10 μM (Figure 3.5). The dimethylpyrrole compounds **6a–f** appeared to be slightly less effective than compounds **4b–f** or **5b–f**, although results were mixed. It was not possible to draw any structure-activity relationship (SAR) conclusions relating to the substituent at the 5-position of the thiophene. It should be noted that some of the error bars in Figure 3.4 are large, and this is due to poor adhesion of cells to the growth surface after CSE stimulation.

The results obtained from this small library of compounds did not allow any discernible SARs to be identified. However, the results are consistent with relevant studies that propose the presence of a lipophilic pocket in the active site (Weber and Heuer, 1989, Walser et al., 1991).

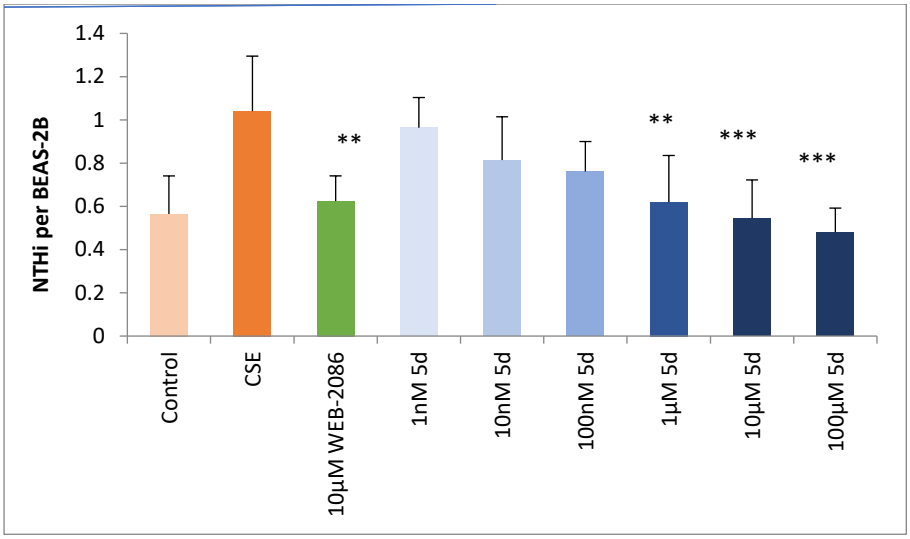


Figure 3.5 Inhibition of CSE-induced NTHi adhesion to BEAS-2B cells by WEB-2086 and compound **5d**.

The BEAS-2B cells were exposed to CSE for 4 h followed by treatment with 10 µM WEB-2086 or compound **5d** at a concentration range of 1 nM to 100 µM for 1 h and subsequent challenge with NTHi for an additional hour. The data represent the mean ± standard error for triplicates; **$p < 0.01$, ***$p < 0.001$ relative to CSE-exposed BEAS-2B cells; one-way analysis of variance with Dunnett's multiple comparison test.

In an Alamar Blue toxicity assay, 2-pyrrolothiophene **5d** was found to be non-toxic; tested up to 100 µM over 24 h (Figure 3.6). When compounds **4a–f**, **5a–f** and **6a–f** were screened against *S. aureus* (MRSA), *E. coli*, *K. pneumonia* (MDR), *A. baumannii*, *P. aeruginosa*, *C. albicans* and *C. neoformans*, no antibacterial or antifungal activity was observed in each case. Importantly, this indicates that observed reduction in NTHi adhesion to human bronchial epithelial cells was not due to any direct bacteriostatic or bactericidal effects of the compounds but is more consistent with an effect on bacterial interaction involving the PAF receptor.

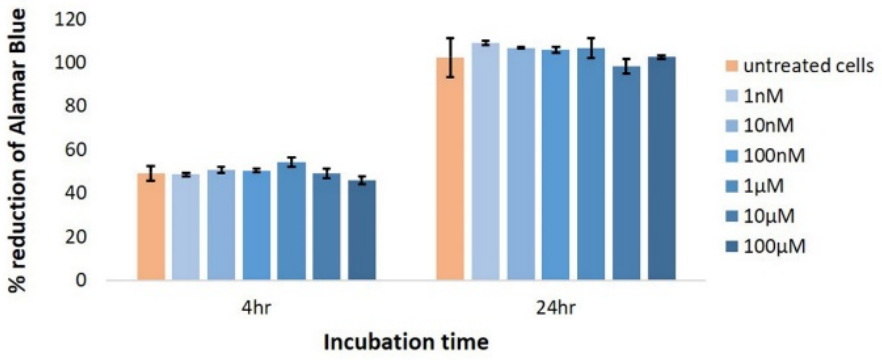


Figure 3.6 Alamar Blue toxicity assay for compound **5d**.

The data represent the mean ± standard error for triplicates; one-way analysis of variance with Bonferroni multiple comparison test.

3.4 Conclusion

We have demonstrated that the predominantly planar thienodiazepine framework of the PAF receptor antagonist WEB-2086 can be replaced by an unfused, less conformationally-restrained 2-aminothiophene core structure and still retain inhibitory activity towards NTHi adhesion to bronchial epithelial cells. Specifically, we have identified that heterocycle **5d**, which we prepared in two steps. represents a potent PAFR antagonist in a human bronchial epithelial cell model of cigarette smoke exposure. Importantly, molecule **5d** was found to be non-toxic. Future studies will focus on constructing a larger library of analogues of structures **4–6** in order to investigate SARs and more extensively evaluate the biological activity of lead compound **5d** and perform screening in a mouse model for COPD.

3.5 Experimental

3.5.1 General

Unless otherwise specified, reactions were performed under an atmosphere of air. Analytical-grade solvents were used, which were purified by standard laboratory procedures. Reagents were purchased from Sigma-Aldrich, AK Scientific, Combi-Blocks, and Oakwood and were used without purification.

Infrared (IR) spectroscopy was performed on a Shimadzu FTIR 8400s spectrometer. Samples for subjected to IR analysis were prepared by the deposition of a thin film of the compound onto NaCl plates following evaporation of CH_2Cl_2 or $CHCl_3$. NMR experiments were performed either on a Bruker Avance III NMR spectrometer operating at 400 MHz (^{1}H) or 100 MHz (^{13}C) or on a Bruker Avance III NMR spectrometer operating at 600 MHz (^{1}H) or 150 MHz (^{13}C). The deuterated solvents used were either $CDCl_3$ or CD_3OD as specified. Chemical shifts were recorded in ppm. Spectra were calibrated by assignment of the residual solvent peak to δ_H 7.26 and δ_C 77.16 for $CDCl_3$ and δ_H 3.31 and δ_C 49.0 for CD_3OD. Coupling constants (*J*) were recorded in Hz. High resolution, accurate mass spectra were acquired using a Thermo LTQ XL Orbitrap tandem mass spectrometer. Samples in methanol were analysed by direct infusion at a rate of 20 μL/min and ionised by Electrospray ionisation (ESI) in the positive mode. Typical ESI conditions were: Sheath gas flow rate (8), Auxiliary gas flow rate (2), Spray Voltage (4.1 kV), Capillary Temperature (320 °C), Capillary voltage (50 V), Tube lens (50 V). Mass spectra were acquired in full scan mode over an appropriate range with an ion inject time of 500 ms and resolution of 30,000. All microwave experiments were carried out in a CEM Microwave reactor. X-ray crystallographic data for the structure determination of compounds **5a** and **6a** were recorded on a Bruker AXS D8 Quest (Cu Kα radiation). Full details of the data collection and refinement are included in the Supporting Information.

Crystallographic data have been deposited with the Cambridge Crystallographic Data Centre (1921328–1921329). TLC was performed using Merck silica gel 60-F_{254} plates. Developed chromatograms were visualised by UV absorbance (254 nm) or through application of heat to a plate stained with cerium molybdate. Flash column chromatography was performed with flash grade silica gel (60 μm) and the indicated eluent in accordance with standard techniques.

3.5.2 Chemistry

General Procedure A: synthesis of compounds 4b–f. 2-Chlorobenzoylacetonitrile (3) (1.0 equiv), aldehyde/ketone (1.1 equiv), sulfur (1.5 equiv), morpholine (1 equiv) and EtOH (0.4 mM) were combined. The mixture was heated at reflux. After 2 h, the mixture was cooled, transferred to a round-bottom flask and concentrated under reduced pressure. The ensuing residue was then subjected to flash column chromatography (silica gel).

General Procedure B: synthesis of compounds 5a–f. Amines 4a–f (1 equiv), 2,5-dimethoxytetrahydrofuran (1.2 equiv) and AcOH (0.3 mM) were added to a microwave reactor tube (10-mL). The mixture was heated in a microwave reactor at 90 °C. After 0.5 h, the mixture was cooled, made alkaline by the addition of $NaHCO_3$ (saturated aqueous solution), and extracted with CH_2Cl_2 (3 x 15 mL). The combined organic phases were dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The ensuing residue was then subjected to flash column chromatography (silica gel).

General Procedure C: synthesis of compounds 6a–f. Amines 4a–f (1 equiv), hexane-2,5-dione (1.2 equiv) and AcOH (0.3 mM) were added to a microwave reactor tube (10-mL). The mixture was heated in a microwave reactor at 90 °C. After 0.5 h, the mixture was cooled, made alkaline by the addition of $NaHCO_3$ (saturated aqueous solution), and extracted with CH_2Cl_2 (3 x 15

mL). The combined organic phases were dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The ensuing residue was then subjected to flash column chromatography (silica gel).

*(2-Aminothiophen-3-yl)(2-chlorophenyl)methanone (**4a**).* 2-Chlorobenzoylacetonitrile (**3**) (993 mg, 5.53 mmol; 1 equiv), dithiane-2,5-dione (856 mg, 5.63 mmol; 1 equiv), NEt_3 (0.8 mL; 1 equiv) and THF (2 mL) were combined. The mixture was heated at reflux. After 1 h, the mixture was cooled, transferred and concentrated under reduced pressure. The ensuing residue was dissolved in EtOAc (15 mL) and then successively washed with $KHSO_4$ (2 x 10 mL of a saturated aqueous solution) and $NaHCO_3$ (2 x 10 mL of a saturated aqueous solution). The organic phase was then dried ($MgSO_4$), filtered, and concentrated under reduced pressure. The ensuing residue was subjected to flash chromatography (silica gel; 20% EtOAc/ hexanes) to provide compound **4a** as a brown semi-solid in 92% yield (259 mg, 1.09 mmol) (Aurelio et al., 2010).

δ_H (600 MHz, $CDCl_3$): 7.43–7.42 (m, 1H), 7.35–7.31 (m, 3H), 6.46 (d, J = 5.8 Hz, 1H), 6.06 (d, J = 5.8 Hz, 1H); δ_C (150 MHz, $CDCl_3$): 189.2, 166.5, 140.2, 130.5, 130.4, 130.0, 128.4, 127.2, 126.6, 114.7, 106.5; ν_{max}/cm^{-1} (NaCl): 3279, 1589, 1447, 1318, 1056, 838, 757, 649. HRMS [M + H]: for $C_{11}H_9ClNOS$, predicted 238.0093, found 238.0088.

*(2-Amino-5-butylthiophen-3-yl)(2-chlorophenyl)methanone (**4b**).* Previously unreported compound **4b** was prepared from substrate 3 (1.00 g, 5.57 mmol) and hexanal (0.70 mL, 5.70 mmol) following General Procedure A (40% EtOAc/ hexanes) as a brown oil in 97% yield (1.59 g, 5.41 mmol).

δ_H (600 MHz, $CDCl_3$): 7.36–7.35 (m, 1H), 7.29–7.23 (m, 3H), 6.01 (s, 1H), 2.43 (t, J = 7.6 Hz, 2H), 1.43 (quin. J = 7.6 Hz, 2H), 1.25 (sext, J = 7.5 Hz, 2H), 0.82 (t, J = 7.3 Hz, 3H); δ_C (150

MHz, $CDCl_3$): 188.7, 165.4, 140.5, 130.6, 130.2, 129.9, 128.4, 126.6, 126.4, 122.5, 115.3, 33.1, 26.5, 22.1, 13.8; ν_{max}/cm^{-1} (NaCl): 3273, 2929, 1594, 1450, 1308, 1122, 1055, 755, 694, 644. HRMS [M + H]: for $C_{15}H_{17}ClNOS$, predicted 294.0719, found 294.0714.

*(2-Amino-5-octylthiophen-3-yl)(2-chlorophenyl)methanone (**4c**).* Previously unreported compound 4c was prepared from substrate 3 (499 mg, 2.78 mmol) and n-decylaldehyde (519 mg, 3.32 mmol) following General Procedure A (15% EtOAc/ hexanes) as a yellow oil in 89% yield (442 mg, 1.26 mmol).

δ_H (600 MHz, CD_3OD): 7.49 (dd, J = 7.9 and 1.2 Hz, 1H), 7.41 (dtd, J = 21.3, 7.4 and 1.8 Hz, 2H), 7.34 (dd, J = 7.4 and 1.7 Hz, 1H), 6.00 (s, 1H), 2.50 (td, J = 7.4 and 0.9 Hz, 2H), 1.54–1.49 (m, 2H), 1.33–1.28 (m, 10H), 0.90 (t, J = 7.0 Hz, 3H); δ_C (150 MHz, CD_3OD): 188.1, 168.1, 140.6, 130.1, 130.0, 129.4, 127.9, 126.6, 125.7, 121.6, 113.8, 31.6, 30.4, 29.1, 28.9 (2C), 28.4, 22.3, 13.0; ν_{max}/cm^{-1} (NaCl): 3277, 2925, 1586, 1447, 1304, 1120, 1056, 754, 644. HRMS [M + H]: for $C_{19}H_{25}ClNOS$, predicted 350.1345, found 350.1338.

*(2-Amino-5-dodecylthiophen-3-yl)(2-chlorophenyl)methanone (**4d**).* Previously unreported compound 4d was prepared from substrate 3 (504 mg, 2.81 mmol) and tetradecylaldehyde (656 mg, 2.64 mmol) following General Procedure A (15% EtOAc/ hexanes) as a brown oil in 88% yield (916 mg, 2.26 mmol).

δ_H (600 MHz, $CDCl_3$): 7.43 (m, 1H), 7.38–7.30 (m, 3H), 6.09 (s, 1H), 2.50 (t, J = 7.5 Hz, 2H), 1.56–1.49 (m, 2H), 1.36–1.24 (complex m, 18H), 0.90 (t, J = 6.8 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.6, 165.8, 140.5, 130.6, 130.2, 129.9, 129.4, 126.6, 126.4, 122.4, 115.1, 31.9, 30.9, 29.8, 29.7 (2C), 29.6 (2C), 29.4, 29.3, 29.0, 22.7, 14.1; ν_{max}/cm^{-1} (NaCl): 3252, 2927, 1704, 1525, 1235, 1037, 753, 646. HRMS [M + H]: for $C_{23}H_{33}ClNOS$, predicted 406.1971, found 406.1963.

(2-Amino-5-phenylthiophen-3-yl)(2-chlorophenyl)methanone ***(4e)***. Previously unreported compound 4e was prepared from substrate 3 (1.00 g, 5.57 mmol) and phenylacetaldehyde (735 mg, 6.12 mmol) following General Procedure A (20% EtOAc/ hexanes) as a yellow oil in 97% yield (1.59 g, 5.41 mmol).

δ_H (600 MHz, $CDCl_3$): 7.66–7.40 (complex m, 10H); δ_C (150 MHz, $CDCl_3$): 186.1, 153.4, 156.0, 134.0, 132.9, 131.8, 131.7, 131.0, 130.5, 130.3, 129.2, 127.7, 127.5, 126.9, 120.4, 117.4, 114.1; ν_{max}/cm^{-1} (NaCl): 3060, 1615, 1433, 1337, 1054, 759, 736, 644. HRMS [M + Na]: for $C_{17}H_{12}ClNOSNa$, predicted 336.0226, found 336.0210.

(2-Amino-4,5,6,7-tetrahydrobenzo[b]thiophen-3-yl)(2-chlorophenyl)methanone ***(4f)***. Previously unreported compound 4e was prepared from substrate 3 (1.00 g, 5.57 mmol) and cyclohexanone (0.65 mL, 6.29 mmol) following General Procedure A (20% EtOAc/ hexanes) as a yellow solid in 90% yield (2.06 g, 7.11 mmol). mp 82 °C.

δ_H (400 MHz, $CDCl_3$): 7.41–7.25 (m, 4H), 2.51–2.48 (m, 2H), 1.72–1.68 (m, 4H), 1.52–1.49 (m, 2H); δ_C (100 MHz, $CDCl_3$): 189.1, 166.6, 141.9, 130.7, 130.3, 129.9, 129.6, 127.8, 126.7, 118.0, 115.2, 25.9, 24.7, 22.9, 22.7; ν_{max}/cm^{-1} (NaCl): 3390, 2933, 1564, 1424, 1293, 754, 644. HRMS [M + H]: for $C_{15}H_{15}ClNOS$, predicted 292.0563, found 292.0558.

(2-(1H-pyrrol-1-yl)thiophen-3-yl)(2-chlorophenyl)methanone ***(5a)***. Previously unreported compound 5a was prepared from substrate 4a (504 mg, 2.12 mmol) following General Procedure B (20% EtOAc/ hexanes) as brown semi-solid in 60% yield (299 mg, 1.26 mmol).

δ_H (400 MHz, $CDCl_3$): 7.29–7.28 (m, 4H), 7.17 (m, 1H), 7.10 (d, J = 5.8 Hz, 1H), 6.76–6.75 (m, 2H), 6.07–6.06 (m, 2H); δ_C (100 MHz, $CDCl_3$): 188.5, 148.7, 138.2, 131.6 (2C), 131.5,

130.2, 129.3, 128.5, 126.5, 123.6 (2C), 120.7, 110.8 (2C); ν_{max}/cm^{-1} (NaCl): 1591, 1515, 1065, 735, 648. HRMS [M + H]: for $C_{15}H_{11}ClNOS$, predicted 288.0250, found 288.0244.

(5-Butyl-2-(1H-pyrrol-1-yl)thiophen-3-yl)(2-chlorophenyl)methanone ***(5b)****.* Previously unreported compound 5b was prepared from substrate 4b (229 mg, 0.78 mmol) following General Procedure B (40% CH_2Cl_2/ hexanes) as a brown oil in 60% yield (159.6 mg, 0.46 mmol).

δ_H (400 MHz, $CDCl_3$): 7.29–7.26 (m, 3H), 7.14 (m, 1H), 6.98 (s, 1H), 6.71 (s, 2H), 6.01 (s, 2H), 2.79 (t, *J* = 7.6 Hz, 2H), 1.74–1.66 (m, 2H), 1.50–1.42 (m, 2H), 0.98 (t, *J* = 7.3 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.6, 146.3, 141.1, 138.4, 131.7, 131.5, 131.3, 130.1, 129.2, 126.4, 124.6, 123.5 (2C), 110.4 (2C), 33.3, 29.8, 22.1, 13.8; ν_{max}/cm^{-1} (NaCl): 2930, 1656, 1514, 1303, 1065, 726, 646. HRMS [M + H]: for $C_{19}H_{19}ClNOS$, predicted 344.0876, found 344.0870.

(2-Chlorophenyl)(5-octyl-2-(1H-pyrrol-1-yl)thiophen-3-yl)methanone ***(5c)****.* Previously unreported compound 5c was prepared from substrate 4c (205 mg, 0.58 mmol) following General Procedure B (30% EtOAc/ hexanes) as a brown oil in 21% yield (49 mg, 1.2 mmol).

δ_H (400 MHz, CD_3OD): 7.32–7.28 (m, 3H), 7.20 (m, 1H), 6.97 (s, 1H), 6.70 (t, *J* = 2.1 Hz, 2H), 5.96 (t, *J* = 2.1 Hz, 2H), 2.81 (t, *J* = 7.4 Hz, 2H), 1.71 (quin, *J* = 7.5 Hz, 2H), 1.44–1.32 (complex m, 10H), 0.92 (t, *J* = 7.0 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.6, 146.3, 141.1, 138.4, 131.6, 131.5, 131.3, 130.1, 129.2, 126.3, 124.5, 123.5 (2C), 110.4 (2C), 31.8, 31.2, 30.1, 29.3, 29.2, 29.0, 22.7, 14.1; ν_{max}/cm^{-1} (NaCl): 2926, 1715, 1436, 1236, 752, 647. HRMS [M + H]: for $C_{23}H_{27}ClNOS$, predicted 400.1502, found 400.1497.

(2-Chlorophenyl)(5-dodecyl-2-(1H-pyrrol-1-yl)thiophen-3-yl)methanone ***(5d)***. Previously unreported compound **5d** was prepared from substrate **4d** (111 mg, 0.27 mmol) following General Procedure B (50% CH_2Cl_2/ hexanes) as a yellow oil in 67% yield (83 mg, 0.18 mmol).

δ_H (400 MHz, $CDCl_3$): 7.28–7.23 (m, 3H), 7.18–7.12 (m, 1H), 6.98 (s, 1H), 6.71 (t, J = 2.2 Hz, 2H), 6.02 (t, J = 2.2 Hz, 2H), 2.78 (t, J = 7.6 Hz, 2H), 1.71 (quin, J = 7.4 Hz, 2H), 1.43–1.30 (m, 18H), 0.91 (t, J = 6.8 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.6, 146.3, 141.2, 138.4, 131.6, 131.5, 131.3, 130.1, 129.2, 126.3, 124.5, 123.5 (2C), 110.4 (2C), 31.9, 31.2, 30.1, 29.7 (2C), 29.6, 29.5, 29.4, 29.3, 29.0, 22.7, 14.1; ν_{max}/cm^{-1} (NaCl): 2925, 1651, 1516, 1234, 753, 646. HRMS [M + Na]: for $C_{27}H_{34}ClNOSNa$, predicted 478.1947, found 478.1935.

(2-Chlorophenyl)(5-phenyl-2-(1H-pyrrol-1-yl)thiophen-3-yl)methanone ***(5e)***. Previously unreported compound **5e** was prepared from substrate **4e** (1389 mg, 0.44 mmol) following General Procedure B (70% CH_2Cl_2/ hexanes) as brown semi-solid in 22% yield (36 mg, 0.10 mmol).

δ_H (400 MHz, $CDCl_3$): 7.61–7.59 (m, 2H), 7.49 (s, 1H), 7.46–7.42 (m, 2H), 7.39–7.28 (m, 4H), 7.20–7.16 (m, 1H), 6.79 (t, J = 2.1 Hz, 2H), 6.07 (t, J = 2.1 Hz, 2H); δ_C (100 MHz, $CDCl_3$): 188.5, 147.4, 139.0, 138.1, 132.8, 132.6, 131.6 (2C), 130.2, 129.3, 129.1 (2C), 128.5, 126.5, 125.8 (2C), 123.5 (2C), 123.4, 110.8 (2C); ν_{max}/cm^{-1} (NaCl): 1652, 1505, 1379, 1222, 1060, 750, 729, 645. HRMS [M + H]: for $C_{21}H_{15}ClNOS$, predicted 364.0563, found 364.0554.

(2-(1H-Pyrrol-1-yl)-4,5,6,7-tetrahydrobenzo[b]thiophen-3-yl)(2-chlorophenyl)methanone ***(5f)***. Previously unreported compound **5f** was prepared from substrate **4f** (103 mg, 0.35 mmol) following General Procedure B (40% CH_2Cl_2/ hexanes) as a brown semi-solid in 61% yield (73 mg, 0.22 mmol).

δ_H (400 MHz, $CDCl_3$): 7.28 (m, 1H), 7.23–7.22 (m, 2H), 7.09 (m, 1H), 6.60 (bs, 2H), 5.93 (bs, 2H), 2.75 (dt, *J* = 20.5 and 5.5 Hz, 4H), 1.95–1.82 (m, 4H); δ_C (100 MHz, $CDCl_3$): 190.5, 144.5, 138.0, 134.7, 132.6, 132.2, 132.1, 131.7, 130.2, 130.1, 126.2, 123.5 (2C), 110.2 (2C), 25.5, 24.9, 23.0, 22.5; ν_{max}/cm^{-1} (NaCl): 2936, 1657, 1514, 1435, 1333, 1218, 1066, 1032, 871, 752, 647. HRMS [M + H]: for $C_{19}H_{17}ClNOS$, predicted 342.0719, found 342.0712.

*(2-Chlorophenyl)(2-(2,5-dimethyl-1H-pyrrol-1-yl)thiophen-3-yl)methanone (**6a**).* Previously unreported compound **6a** was prepared from substrate **4a** (101 mg, 0.43 mmol) following General Procedure C (60% CH_2Cl_2/ hexanes) as a pale-yellow solid in 80% yield (107 mg, 0.34 mmol). mp 109 °C.

δ_H (600 MHz, $CDCl_3$): 7.25–7.18 (m, 4H), 7.13–7.08 (m, 2H), 5.61 (s, 2H), 1.96 (s, 6H); δ_C (150 MHz, $CDCl_3$): 188.8, 145.6, 138.3, 136.9, 131.2, 130.9, 130.1, 129.8, 128.6, 128.3 (2C), 126.5, 123.6, 107.3 (2C), 12.7 (2C); ν_{max}/cm^{-1} (NaCl): 1656, 1534, 1382, 1306, 1262, 861, 761, 738, 648. HRMS [M + H]: for $C_{17}H_{15}ClNOS$, predicted 316.0563, found 316.0555.

*(5-Butyl-2-(2,5-dimethyl-1H-pyrrol-1-yl)thiophen-3-yl)(2-chlorophenyl)methanone (**6b**).* Previously unreported compound **6b** was prepared from substrate **4b** (110 mg, 0.37 mmol) following General Procedure C (40% CH_2Cl_2/ hexanes) as a brown oil in 47% yield (66 mg, 0.18 mmol).

δ_H (400 MHz, $CDCl_3$): 7.31–7.26 (m, 2H), 7.18 (m, 2H), 7.00 (s, 1H), 5.67 (s, 2H), 2.82 (t, *J* = 7.6 Hz, 2H), 2.06 (s, 6H), 1.72 (quin, *J* = 7.5 Hz, 2H), 1.46 (sext, *J* = 7.4 Hz, 2H), 0.99 (t, *J* = 7.3 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.9, 143.9 (2C), 142.9, 138.5, 136.9, 131.0, 130.8, 130.1, 129.7, 128.5, 126.4, 124.4, 107.0 (2C), 33.2, 30.0, 22.2, 13.8, 12.7 (2C); ν_{max}/cm^{-1} (NaCl): 2958, 1675, 1521, 1436, 1235, 754, 644. HRMS [M + H]: for $C_{21}H_{23}ClNOS$, predicted 372.1189, found 372.1180.

(2-Chlorophenyl)(2-(2,5-dimethyl-1H-pyrrol-1-yl)-5-octylthiophen-3-yl)methanone ***(6c)***. Previously unreported compound **6c** was prepared from substrate **4c** (147 mg, 4.2 mmol) following General Procedure C (15% EtOAc/ hexanes) as a brown oil in 20% yield (36 mg, 0.09 mmol).

δ_H (400 MHz, CD_3OD): 7.36–7.34 (m, 2H), 7.24–7.22 (m, 2H), 7.01 (s, 1H), 5.59 (s, 2H), 2.86 (t, J = 7.5 Hz, 2H), 2.02 (s, 6H), 1.73 (quint, J = 7.5 Hz, 2H), 1.44–1.32 (complex m, 10H), 0.92 (t, J = 7.0 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.9, 144.0, 142.9, 138.5, 136.9, 131.0 (2C), 130.8, 130.1, 129.7, 128.5, 126.4, 124.4, 107.0 (2C), 31.8, 31.1, 30.3, 29.2 (2C), 29.0, 22.7, 14.1, 12.7 (2C); ν_{max}/cm^{-1} (NaCl): 2927, 1669, 1521, 1436, 1236, 754, 646. HRMS [M + H]: for $C_{25}H_{31}ClNOS$, predicted 428.1815, found 428.1804.

(2-Chlorophenyl)(2-(2,5-dimethyl-1H-pyrrol-1-yl)-5-dodecylthiophen-3-yl)methanone ***(6d)***. Previously unreported compound **6d** was prepared from substrate 4d (92 mg, 2.3 mmol) following General Procedure C (20% EtOAc/ hexanes) as a brown oil in 44% yield (48 mg, 0.01 mmol).

δ_H (400 MHz, $CDCl_3$): 7.33–7.28 (m, 2H), 7.19–7.17 (m, 2H), 7.00 (s, 1H), 5.67 (s, 2H), 2.81 (t, J = 7.6 Hz, 2H), 2.06 (s, 6H), 1.71 (quin, J = 7.6 Hz, 2H), 1.42–1.29 (complex m, 18H), 0.91 (t, J = 6.8 Hz, 3H); δ_C (100 MHz, $CDCl_3$): 188.9, 144.0, 142.9, 138.5, 136.9, 131.0, 130.8, 130.1 (2C), 129.7, 128.5, 126.3, 124.4, 107.0 (2C), 31.9, 31.1, 30.3, 29.7, 29.6 (2C), 29.5, 29.4, 29.3, 29.0, 22.7, 14.1, 12.7 (2C); ν_{max}/cm^{-1} (NaCl): 2925, 1669, 1520, 1436, 1235, 756, 644. HRMS [M + H]: for $C_{29}H_{39}ClNOS$, predicted 484.2440, found 484.2431.

(2-Chlorophenyl)(2-(2,5-dimethyl-1H-pyrrol-1-yl)-5-phenylthiophen-3-yl)methanone ***(6e)***. Previously unreported compound **6e** was prepared from substrate **4e** (113 mg, 0.36 mmol)

following General Procedure C (15% EtOAc/ hexanes) as a yellow oil in 19% yield (24 mg, 0.06 mmol).

δ_H (400 MHz, CD_3OD): 7.68 (d, J = 8.0 Hz, 2H), 7.60 (s, 1H), 7.46 (t, J = 7.4 Hz. 2H), 7.41–7.35 (m, 3H), 7.30–7.24 (m, 2H), 5.61 (s, 2H), 2.08 (s, 6H); δ_C (150 MHz, CD_3OD): 189.3, 143.9, 142.3, 138.5, 138.0, 132.7, 131.3, 130.3, 129.7, 129.5, 128.9 (2C), 128.4 (3C), 126.5, 125.4 (2C), 122.6, 107.2 (2C), 11.5 (2C); ν_{max}/cm^{-1} (NaCl): 1670, 1591, 1505, 1435, 1235, 1109, 754, 736, 647. HRMS [M + H]: for $C_{23}H_{19}ClNOS$, predicted 392.0876, found 392.0868.

*(2-Chlorophenyl)(2-(2,5-dimethyl-1H-pyrrol-1-yl)-4,5,6,7-tetrahydrobenzo[b]thiophen-3-yl)methanone (**6f**)*. Previously unreported compound **6f** was prepared from substrate **4f** (108 mg, 0.37 mmol) following General Procedure C (40% CH_2Cl_2/ hexanes) as a brown oil in 46% yield (63 mg, 0.17 mmol).

δ_H (400 MHz, $CDCl_3$): 7.24–7.22 (m, 2H), 7.09–7.08 (m, 2H), 5.52 (s, 2H), 2.78 (dt, J = 14.7 and 5.6 Hz, 4H), 1.98–1.86 (complex m, 10H); δ_C (100 MHz, $CDCl_3$): 192.3, 140.5, 138.8, 137.8, 135.2, 134.5, 131.2 (2C), 130.8, 130.2, 129.3, 128.4, 126.5, 106.9 (2C), 25.2, 25.1, 23.2, 22.4, 12.6 (2C); ν_{max}/cm^{-1} (NaCl): 2943, 1662, 1436, 1037, 754, 647. HRMS [M + Na]: for $C_{21}H_{20}ClNOSNa$, predicted 392.0852, found 392.0847.

3.5.3 Biology

Human bronchial epithelial cells (BEAS-2B) were cultured in the serum free bronchial epithelial basal medium (BEBM) supplemented with the BulletKit (BEGM) at 37°C and 5% CO_2. The BEAS-2B cells were sub-cultured in a T-75 and were used for the experiments at

passage numbers ≤15 passages. The clinical strain of NTHi bacteria, RHH3 were labelled with fluorescein isothiocyanate (FITC). Briefly, bacteria were grown overnight in brain heart infusion medium supplemented with hemin (1 mg/mL) and nicotinamide adenine dinucleotide (NAD, 10 mg/mL) and then bacteria (108 CFU/mL) were incubated for 1 h at 4^{o}C with 1 mg/mL (w/v) FITC prepared in a carbonate buffer (0.9 M sodium carbonate, 0.015 M sodium bicarbonate and 0.1 M sodium chloride). The FITC-labelled bacteria were washed three times with phosphate buffered saline (PBS) and resuspended in BEGM without antibiotics at a density of 5 x 10^7 CFU/mL. BEAS-2B cells were transferred to sterile 8-well chambered glass slides pre-coated with 5% (v/v) bovine collagen I at a density of 30,000 cells per well and cultured for an additional 24 h to form a confluent monolayer. The cell monolayer was then exposed to CSE at a concentration of 87.5 µg/mL. After exposure to the CSE, the BEAS-2B cells were washed with pre-warmed PBS and treated with PAFR antagonists. Compounds were prepared in DMSO at 10mM concentration and then diluted in antibiotic-free BEGM to a final concentration of 10 µM for initial screening and 100 µM, 10 µM, 1 µM, 100 nM, 10 nM and 1 nM for compound 5d. The BEAS-2B cells were incubated with 200 µL of different concentrations of compound for 1 h at 37^{o}C and 5% CO_2. FITC-tagged NTHi (100 µL) was added to the monolayer at a multiplicity of infection of 50:1. The cells were incubated for 1 h at 37^{o}C and 5% CO_2 and then washed three times with pre-warmed PBS.

Cells were analysed via immunocytochemistry and analysis of microscopy images. BEAS-2B cells were fixed with 200 µL of 4% (w/v) paraformaldehyde for 20 min at r.t. After two washes with PBS, cells were permeabilised with 100 µL of chilled (–20^{o}C) acetone for 10 min at r.t. Then cells were washed three times with PBS and blocked with 200 µL of 1% (w/v) bovine serum albumin (BSA), prepared in 1% (v/v) Tween-20-PBS for 1 h at r.t. The cells were then incubated with mouse anti-human platelet activating factor receptor monoclonal antibody (anti-PAFR mAb, 2.5 µg/mL) overnight in the dark at 4^{o}C. After here washed with 0.1% (w/v) BSA,

cells were incubated with 1:100 dilutions of Alexa Fluor 594-conjugated goat anti-mouse IgG (H+L) secondary antibody for 1 h at r.t. The cells were then rinsed three times with 0.1% (w/v) BSA, followed by staining with 4',6-diamidino-2-phenylindole (DAPI, 1 µg/ mL) in the dark at r.t for 0.25 h. Finally, the cells were washed three times with PBS and air-dried before the slides were mounted with Dako fluorescence mounting media. Cells were examined with an Olympus BX50 epifluorescence microscope equipped with NIS elements software and Cool Snap Hq2 CCD camera under 400x magnification. Ten images were taken per well from different points using three different fluorescence channels for DAPI (violet excitation and blue emission, 200 ms exposure), Alexa Fluor 594 (green excitation and red emission, 300 ms exposure) and FITC (blue excitation and green emission, 500 ms exposure). Image merging and bacterial and BEAS-2B cell counting was performed using the software ImageJ.

3.5.4 Compound toxicity

Cells were seeded into sterile clear flat-bottom 96-well plates at a density of 5,000 per well and incubated overnight at 37^{o}C and 5% CO_2. The cells were exposed to different concentrations of the compound, followed by Alamar Blue at a concentration of 10% (v/v). The cells were then incubated for 4 h and 24 h at 37^{o}C and 5% CO_2 to allow cells to convert resazurin (blue) to resorufin (red).

4 Chapter 4: Cow Dung Biomass Smoke Exposure Increases Adherence of Respiratory Pathogen Nontypeable *Haemophilus influenzae* to Human Bronchial Epithelial Cells

In the previous chapter (Chapter 3), we identified a compound that exhibited significant inhibitory activity towards adhesion of nontypeable *H. influenzae* to cigarette smoke exposed human bronchial epithelial cells. This chapter aimed to investigate the *in vitro* effect of biomass smoke exposure on adherence of nontypeable *H. influenzae* to airway epithelial cells, and utilise the compound (identified in Chapter 3) to establish a relationship between bacterial attachment and the host cell receptor, platelet activating factor receptor.

This Chapter is presented as a Word document unchanged from the final accepted version published in a peer-reviewed journal with an exception of a minor change in Figure 4.3 in response to the Examiner's recommendation. In Figure 4.3, information on the type of smoke has been added in the graphs and Y-axes of all the graphs have been modified to a similar scale of 0-5.

KC R, Hyland IK, Smith JA, Shukla SD, Hansbro PM, Zosky GR, Karupiah G, O'Toole RF. Cow Dung Biomass Smoke Exposure Increases Adherence of Respiratory Pathogen Nontypeable *Haemophilus influenzae* to Human Bronchial Epithelial Cells. 2020. Reused by permission from **Springer Nature** *Exposure and Health* (KC et al., 2020a), © 2020.

4.1 Abstract

Biomass smoke exposure is associated with a heightened risk of development of respiratory diseases that include chronic obstructive pulmonary disease (COPD). The aim of this study was to increase our understanding of how biomass smoke could contribute to an increased susceptibility to respiratory infection. We investigated the effects of cow dung and wood smoke exposure on human bronchial epithelial cells with respect to adherence of a major respiratory bacterial pathogen in COPD, nontypeable *Haemophilus influenzae* (NTHi), using immunofluorescence microscopy. In addition, expression of a known receptor of NTHi, platelet activating factor receptor (PAFR), and two proinflammatory cytokines, interleukin-6 (IL-6) and interleukin-8 (IL-8), were determined using quantitative polymerase chain reaction. We observed a dose-dependent increase in NTHi adhesion to human bronchial epithelial cells following exposure to cow dung but not to wood smoke extracts. Pretreatment with PAFR antagonists, WEB-2086 and its analogue, C17, decreased adherence by NTHi to bronchial epithelial cells exposed to cow dung smoke. Both cow dung and wood smoke induced expression of PAFR, as well as of IL-6 and IL-8, which was inhibited by WEB-2086 and C17. In conclusion, biomass smoke from combustion of cow dung and wood induced expression of PAFR and airway inflammatory markers in human bronchial epithelial cells. Cow dung exposure, but not wood smoke exposure, mediated a measurable increase in NTHi adhesion to bronchial epithelial cells that was inhibited by PAFR antagonists. This work highlights the potential of PAFR as a therapeutic target for reducing the impact of hazardous biomass smoke exposure on respiratory health.

4.2 Introduction

Almost half of the world's population is reliant on biomass, such as wood, animal dung, and crop residues as their primary source of fuel for cooking and heating purposes (WHO, 2018c). Rural human populations of low- and middle- income countries, in particular women and young children, are constantly exposed to extremely high levels of smoke from wood/cow dung that is used for cooking. For instance, in a rural Costa Rican setting, Park and Lee (Park and Lee, 2003) monitored the quality of indoor air of the households with wood burning stoves for 24-h and reported elevated concentrations of $PM_{2.5}$ (44 µg/m^3), PM10 (132 µg/m^3), carbon monoxide (0.5 and 3.3 ppm). The concentration of these parameters considerably peaked during cooking, with $PM_{2.5}$ and PM_{10} levels reaching a concentration of approximately 310 to 8170 µg/m3 and 500 to 18900 µg/m3 respectively (Park and Lee, 2003).

Household biomass smoke exposure has been linked to an increased risk of airway disease including chronic obstructive pulmonary disease (COPD) (Kc et al., 2018a, Capistrano et al., 2017). An Indian study involving non-smoking women reported an increased prevalence of respiratory symptoms (chronic bronchitis and dyspnoea) in those who used biomass fuel for cooking (Behera and Jindal, 1991). Another recent Indian study reported that a substantial proportion (>60%) of biomass fuel (cow dung cakes and firewood) users are at higher risks of developing COPD, bronchial asthma and tuberculosis (Faizan and Thakur, 2019). Worldwide, COPD is responsible for 3 million deaths each year making it the third largest cause of human mortalities after heart disease and stroke (WHO, 2018d). While tobacco smoking is still the primary causative factor of COPD in high-income countries, in low and middle countries over 45% of COPD patients have been found to be never smokers (Kc et al., 2018a). There is emerging evidence that smoke from biomass fuels is an independent risk factor for the development of COPD where reliance on biomass fuel is high (Ehrlich et al., 2004, Magitta et

al., 2018). However, the exact mechanism how biomass smoke exposure contributes to the development and progression of COPD is still unknown. Since, biomass fuel smoke has been linked to acute respiratory infections (Po et al., 2011) we hypothesized that the exposure of bronchial epithelial cells to biomass smoke extracts increases adherence of nontypeable *Haemophilus influenzae* (NTHi), the major respiratory pathogen involved in the pathogenesis of COPD.

This study was conducted to advance our understanding of how biomass smoke could contribute to an increased susceptibility to respiratory infection, In this work we investigated the response of human bronchial epithelial cells to cow dung and wood smoke extract exposures, in particular, the expression of platelet activating factor receptor (PAFR) and pro-inflammatory cytokines, interleukin 6 (IL-6) and IL-8, and the level of adhesion of NTHi to smoke-extract treated cells. . We demonstrated that exposure of bronchial epithelial cells to cow dung smoke extract results in the increased adhesion of NTHi. We further established an association between PAFR expression and the adhesion of NTHi to the bronchial epithelial cells. Interestingly the same increased adherence of NTHi was not observed in response to wood smoke extract.

4.3 Methods

4.3.1 Generation of cigarette, cow dung and wood smoke extracts

In our previous paper, we described in detail the cost-effective and reproducible method for the generation of different smoke extracts, including cigarette, cow dung and wood smoke extracts (CSE, CDSE and WSE, respectively) (KC et al., 2018b). Briefly, three cigarette rolls, four rolls of cow dung powder, and four rolls of wood shavings were smoked using a water

aspirator. The cigarette, cow dung and wood smoke extracts were prepared by vortexing cotton filter containing the trapped smoke particles in 1 mL dimethyl sulfoxide (DMSO). The smoke extracts were then filter-sterilized through a 0.22 µm membrane filter and quantified by weight measurement. Finally, the prepared extracts were normalized to the same concentration (8.75 mg/mL) and stored at -20^{o}C. The extracts were diluted in bronchial epithelial cell growth medium (BEGM) for use in exposure experiments.

4.3.2 *In vitro* BEAS-2B cell culture and fluorescence labelling of bacteria

Airway epithelial cells are the primary cells to respond to environmental insults to the respiratory system, such as smoke and dust particles (Crystal et al., 2008). Undoubtedly, these cells are of considerable interest to investigate the biological and/or pathological effects of environmental air pollutants/pathogens (Gruenert et al., 1995). The *in vitro* system comprising airway epithelial cells, including bronchial and alveolar epithelial cells are routinely utilised in both basic and clinical research settings, with the advantage to exposing these cells directly to environmental stimuli. Bronchial epithelial cells are the primary sites in the respiratory tract to interact and respond to pathogenic microbes and environmental substances. Primary human bronchial epithelial cells are mortal, show high degree of variability and are difficult to isolate, whereas, commercially available bronchial epithelial cells, such as BEAS-2B cells, are immortal, much more homogenous under normal conditions and are easily available. Thus, we utilized BEAS-2B cells to evaluate the effect of smoke extract exposures. The BEAS-2B cells (Catalogue no 95102433, Sigma-Aldrich) were cultured in serum free bronchial epithelial basal medium (BEBM) supplemented with the BulletKit (BEGM) (Lonza, Basel, Switzerland) at 37^{o}C and 5% CO_2. The BEAS-2B cells were sub-cultured in a T-75 flask (Corning, NY, USA), and were used for experiments at ≤ 15 passages. The NTHi strain RHH3 was collected as part

of routine hospital laboratory diagnosis at the Royal Hobart Hospital, Hobart, Tasmania, Australia and no research participants or patients were recruited for this purpose. The study was approved by the Tasmanian Health and Medical Human Research Ethics Committee (EC00337) and conducted in accordance with the Declaration of Helsinki. The bacteria were labelled with fluorescein isothiocyanate (FITC) (Sigma-Aldrich, USA) as previously described (Geelen et al., 1993). Briefly, bacteria were grown overnight in brain heart infusion medium supplemented with hemin (1 mg/mL) and nicotinamide adenine dinucleotide (10 mg/mL). Subsequently, bacteria (10^8 CFU/mL) were incubated for 1 h at 4^{o}C with 1 mg/mL (w/v) FITC prepared in a carbonate buffer (0.09M sodium carbonate, 0.015 M sodium bicarbonate and 0.1 M sodium chloride). The FITC-labelled bacteria were washed three times with phosphate buffered saline (PBS) and resuspended in BEGM without antibiotics at a density of 5×10^7 CFU/mL.

4.3.3 Smoke extract exposure

For adherence assays, the BEAS-2B cells were transferred to sterile 8-well chambered glass slides (Millipore, MA, USA) pre-coated with 5% (v/v) bovine collagen I (Thermo Fisher Scientific, USA) at a density of 30,000 cells per well and cultured for an additional 24 h to form a confluent monolayer. The cell monolayer was then exposed to CSE, CDSE and WSE at concentrations ranging from 87.5 ng/mL to 87.5 µg/mL for 4 h at 37^{o}C and 5% CO_2 as described in our previous paper (KC et al., 2018b).

4.3.4 Pre-treatment with PAFR antagonists

After exposure to smoke extracts, the BEAS-2B cells were washed with pre-warmed PBS and treated with PAFR antagonists, WEB-2086 (Sigma-Aldrich, USA) and a new analogue C17 that was synthesized at the University of Tasmania whereby the fused triazolo-diazepine ring and oxopropylmorpholine side chain group were replaced by a pyrrole ring and a dodecyl group, respectively (Figure 4.1). A 10 mM stock solution was prepared in DMSO (Sigma, USA) and diluted in antibiotic-free BEGM to final concentrations of 100 µM, 10 µM, 1 µM and 100 nM. The BEAS-2B cells were pre-incubated with 200 µL of 100 µM, 10 µM, 1 µM and 100 nM concentrations of WEB-2086 or C17 for 1 h at 37^{o}C and 5% CO_2.

4.3.5 Bacterial adhesion assays

Following pre-treatment of the BEAS-2B cell monolayer with PAFR antagonists, 100 µL of FITC-tagged NTHi was added to the monolayer at a multiplicity of infection of 100:1. We conducted the smoke extract exposure and bacterial adhesion experiments at 37^{o}C mimicking the *in vivo* body temperature. NTHi being an obligate human pathogen exhibits optimal infectivity at 37^{o}C, therefore, we conducted this experiment only at 37^{o}C. The cells were incubated for 1 h at 37^{o}C and 5% CO_2 and subsequently washed three times with pre-warmed PBS. BEAS-2B cells were fixed with 200 µL of 4% (w/v) paraformaldehyde (Sigma-Aldrich, USA) for 20 minutes at room temperature. After 2 washes with PBS, cells were permeabilized with 100 µL of chilled (-20^{o}C) acetone for 10 minutes at room temperature. Subsequently, the cells were washed three times with PBS and blocked with 200 µL of 1% (w/v) bovine serum albumin (BSA) (Sigma-Aldrich, USA), prepared in PBS containing 1% (v/v) Tween-20 (Sigma-Aldrich, USA) for 1 h at room temperature. The cells were then incubated with mouse

anti-human platelet activating factor receptor monoclonal antibody (anti-PAFR mAb, 2.5 µg/mL) (11A4, Clone 21, Cayman Chemical Company, USA) overnight in the dark at 4ºC. After 3 washes with 0.1% (w/v) BSA, cells were incubated with a 1:100 dilution of Alexa Fluor 594-conjugated goat anti-mouse IgG (H+L) secondary antibody (Thermo Fisher Scientific, USA) for 1 h at room temperature. The cells were then rinsed three times with 0.1% (w/v) BSA, followed by staining with 4′,6-diamidino-2-phenylindole (DAPI, 1 µg/mL) (Thermo Fisher Scientific, USA) in the dark at room temperature for 15 min. Finally, the cells were washed three times with PBS and air-dried before the slides were mounted with Dako fluorescence mounting media (Agilent, USA).

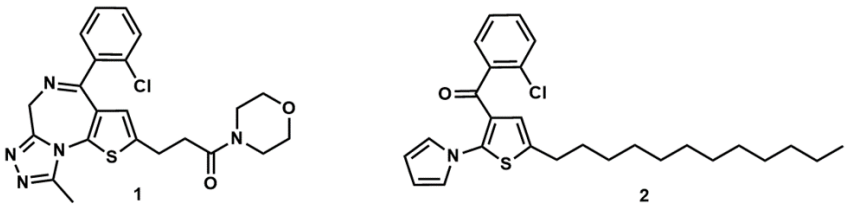


Figure 4.1 Chemical structure of WEB-2086 (1) and its analogue C17 (2).

WEB-2086 is a PAFR antagonist and belongs to the group of thienotriazolo-1,4-diazepines (hetrazepines). C17 is an analogue of WEB-2086 in which the fused triazolo-diazepine ring and the oxopropylmorpholine side chain group are replaced by a pyrrole ring and a dodecyl group, respectively.

4.3.6 Microscopy and image analysis

The cells were examined with an Olympus BX50 epifluorescence microscope equipped with NIS elements software (Nikon, Japan) and a Cool Snap Hq2 CCD camera (Photometrics, USA) under 400× magnification. Ten images were taken per well from different points using three different fluorescence channels for DAPI (violet excitation and blue emission, 200 ms

exposure), Alexa Fluor 594 (green excitation and red emission, 300 ms exposure) and FITC (blue excitation and green emission, 500 ms exposure). Image merging and bacterial and BEAS-2B cell counting were performed using ImageJ (NIH, USA).

4.3.7 Quantitative real-time polymerase chain reaction (qRT-PCR) analysis

To determine the transcriptional level of expression of PAFR, IL-6 and IL-8, the BEAS-2B cells were seeded into sterile clear flat-bottom 12-well plates (Corning, USA), not coated with collagen, at a density of 2×10^5 cells per well and incubated overnight at 37ºC and 5% CO_2. The cells were then exposed to smoke extracts CSE, CDSE or WSE for 4 h at 37ºC and 5% CO_2. After 4 h, total RNA was extracted using Tri-reagent (Sigma-Aldrich, USA), according to the manufacturer's instructions. Any traces of genomic DNA were degraded by treating the total RNA extract with DNase (Promega, USA). This was followed by first-stranded cDNA synthesis using a SensiFAST cDNA synthesis kit (Bioline, UK). Subsequently, the cDNA generated was amplified on a LightCycler 480 System (Roche, Switzerland) using SensiFAST Probe No-ROX kit (Bioline, UK). Finally, mRNA expression was normalized to the expression of a housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and the fold-changes in PAFR, IL-6 and IL-8 mRNA levels in smoke extract exposed cells were expressed relative to mRNA levels in mock (1% DMSO) treated cells using the primer pairs: GAPDH_F, GCCAAGGTCATCCATGACAACT; GAPDH_R, GGGCCATCCA-CAGTCTTCTG; PAFR_F, GACAGCATAGAGGCTGAGGC; PAFR_R, TAGCCATT-AGCAATGACCCC; IL-6_F, ACAGCCACTCACCTCTTCAG; IL-6_R, CCATCTTT- TTCAGCCATCTTT; IL-8_F, CTTGGCAGCCTTCCTGATTT; IL-8_R, TTCTTTAG- CACTCCTTGGCAAAA, respectively.

4.3.8 Cytotoxic effect of WEB-2086 and C17 on BEAS-2B cells

To ascertain whether the compounds WEB-2086 and C17 exhibited toxicity towards bronchial epithelial cells, BEAS-2B cells were seeded into sterile clear flat-bottom 96 well plates (Sigma-Aldrich, USA), not coated with collagen, at a density of 5,000 cells per well and incubated overnight at 37^{o}C and 5% CO_2. The cells were then treated in triplicate with different concentrations of WEB-2086 or C17 ranging from 100 nM to 100 µM, followed by the addition of Alamar Blue (Life Technologies, USA) to each well at a final concentration of 10% (v/v). The cells were then incubated for 4 h at 37^{o}C and 5% CO_2 to allow reduction of resazurin (blue) to resorufin (red). The absorbance signal was then measured at 570 nm with 600 nm as a reference wavelength after 4 and 24 h of treatment with WEB-2086 or C17, using a Spectromax Spectrophotometer Microplate Reader (Molecular Devices, USA). The percent reduction of Alamar Blue was calculated as previously described (KC et al., 2018b).

4.3.9 Statistical analysis

Two independent experiments, each including two or three replicates, were analyzed. Data from four or six biological replicates from two experiments were included in the analysis and were expressed as the mean ± standard error. Differences between groups were tested by one-way or two-way analysis of variance (ANOVA) with Dunnett's or Bonferroni post-hoc analysis using GraphPad Prism version 5.0 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

4.4 Results

4.4.1 CDSE but not WSE significantly upregulated NTHi adherence to human bronchial epithelial cells

We investigated whether exposure to cow dung and wood smoke extracts increases the susceptibility of bronchial epithelial cells to NTHi adhesion. For this BEAS-2B cells were exposed to CDSE and WSE at concentrations ranging from 87.5 ng/mL to 87.5 µg/mL for 4 h. Parallel exposures of BEAS-2B cells to CSE were performed for comparison. Treatment of BEAS-2B cells with CSE, CDSE and WSE correlated with increased PAFR expression as measured by immunofluorescence (KC et al., 2018b) Challenge of BEAS-2B with FITC-labelled NTHi for 1 h revealed that bacterial adhesion to unstimulated BEAS-2B cells was low, *i.e.* 1.859 ± 0.023 bacteria per cell. The number of bacteria adhering to BEAS-2B cells exposed to CSE and CDSE increased in a dose-dependent manner. NTHi adherence to CSE- and CDSE-exposed BEAS-2B cells was 2.0 ($p < 0.001$), and 1.6 ($p < 0.01$) fold higher, respectively, with respect to 1% DMSO (mock) treated BEAS-2B cells, at the 87.5 µg/mL concentration (Figure 4.2). At lower concentrations of CSE and CDSE, *i.e.* 875 ng/mL and 87.5 ng/mL, NTHi adhesion did not differ significantly between mock treated and smoke-extract exposed BEAS-2B cells. Regression analyses were performed on levels of PAFR expression based on PAFR fluorescence intensity with respect to NTHi adhesion. The adherence of NTHi to CSE and CDSE exposed BEAS-2B cells was positively correlated with levels of PAFR expression ($r = 0.96$ $p = 0.04$ and $r = 0.979$ $p = 0.021$, respectively) (Figure 4.3). WSE exposure did not result in increased attachment of NTHi to BEAS-2B cells across the concentration range tested, *i.e.* 87.5 ng/mL to 87.5 µg/mL. Compared to 87.5 µg/mL WSE exposure, NTHi adhesion to bronchial epithelial cells was 80% ($p < 0.001$) and 42% ($p < 0.05$) higher in CSE and CDSE exposures, respectively.

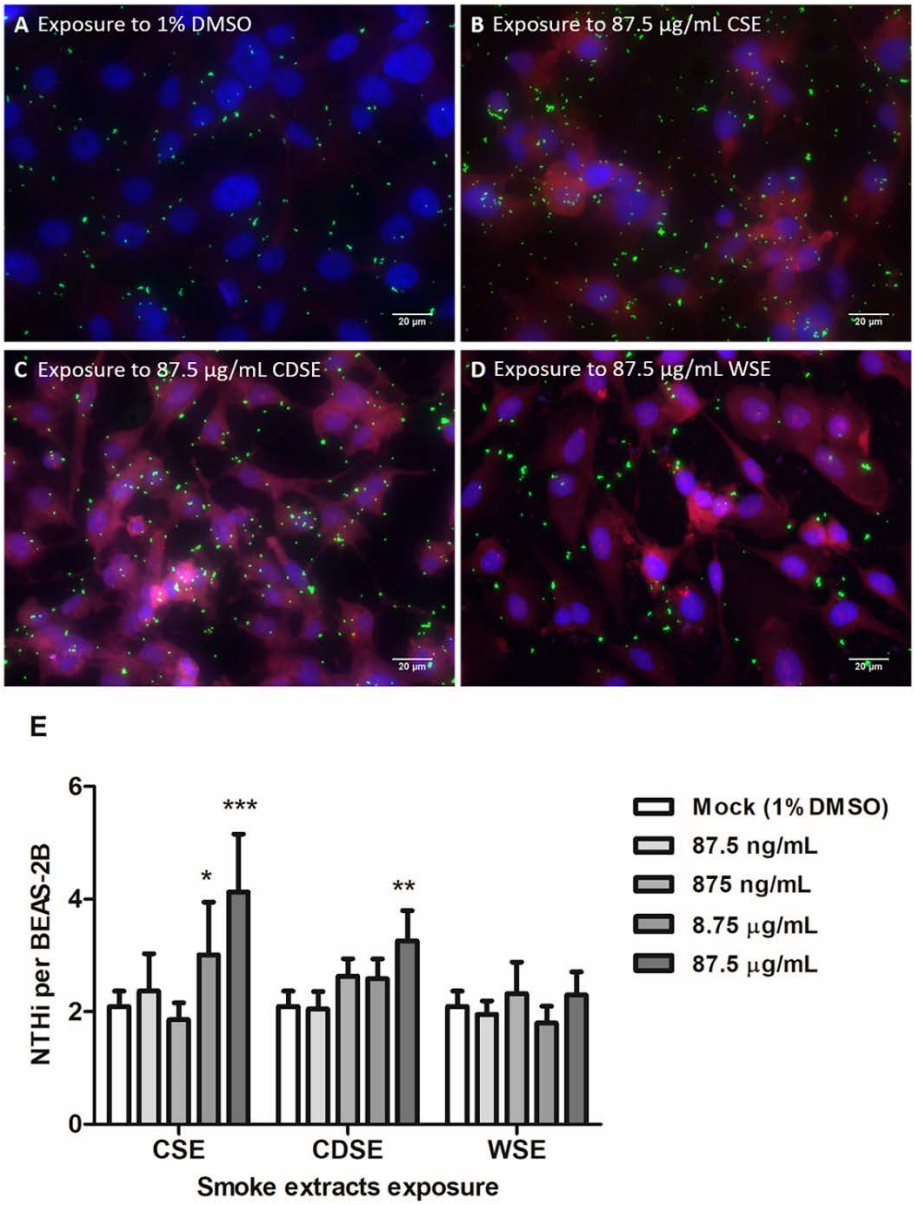


Figure 4.2 Effect of biomass smoke exposure on the adhesion of RHH3 strain of nontypeable *H. influenzae* to human bronchial epithelial cells.

(A) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (B) BEAS-2B cells exposed to 87.5 µg/mL CSE for comparison. (C) BEAS-2B cells exposed to 87.5 µg/mL CDSE. (D) BEAS-2B cells exposed to 87.5 µg/mL WSE. All immunofluorescence micrographs show NTHi labelled with FITC (1mg/mL, green) and BEAS-2B cells stained for PAFR expression (anti-PAFR monoclonal antibody; 2.5 µg/mL, red), and nuclei counterstained with DAPI (1 µg/mL, blue). Magnification = 400×. (E) Exposure of BEAS-2B cells to different concentrations of CSE, CDSE and WSE ranging from 87.5 ng/mL to 87.5 µg/ml. NTHi

adherence was significantly increased in CSE (8.75 and 87.5 µg/ml) and CDSE (87.5 µg/ml) exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; *$p < 0.05$, **$p < 0.001$, ***$p < 0.001$, relative to mock treated cells, one-way ANOVA with Dunnett's multiple comparison test. CSE, cigarette smoke extract; CDSE, cow dung smoke extract; WSE, wood smoke extract; NTHi, nontypeable *Haemophilus influenzae*; PAFR, platelet activating factor receptor; DMSO, dimethyl sulfoxide; FITC, fluorescein isothiocyanate; DAPI, 4′, 6-diamidino-2-phenylindole.

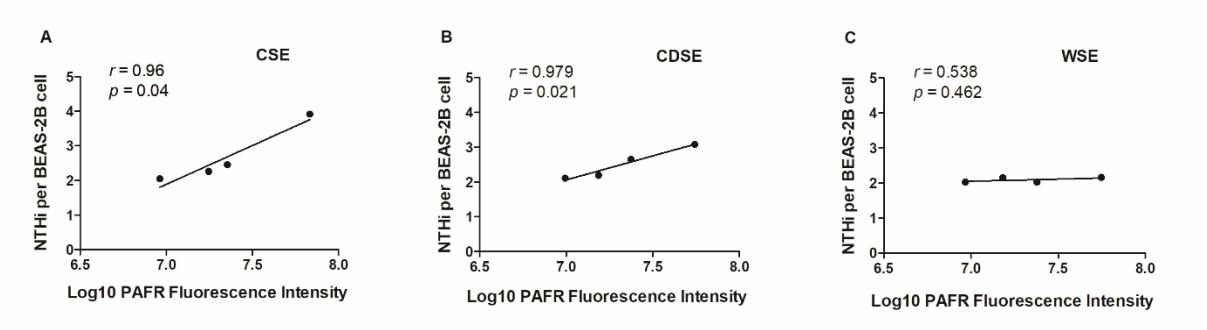


Figure 4.3 Correlation between PAFR expression and NTHi adhesion on BEAS-2B cells exposed to (A) CSE (B) CDSE (C) WSE.

PAFR expression corresponds to log10 of fluorescence intensity, which is significantly positively correlated with the number of NTHi attached per BEAS-2B cells for CSE and CDSE exposures. For WSE exposure, NTHi adhesion is not associated with expression of PAFR. Data are presented from four independent replicates from two experiments. CSE, cigarette smoke extract; CDSE, cow dung smoke extract; WSE, wood smoke extract; NTHi, nontypeable *Haemophilus influenzae*; PAFR, platelet activating factor receptor.

4.4.2 WEB-2086 and C17 significantly reduced the adhesion of NTHi to cow dung smoke exposed BEAS-2B cells

To determine whether PAFR played a role in adhesion by NTHi to cow dung smoke exposed bronchial epithelial cells, BEAS-2B cells were treated with a well-known PAFR antagonist, WEB-2086 at concentrations ranging from 100 nM to 100 µM for 1 h. We observed significant

decrease in CDSE-induced PAFR expression upon treatment with WEB-2086 as shown in Figure 4.4. The treated BEAS-2B cells were then challenged with FITC-tagged NTHi. The adherence of NTHi to both CSE- and CDSE- exposed BEAS-2B cells was reduced in the presence of WEB-2086 in a dose-dependent manner (Figure 4.5). Compared to mock stimulated cells, NTHi adhesion to WSE-exposed BEAS-2B cells was not affected by treatment with WEB-2086 ($p > 0.05$) (Figure 4.5). C17, an analogue of WEB-2086 was also tested for its activity towards the expression of PAFR on and the adhesion of NTHi to CSE, CDSE and WSE exposed bronchial epithelial cells. Consistent with WEB-2086 results, C17 exhibited significant inhibition of CSE-, CDSE- and WSE-induced expression of PAFR on the bronchial epithelial cells (Figure 4.4). Notably, C17 exhibited significant abrogation of attachment of NTHi to both CSE- and CDSE-treated bronchial epithelial cells in a dose dependent manner (Figure 4.6).

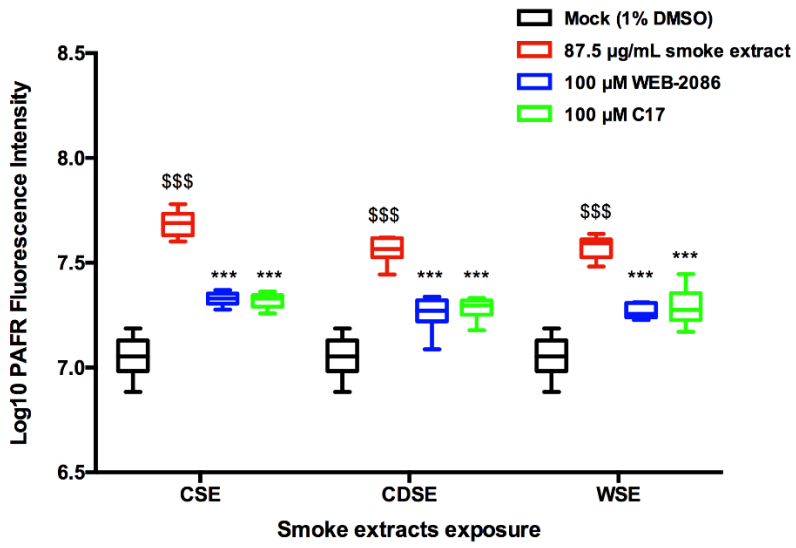


Figure 4.4 Smoke extracts (CSE, CDSE and WSE) exposure induced PAFR expression on bronchial epithelial cells and their inhibition by PAFR antagonists WEB-2086 and C17.

PAFR expression corresponds to log10 of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb in immunofluorescence assay. Data are representative of two independent experiments. $\$\$\$p < 0.001$ relative to 1% DMSO

treated control cells, ***$p < 0.001$ relative to vehicle-treated smoke extracts-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. CSE, cigarette smoke extract; CDSE, cow dung smoke extract; WSE, wood smoke extract; PAFR, platelet activating factor receptor; DMSO, dimethyl sulfoxide.

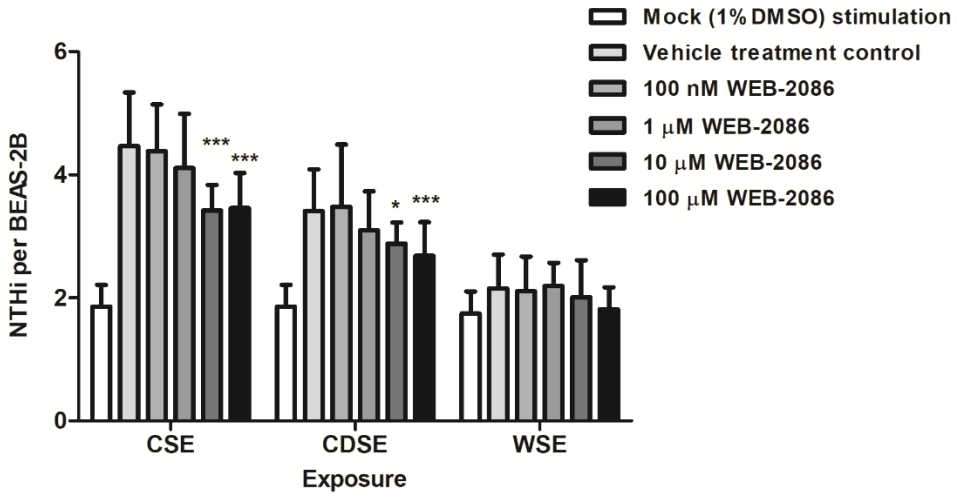


Figure 4.5 Effects of PAFR antagonist, WEB-2086, on attachment of NTHi to CSE and CDSE exposed BEAS-2B cells.

The BEAS-2B cells were exposed to 1% DMSO or 87.5 µg/mL of CSE, CDSE, or WSE for 4 h. This was followed by treatment with vehicle (1% DMSO) or PAFR antagonist, WEB-2086, at a concentration range of 100 nM to 100 µM for 1 h and subsequent challenge with FITC-labelled NTHi for an additional hour. WEB-2086 exhibited dose dependent inhibition of NTHi adhesion to the CSE and CDSE exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; *$p < 0.05$, ***$p < 0.001$, relative to vehicle-treated CSE-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. CSE, cigarette smoke extract; CDSE, cow dung smoke extract; WSE, wood smoke extract; NTHi, nontypeable *Haemophilus influenzae*; PAFR, platelet activating factor receptor; DMSO, dimethyl sulfoxide; FITC, fluorescein isothiocyanate.

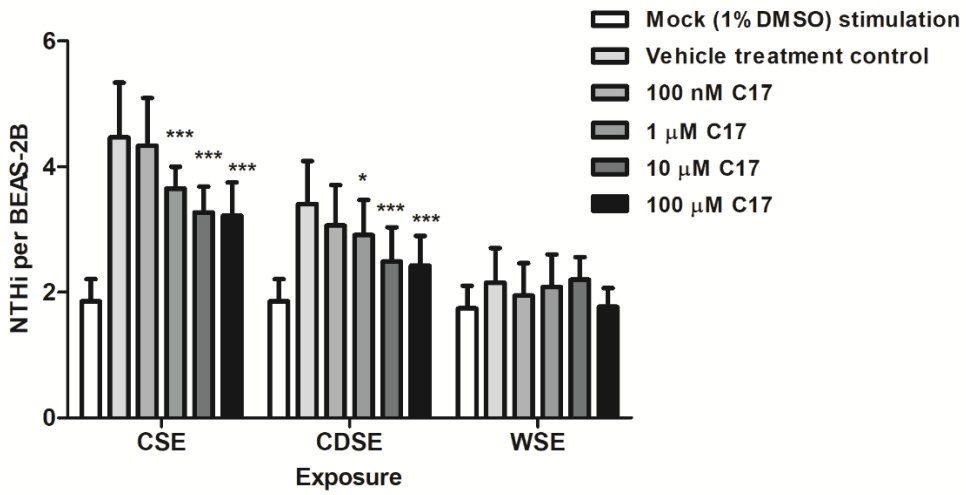


Figure 4.6 Effect of new compound, C17, on attachment of NTHi to CSE and CDSE exposed BEAS-2B cells.

The BEAS-2B cells were exposed to 1% DMSO or 87.5 µg/mL of CSE, CDSE, or WSE for 4 h. This was followed by treatment with vehicle (1% DMSO) or compound C17 at a concentration range of 100 nM to 100 µM for 1 h and subsequent challenge with FITC-labelled NTHi for an additional hour. C17 exhibited dose dependent inhibition of NTHi adhesion to the CSE and CDSE exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; *p < 0.05, ***p < 0.001, relative to vehicle-treated CSE-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. CSE, cigarette smoke extract; NTHi, nontypeable Haemophilus influenzae; PAFR, platelet activating factor receptor; DMSO, dimethyl sulfoxide; FITC, fluorescein isothiocyanate.

4.4.3 Transcriptional response of BEAS-2B cells following CSE, CDSE, and WSE exposure

The effect of CSE-, CDSE-, and WSE-exposure on BEAS-2B cells expression of a known receptor of NTHi, PAFR, was investigated at the transcriptional level. PAFR mRNA levels in 87.5 µg/mL CSE-, CDSE-, and WSE-exposed BEAS-2B cells increased by 9.0 ($p < 0.001$), 7.2 ($p < 0.01$) and 9.5-fold ($p < 0.001$), respectively, compared to mock (1% DMSO) treated

BEAS-2B cells (Figure 4.7). mRNA transcripts for the pro-inflammatory cytokines, interleukin-6 (IL-6) and interleukin-8 (IL-8), were significantly induced in BEAS-2B cells by 87.5 µg/mL CSE, CDSE, and WSE as illustrated in Figure 4.7. In comparison to the mock (1% DMSO) treated control BEAS-2B cells, the mean IL-6 and IL-8 levels of expression were approximately 8.4 ($p < 0.001$) and 11.2 ($p < 0.001$) times higher in CSE-treated cells, respectively; 7.2 ($p < 0.001$) and 11.9 ($p < 0.001$) times higher in CDSE stimulated cells, respectively; and 6.6 ($p < 0.001$) and 12.2 ($p < 0.001$) times higher in WSE exposed cells, respectively.

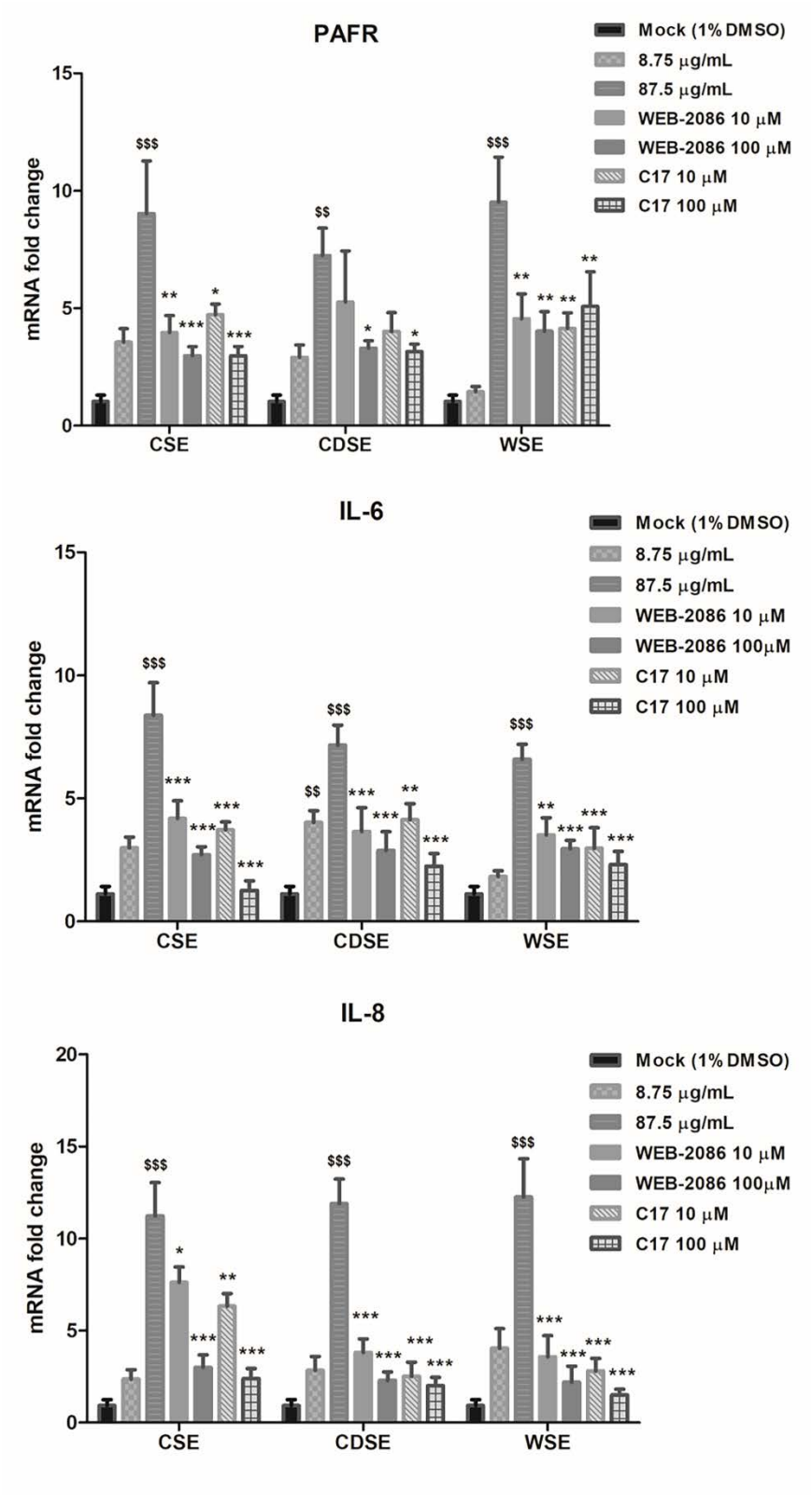


Figure 4.7 Relative fold change in PAFR, IL-8, and IL-6 mRNA normalized with GAPDH.

BEAS-2B cells exposed to CSE, CDSE, and WSE for 4 h at concentrations 8.75 and 87.5 µg/mL following treatment with PAFR antagonists, WEB-2086 and C17 at concentrations 10 µM and 100 µM for 1 h. The data are presented as the mean of the observed fold-change ± standard error for four independent replicates from two experiments; $^{\$\$}p < 0.01$, $^{\$\$\$}p < 0.001$, relative to mock (1% DMSO) treated control; $*p < 0.05$, $** p < 0.01$, $*** p < 0.001$, relative to the 87.5 µg/mL smoke extract exposed BEAS-2B cells; two-way ANOVA with Bonferroni post-hoc test. CSE, cigarette smoke extract; CDSE, cow dung smoke extract; WSE, wood smoke extract; NTHi, nontypeable *Haemophilus influenzae*; PAFR, platelet activating factor receptor; IL-6, interleukin 6; IL-8, interleukin 8; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

4.4.4 Effects of the WEB-2086 and the C17 treatment on induced transcriptional expression of PAFR and inflammatory cytokines IL-6 and IL-8

The effects of WEB-2086 and C17 on the expression of PAFR, IL-6 and IL-8 in CSE-, CDSE-, and WSE-treated BEAS-2B cells were also examined. Treatment of smoke extract-exposed BEAS-2B cells with WEB-2086 or C17 for 1 h significantly downregulated the expression of PAFR mRNA (Figure 4.7). WEB-2086 and C17 at a concentration of 100 µM reduced the levels of PAFR mRNA by 54% ($p < 0.05$) and 56% ($p < 0.05$), respectively. Compared to CSE-exposed BEAS-2B cells (87.5 µg/mL) that were treated with vehicle (1% DMSO), transcriptional levels of IL-6 and IL-8 were significantly inhibited by 68% ($p < 0.001$) and 73% ($p < 0.001$), respectively, with 100 µM WEB-2086 treatment, and by 85% ($p < 0.001$) and 79% ($p < 0.001$), respectively, with 100 µM C17 treatment. In addition, both WEB-2086 and C17 treatment downregulated the CDSE-induced expression of IL-6 in a dose dependent manner *i.e.* for 10 µM versus 100 µM of WEB-2086 and C17, 49% ($p < 0.001$) versus 60% ($p < 0.001$) and 43% ($p < 0.01$) versus 69% ($p < 0.001$), respectively. Similarly, the levels of IL-8 mRNA were also reduced in a dose dependent manner by both WEB-2086 and C17 *i.e.* 68% ($p < 0.001$) versus 81% ($p < 0.001$) and 79% ($p < 0.001$) versus 83% ($p < 0.001$), for 10 µM versus

100 µM of WEB-2086 and C17 treatments, respectively (Figure 4.7). Furthermore, the compounds were also found to suppress the expression of IL-6 and IL-8 in WSE-treated cells.

4.4.5 Tolerance of compounds WEB-2086 and C17 by human bronchial epithelial cells

WEB-2086 and C17 were tested for toxicity towards BEAS-2B cells across the concentration range of 100 nM to 100 µM using the Alamar Blue assay. No significant signs of cytotoxicity were seen after 24 h treatments. The reduction of Alamar Blue by BEAS-2B cells was similar for treatment with both vehicle (1% DMSO) and 100 µM WEB-2086, *i.e.* 49.2% versus 44.8% at 4 h ($p > 0.05$), and 102.3% versus 104.1% at 24 h ($p > 0.05$). Furthermore, the viability of the bronchial epithelial cells was similar across treatment with vehicle and 100 µM C17, *i.e.* 49.2% versus 45.9% at 4 h ($p > 0.05$), and 102.3% versus 102.5% at 24 h ($p > 0.05$). Thus, both compounds, WEB-2086 and C17, were found to be tolerated by BEAS-2B cells up to the highest concentration tested of 100 µM.

4.5 Discussion

Nearly 4 million deaths annually have been attributed to household air pollution (WHO, 2018c). A number of meta-analyses of epidemiological studies have reported an overall 1.4 to 2.8-fold increased risk of COPD development in solid biomass fuel users compared to clean fuel users (Hu et al., 2010, Yang et al., 2017, Kurmi et al., 2010, Sana et al., 2018, Po et al., 2011). A recent PUMA (Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America) study conducted in primary care settings in four Latin American countries reported a higher proportion of subjects with COPD were exposed to biomass smoke compared to those without

COPD (Montes de Oca et al., 2017). Da Silva and colleagues examined the dose-response relationship and reported a significant correlation between reduced pulmonary function and the level and duration of biomass exposure (da Silva et al., 2012). Several short-term trials using improved cooking stoves and kitchen ventilation have provided support for the efficacy of interventions, suggesting an association between biomass smoke exposure and lung function impairment (Zhou et al., 2014, Romieu et al., 2009, Pope et al., 2015).

Smith et al., in a population-based study of ~317,000 nonsmoking Chinese adults, reported airflow obstruction to be positively associated with cooking with coal but not with other sources of household air pollution (Smith et al., 2014). Recently, two large-scale multinational studies were conducted to address the role of household air pollution in airflow obstruction and were published in the same issue of the *American Journal of Respiratory and Critical Care Medicine* in 2018 (Amaral et al., 2018, Siddharthan et al., 2018). Amaral and colleagues, using data from the BOLD (Burden of Obstructive Lung Disease) study, reported no evidence of association between airflow obstruction and the use of solid fuels (Amaral et al., 2018). However, they reported a 28% higher risk of chronic phlegm in never-smoker women who had used solid fuels for cooking or heating over twenty years or more (Amaral et al., 2018). Siddharthan and co-investigators analyzed population data from the PRISA (Pulmonary Risk in South America) and CRONICAS cohort studies in South America, a longitudinal study conducted by the Centre for Control of Chronic Diseases in Bangladesh, and the LINK (Lung Function Study in Nakaseke and Uganda) and FRESH AIR studies in Africa. They reported that study participants with household air pollution exposure from biomass fuel use were 41% more likely to have COPD than non-exposure individuals and that the association between household air pollution exposure and COPD was more pronounced in women (adjusted odds ratio 1.70; 95% confidence interval, 1.24 - 2.32) than in men (1.21; 0.92 - 1.58) (Siddharthan et al., 2018). The findings from the above two studies are therefore not in agreement and it has

been acknowledged that the lack of direct measurement of exposure and level of ventilation may have impacted the studies (Balmes and Eisen, 2018). The role of biomass smoke in airway obstructive disease requires further extensive research including examining the effect of biomass smoke at a cellular level, which was the focus of our study.

In this work, we report for the first time that exposure of bronchial epithelial cells to biomass smoke in the form of cow dung smoke exposure results in increased adhesion by NTHi, a major respiratory bacterial pathogen in the chronic respiratory disease COPD. NTHi are opportunistic bacterial pathogens that are normal commensals of the upper respiratory tract (Bakaletz and Novotny, 2018). However, in COPD patients, due to impaired immunity they colonize the lower airways of COPD patients, thereby inducing considerable inflammation, resulting in a worsening of symptoms and an increased frequency of exacerbations (Leung et al., 2017). More than half of acute COPD exacerbations are caused by bacterial infections, particularly NTHi (Sethi et al., 2002). This study suggests that cow dung smoke exposure could translate into an increased susceptibility to NTHi infection in the respiratory tract of patients with biomass smoke associated COPD.

We further investigated the effect of CDSE and WSE on bronchial epithelial cells. As shown in Figure 4.7, expression of both IL-6 and IL-8 was upregulated several-fold in bronchial epithelial cells following exposure to CDSE and WSE with respect to controls. IL-6 is a major stimulator of the production of acute phase proteins (Castell et al., 1989) and mediates upregulation of the Th17/Treg balance disrupting immunological tolerance and thus, contributes to the development of autoimmune and chronic inflammatory diseases (Kimura and Kishimoto, 2010). IL-8, on the other hand, is a key mediator in neutrophil-mediated acute inflammation via neutrophil transmigration and activation, including degranulation and the release of leukotrienes and platelet activating factor (Mukaida, 2003, Schröder, 1989, Mukaida,

2000). A significant upregulation of expression of pro-inflammatory cytokines, IL-6 and IL-8 upon exposure to biomass smoke play an important role in the pathogenesis of chronic inflammatory diseases, particularly in COPD (Tamimi et al., 2012, Gabay, 2006, Oishi et al., 1994). The results are in agreement with other human and animal model studies. Increased sputum and serum levels of IL-6 and IL-8 have been observed in women cooking with biomass fuels compared to the clean fuel users (Dutta et al., 2012, Dutta et al., 2013). Similar upregulation of cytokines was also observed in a mouse model of biomass exposure and in the *in vitro* exposure of biomass smoke to primary human small airway epithelial cells (Sussan et al., 2014, McCarthy et al., 2016a). This may in part be linked to the potential role of biomass smoke exposure in airway obstructive disease in never smokers, particularly, in low- and middle-income countries where biomass fuel is the primary source of domestic energy (Kc et al., 2018a).

Similar to CDSE, WSE exposure also increases the expression of PAFR (Figure 4.4 & 4.7). However, WSE exposure did not increase the attachment of NTHi to respiratory epithelial cells. There are two possible reasons for a lack of a measured effect of WSE on NTHi adhesion to bronchial epithelial cells. Firstly, the possible presence of compounds in WSE that competitively inhibit the binding of NTHi to PAFR. Secondly, the utilization of other potential epithelial receptors besides PAFR in CDSE-induced NTHi adhesion (Fink et al., 2002, Novotny and Bakaletz, 2016, Avadhanula et al., 2006, Klaile et al., 2013). PAFR has been reported to be the major receptor for NTHi binding via lipooligosaccharide glycoforms containing phosphorylcholine (Shukla et al., 2015, Shukla et al., 2016a, Swords et al., 2000). The temporal upregulation of PAFR expression in airway cells upon exposure to biomass smoke, air pollutants as well as cigarette smoke (KC et al., 2018b, Shukla et al., 2016b, Miyashita et al., 2017) appears to increase the susceptibility to NTHi infection (Shukla et al., 2015). Besides PAFR, NTHi has been reported to utilize the host extracellular matrix (ECM)

proteins; fibronectin, laminin and collagen IV for colonization in the respiratory tract via the Hap autotransporter (Fink et al., 2002). In addition, the NTHi Type IV pilus and the outer membrane protein P1 interacts with the intercellular adhesion molecule-1 (ICAM-1) and the carcinoembryonic antigen-related cell adhesion molecules (CEACAMs), respectively to adhere to respiratory tract epithelial cells (Novotny and Bakaletz, 2016, Avadhanula et al., 2006, Klaile et al., 2013). Thus, further work is needed to investigate the effect of CDSE and WSE on the expression of ECM proteins, ICAM-1 or CEACAMs, and how they relate to adherence of NTHi to the airway epithelium. While a number of studies have reported differential effects of wood and cigarette smoke exposure in terms of COPD phenotypes (Camp et al., 2014, Rivera et al., 2008, Fernandes et al., 2017), investigations are also needed to compare the effects of cow dung and cigarette smoke on the manifestation of COPD.

In terms of treatment of respiratory bacterial infections, the major challenges to antimicrobial therapy are the acquisition and spread of antibiotic resistance and the limited number of new antibiotics being developed and clinically approved for human use (Ventola, 2015). Therefore, a novel alternative approach is needed to prevent and treat bacterial airway infections through interfering with bacterial/host tissue interfaces. A number of PAFR antagonists, including WEB-2086, Ginkgolide-B, CV-3988, PCA-4248, and CAS-99103-16–9 have been reported to block the attachment of bacterial pathogens to respiratory epithelium (Grigg et al., 2012, Barbier et al., 2008, Shukla et al., 2016a, Hergott et al., 2015, Negro Alvarez et al., 1997). Here, WEB-2086 and its structural analogue, C17, were found to reduce the adhesion of NTHi to CSE- and CDSE-exposed human bronchial epithelial cells in a dose dependent manner. In addition, they both inhibited biomass and cigarette smoke-induced expression of pro-inflammatory cytokines IL-6 and IL-8, which are a hallmark feature of COPD.

A major limitation of this study is the use of a commercial airway epithelial cell line. Further work is now needed to evaluate the effects of biomass smoke exposure on PAFR expression and adherence of bacterial pathogens *in vivo*. Nevertheless, the results of the work reveal that biomass smoke exposure increases the adherence of a bacterial pathogen to bronchial epithelial cells. Furthermore, through the use of antagonists, we identified PAFR as a potential therapeutic target for alleviating the impact of biomass smoke exposure and reducing the risk of NTHi exacerbations. Translation of these findings would contribute to the development of new management strategies for biomass smoke-related obstructive airway disease.

4.6 Recommendation

This study has highlighted the need for global immediate action towards reducing the exposure to and impact of smoke generated from combustion of biomass fuel. National policies, and community awareness and education on preventive interventions, including the use of efficient cooking stoves and cleaner fuels such as biogas, liquefied petroleum gas and electricity are necessary to substantially reduce the number of deaths and illnesses from hazardous effects of biomass smoke exposure. Furthermore, therapeutic interventions, including the use of anti-oxidants, anti-inflammatory and anti-microbial drugs may be promising in limiting the effects of biomass exposure related respiratory diseases, including COPD. This will thus, significantly contribute to achieving the primary goal target of the Sustainable Development Goal 3; Ensure healthy lives and promote well-being for all at all ages.

5 Chapter 5: Whole-genome analyses reveal gene content differences between nontypeable *Haemophilus influenzae* isolates from chronic obstructive pulmonary disease compared to other clinical phenotypes

In the previous chapter (Chapter 4), we showed how temporal upregulation of PAFR during biomass exposure provides a gateway for NTHi adherence to human bronchial epithelial cells. In addition to PAFR mediated adherence, NTHi leverages a wide array of factors to survive and colonise in the COPD airways. It is still not clear what makes NTHi competent to establish infection in lower airways of COPD patients and how the COPD strains are different from NTHi isolates collected from other clinical diseases. This chapter aimed to elucidate genetic distinction between bacterial strains isolated from COPD and other clinical diseases.

This Chapter is presented as a Word document unchanged from the manuscript version submitted to *Microbial Genomics* for publication as

KC R, Leong KWC, Harkness NM, Gautam SS, Cooley LA, Lachowicz J, McEwan B, Petrovski S, Karupiah G, O'Toole RF. Whole-genome analyses reveal gene content differences between nontypeable *Haemophilus influenzae* isolates from chronic obstructive pulmonary disease compared to other clinical phenotypes. *Microbial Genomics*. 2020 Jul 24. doi: 10.1099/mgen.0.000405 (Kc et al., 2020b)

5.1 Abstract

Nontypeable *Haemophilus influenzae* (NTHi) colonises human upper respiratory airways and plays a key role in the course and pathogenesis of acute exacerbations of chronic obstructive pulmonary disease (COPD). Currently, it is not possible to distinguish COPD isolates of NTHi from other clinical isolates of NTHi using conventional genotyping methods. Here, we analysed the core and accessory genome of 568 NTHi isolates, including 40 newly sequenced isolates, to look for genetic distinctions between NTHi isolates from COPD with respect to other illnesses including otitis media, meningitis and pneumonia. Phylogenies based on polymorphic sites in the core genome did not infer discrimination between NTHi strains collected from different clinical phenotypes. On the other hand, pan genome wide association studies identified 79 unique NTHi accessory genes which were significantly associated with COPD. Furthermore, many of the COPD-related NTHi genes have known or predicted roles in virulence, transmembrane transport of metal ions and nutrients, cellular respiration and maintenance of redox homeostasis. This indicates that specific genes may be required by NTHi for its survival or virulence in the COPD lung. These results advance our understanding of the pathogenesis of NTHi infection in COPD lungs.

5.2 Impact Statement

Chronic obstructive pulmonary disease (COPD) is emerging as the third leading cause of human mortalities worldwide. Nontypeable *Haemophilus influenzae* (NTHi) is a major pathogen causing acute exacerbations resulting in diminished quality of life, hospitalisation, and increased risk of death in COPD patients. We analysed the core and accessory genome of 568 NTHi isolates, including 40 newly sequenced isolates, to genotypically distinguish between NHTi from COPD and other clinical phenotypes. This genome-wide analysis

identified accessory gene content differences between COPD and non-COPD strains. It highlighted a set of virulence and metabolic functions that may be differentially required by COPD strains. This knowledge is important for developing improvements in the management of NTHi infections that may lead to acute exacerbations in COPD patients.

5.3 Data Summary

1. Sequence read files for all 40 isolates sequenced in this work have been deposited in SRA and are accessible through NCBI BioSample and SRA accession numbers SAMN13942196 and SRS6166046, respectively.

2. All supporting data have been provided through supplementary data files. Five supplementary datasets (Dataset 1-5) and a multifasta file of non-redundant pan-gene sequences (pangene_sequence.fa) are available on Figshare (https://doi.org/10.6084/m9.figshare.12545957.v1).

5.4 Introduction

While nontypeable *Haemophilus influenzae* (NTHi) is a common commensal of the human nasopharynx, this bacteria is also associated with a spectrum of diseases including otitis media, sinusitis as well as hospital- and community-acquired pneumonia (Agrawal and Murphy, 2011). In addition, NTHi is the most common bacterial cause of COPD exacerbations (Moghaddam et al., 2011, Murphy et al., 2004, Sriram et al., 2018). NTHi has developed mechanisms to thrive in the hostile environment of different anatomical regions, such as the middle ear, upper and lower respiratory tracts, blood, and the meninges (Wong and Akerley, 2012). Evolutionary and ecological forces drive bacteria to adapt and grow in different niches (Messer et al., 2017, Pettigrew et al., 2018, Harrison et al., 2020, Elhenawy et al., 2019) by utilising the basic nutrients available and resisting toxic products present in its environment (Clementi and Murphy, 2011). This evolutionary adaptation typically involves two fundamental processes. The first is through mutations in genes, such as single nucleotide polymorphisms or nucleotide insertions/deletions, that can potentially alter the antigenicity of surface proteins or change the activity of enzymatic and transport proteins (LeClerc et al., 1996). A related mechanism is phase variation in which loci susceptible to hypermutation can undergo slipped-strand mispairing, due to changes in simple sequence repeats, which can rapidly modulate the expression levels of genes (Torres-Cruz and van der Woude, 2003, Pettigrew et al., 2018). The second process is the acquisition of entirely new genetic sequences via horizontal gene transfer, which can undergo homologous or non-homologous recombination into the recipient genome (Ochman et al., 2000). During homologous recombination, a chromosomal fragment of a genome is replaced with a homologous sequence from another genome whereas non-homologous recombination results in gain and loss of genetic material (Didelot et al., 2012). These processes can contribute to phenotypic changes including increased virulence, antibiotic resistance, and adaptations to the host

microenvironments such as immune evasion and greater metabolic capacity (Wilson et al., 2016, Leimbach et al., 2013, Barrick et al., 2009, Lieberman et al., 2011, Pettigrew et al., 2018, Didelot et al., 2016, Harrison et al., 2020).

Conventional typing methods such as multi-locus sequence typing (MLST) cannot distinguish between commensal and pathogenic strains of NTHi (LaCross et al., 2013, De Chiara et al., 2014, Erwin et al., 2008). Furthermore, a previous study by De Chiara and colleagues reported that phylogeny provides insufficient resolution to discriminate between strains isolated from different clinical sources based on an analysis of 97 NTHi isolates (De Chiara et al., 2014). Here, we hypothesised that NTHi associated with COPD may exhibit genetically encoded functional variances when compared to the isolates from non-COPD clinical illnesses. We therefore performed an analysis on a larger set of 568 NTHi genomes, which included 40 Australian isolates that were newly sequenced in this study. Our analyses involved the application of pan genome wide association studies (pan-GWAS) of genes to determine whether NTHi from COPD could be discriminated from isolates from other types of clinical disease.

5.5 Materials and methods

5.5.1 Bacterial strains collection, DNA extraction, and genome sequencing

Forty NTHi isolates were retrieved from sputum samples of patients admitted to the Royal Hobart Hospital (RHH) in Tasmania, Australia, from 2017 to 2018. Of these, thirteen isolates were collected from COPD patients with the remaining twenty-seven collected from patients with other disease presentations listed as non-COPD as shown in Dataset 1. DNA extraction and genome sequencing were performed using the protocol that has been described in detail in

our previous methods paper (Gautam et al., 2019). Briefly, for DNA preparation, pure cultures of NTHi isolates grown on chocolate agar were suspended in phosphate buffered saline. Chromosomal DNA was extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) and was treated with RNase for the removal of RNA from genomic DNA samples. RNase-treated DNA was further purified using the High Pure PCR Template Preparation Kit (Roche, Basel, Switzerland) as described previously (Gautam et al., 2019). The genomic DNA library of these isolates was then prepared using the Illumina Nextera XT Library Preparation Kit t (Illumina Inc., San Diego, CA, United States) and loaded into an Illumina MiSeq v2 (2 × 150-bp paired-end reads) cartridge for sequencing using an Illumina MiSeq platform at La Trobe University, Australia.

In addition to the 40 Australian NTHi isolates that were sequenced, 528 publicly-available NTHi genomes (Harrison et al., 2005, Pettigrew et al., 2018, Moleres et al., 2018, De Chiara et al., 2014, Kappler et al., 2017, Aziz et al., 2019, Atack et al., 2018, Kc et al., 2020c) were also downloaded from the National Centre for Biotechnology Information (NCBI) (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Haemophilus_influenzae/all_assembly_versions/) on September 10, 2019 for analysis. The collection was composed of a heterogeneous group of isolates, in terms of geographical coverage. Based on the clinical source, the isolates were classified into two groups, COPD (*n*=373 isolates) and non-COPD (*n*=195 isolates).

5.5.2 *De novo* genome assembly, annotation and pan genome analysis

Raw Fastq files generated from the Illumina sequencer were uploaded to the Galaxy web platform (Afgan et al., 2018) and the St. Petersburg genome assembler (SPAdes) tool (Bankevich et al., 2012) was used for the *de novo* assembly of the sequence reads (Afgan et

al., 2018). The default settings for all parameters were used, except for the size of k-mers, which were manually chosen as 21, 33, 43, 53, 63, 75. The quality of genome assembly was evaluated using the Quality Assessment Tool (QUAST) (Gurevich et al., 2013). Coverage of the reference genome was determined by aligning all of the sequenced genomes to the reference complete genome of the strain 86-028NP (Harrison et al., 2005). The 40 isolates sequenced in this study had on average 120 contigs (>500 bp), with mean genome size of 1,906,568 bp, reference genome coverage of 87.3%, and average read depth of 118.8 -fold.

The assembled FASTA/FNA files of 40 RHH isolates and 528 publicly available datasets were uploaded to the NeCTAR research cloud (http://cloud.nectar.org.au/) for subsequent analyses. We annotated relevant genomic features on the assembled/downloaded contigs using command line software tool, Prokka version 1.12 with default parameters (Seemann, 2014). An e-value threshold of 10^{-6} was used to determine the best match to known proteins in the databases, which includes UniProt, Pfam and TIGRFAMs. If no matches were found, an open reading frame was labelled as an 'hypothetical protein'.

The annotated genome assembly outputs from Prokka (in GFF3 format) were aligned to build large-scale pan genomes using a rapid large-scale prokaryote pan genome analysis pipeline, Roary version 3.12 (Page et al., 2015). An additional argument '-e --mafft' was added to generate a multiFASTA alignment of core genes using MAFFT (Katoh et al., 2002). All other Roary parameters were used as default; minimum blastp identity of 90%; Markov Clustering Algorithm (MCL) inflation value of 1.5. The genes identified within the genome alignment were classified with respect to their presence among the isolates: core (≥ 95%), shell (≥ 15% and < 95%), and unique or cloud genes (<15%). The output files from Roary, *gene_presence_absence.csv* (Dataset 2) and *accessory_binary_genes.fa.newick*, were

visualised using a Python script *roary_plots.py*, developed by Marco Galardini (Galardini, 2013).

5.5.3 In silico MLST

The draft and complete genomes of 568 NTHi isolates were genotyped based on the allelic profile of seven housekeeping genes (*adk, atpG, frdB, fucK, mdh, pgi* and *recA*) hosted at https://pubmlst.org/hinfluenzae/ (Jolley et al., 2018). Each isolate was assigned with a sequence type using *in silico* MLST typing (Seeman, 2014). Minimum spanning tree (MST) was then generated based on the MLST profiles using the goeBurst algorithm (Francisco et al., 2009) and the tree was visualised using Phyloviz (Francisco et al., 2012).

5.5.4 Core genome SNP extraction and identification of genetic clusters

The core genome alignment file was converted into a genlight object using the function *fasta2genlight* in the R package *adegenet* (version 2.1.1, RStudio version 1.0.143) (Jombart, 2008). The core genome SNPs (cgSNPs) were extracted and analysed to determine clusters of genetically-related individuals using the multivariate analysis method called Discriminant Analysis of Principal Components (DAPC) (Jombart et al., 2010). The cgSNPs raw data were initially transformed using principal component analysis (PCA), followed by the identification of genetic clusters using the *k-means* clustering algorithm. *k-means* determines a given number of groups (clusters) such that sequences belonging in the same cluster are more similar to each other than to sequences in other clusters. This was achieved using the function *find.clusters* with 150 principal components (PCs) retained, accounting for >90% of the sample variability. The optimal number of clusters was then inferred using Bayesian Information Criterion (BIC)

as eight genetic clusters which were then efficiently described using the *dapc* function with 60 PCs retained. The α-score was used to choose the optimal number of PCs retained for DAPC analysis. The α-score is the difference between reassignment probabilities for the true cluster and randomly permuted clusters. The average α-score for 60 retained PCs was found to be 0.70. The first three eigenvalues (discriminant functions) were then selected for visualisation and interpretation.

In addition, we performed DAPC analysis on the presence and absence profile of accessory genes to assess if the composition of the accessory genome supports the partitioning of the collection into the identified clades. The same predefined clusters (Clade I-VIII) identified by the core genome-based k-means clustering, were used to group the isolates. The accessory genome DAPC was carried out using the same methodology used for the core gene DAPC analysis except that the gene presence and absence matrix file was used instead of the core genome alignment file. We retained the first 60 PCs and three eigenvalues to examine the genetic clusters. The mean α-score for 60 retained PCs was found to be 0.68.

5.5.5 Phylogenetic analysis

cgSNPs were further utilised for performing phylogenetic analyses using the MEGAX software (Kumar et al., 2018). Evolutionary genetic distances between the strains were computed using the Maximum Composite Likelihood method (Tamura et al., 2004). The evolutionary relationship was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). Reliability of tree topology was tested using a bootstrap interior-branch test (Sitnikova, 1996). All the branches were supported by the bootstrap values >90%.

5.5.6 Identification of COPD-associated genes

The distribution of COPD strains in the subpopulation (clades) of isolates was assessed. We also evaluated the genetic distinction between COPD and non-COPD strains using DAPC analysis on cgSNPs and the 'presence/absence' profile of accessory genes. A pan-genome-wide association study (pan-GWAS) approach was applied to determine if any genes in the accessory genome were linked to a particular patho-phenotype including COPD using Scoary (Brynildsrud et al., 2016). Scoary was implemented in Python using *gene_presence_absence.csv* (output from Roary) and *trait.csv* for the genotype and phenotype input files, respectively. Each candidate gene in the pan genome was scored according to its apparent correlation to the clinical phenotype using a 2×2 contingency table of the 'presence/absence' profile of each gene for the clinical phenotype. A Fisher's exact test was performed on each gene in a population-agonistic manner. The Benjamini-Hochberg false discovery rate (FDR) adjustment was applied to correct for multiple comparisons (Benjamini and Hochberg, 1995). The cut-off for a significant association was a *p*-value lower than 0.05. Furthermore, for casual inference, pairwise comparison algorithm was implemented that controls for spurious associations dependent on the population structure (Maddison, 2000). The phylogenetic tree calculated internally by Scoary from the Hamming distances in the genotype matrix was used for the pairwise comparisons (Brynildsrud et al., 2016). The causal association was scored as significant when the *p*-values for both the best and the worst pairings were lower than 0.05. Finally, an additional label-switching permutation was implemented by running pan-GWAS on randomly permutated phenotypic values between individuals for 1000 times and retaining the 5% quantile, referred to as an empirical *p*-value (North et al., 2003). A minimum allele frequency threshold of 5% was used so that genes present in either more than 28 isolates or less than 540 isolates were included in the pan-GWAS analysis to avoid assigning too high importance to very rare genes/variants.

5.5.7 Functional annotation and classification of candidate COPD genes

The nucleotide sequences of candidate genes were translated to their corresponding peptide sequences using EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). The peptide sequences were then queried against the UniProt database using the basic local alignment search tool with an E-threshold of 0.001 (https://www.uniprot.org/blast/). In order to better understand the underlying biological processes, we performed functional classification of the COPD-associated genes using the Gene List Analysis Tool that is accessed through the web-based PANTHER version 14 classification system (http://pantherdb.org/) (Mi et al., 2019). The complete Gene Ontology (GO) annotation system, which consists of three datasets, was used for mapping the functions of the genes of interest (Ashburner et al., 2000). We classified our genes based on the GO molecular function and biological process.

5.5.8 Ethical approval

This work was conducted in accordance with Ethics Approval #H0016214 from the Tasmanian Health and Medical Human Research Ethics Committee.

5.6 Results

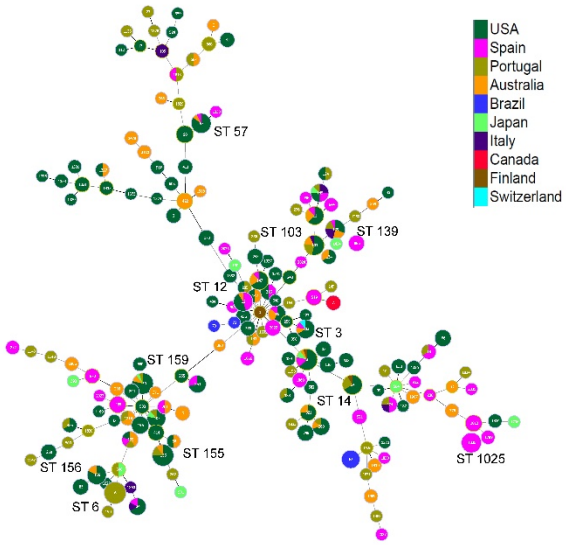
5.6.1 Global collection of 568 NTHi strains, including 40 newly sequenced Australian isolates

For the initial assessment of the genetic diversity of the collection of NTHi isolates, we determined their sequence type (ST) using *in silico* MLST allelic profile. Based on the MLST profile, the 568 NTHi isolates were assigned to 174 unique sequence types (STs) (Jolley et al.,

2018). Of which, 70 STs were associated with COPD, 34 STs contained both COPD and non-COPD strains and the remaining 70 STs included non-COPD strains only (Figure 5.1B). Sixty-one STs contained more than two NTHi isolates, of which 27 STs were found to be associated with COPD (Figure 5.1B). Some of the important COPD-associated STs that contained 5 or more NTHi isolates were ST12 (*n*=11), ST98 (*n*=6), ST196 (*n*=7), ST349 (*n*=5), ST485 (*n*=5), ST1025 (*n*=13), and ST1812 (*n*=5).

A minimum spanning tree (MST) overview of all the isolates was generated by Phyloviz (Francisco et al., 2012) (Figure 5.1) using the goeBurst full MST algorithm based on MLST profile (Francisco et al., 2009). The MST overview was overlaid by the isolation data based on the geographical (Figure 5.1A) and clinical (Figure 5.1B) sources of the isolates. The distribution of NTHi collection was uniform over the entire MST, in terms of geographical and clinical isolation (Figure 5.1A and 5.1B). There was no clustering specific to a geographical area. The US, Spanish, Portuguese and Australian isolates were present in all groups and scattered throughout the entire tree. Moreover, COPD isolates were also dispersed all over the MST tree. By applying the traditional definition of clonal complex (CC) (genotypes which have allelic profiles that differ from the ST genotype at only one of the seven MLST loci *i.e.*, having six out of seven identical MLST genes) (Feil et al., 2004), we found 119 different CCs, of which 74 included a single ST. The largest clonal complex consisted of only five STs and contained both COPD and non-COPD strains. Nine CCs included 3 or more STs; none of which were found to be specific to COPD phenotype. This indicates a weak or no association between MLST genotype and the disease phenotype (COPD).

A



B

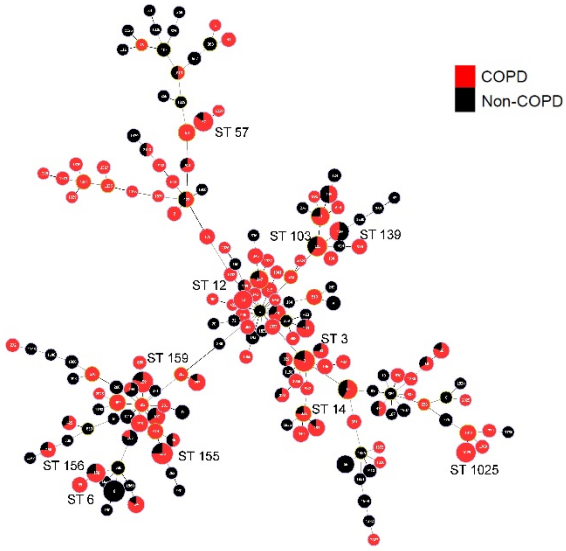


Figure 5.1 Minimum spanning tree (MST) overview of 568 NTHi isolates based on Multi-Locus Sequence Typing (MLST), i.e. allelic profiles of seven housekeeping genes present in the PubMLST database.

This was generated using the goeBurst full MST algorithm and was visualised with Phyloviz 2. Each node is a sequence type (ST), and it is coloured according to the (A) geographical and (B) clinical sources of the isolates. The size of each node is proportional to the number of isolates. The larger STs containing more than 10 NTHi isolates are labelled in-text. There was no absolute separation of the strains according to geography. (A) Based on the MLST profiling, strains from the same STs were common to a wide range of geographical locations. (B) Clusters of COPD-associated strains were scattered over the MST-based on the MLST allelic profile.

Out of 174 unique STs assigned to our collection, 70 STs were associated with COPD strains and 34 STs included strains from mixed infections, *i.e.* COPD and other clinical illnesses.

5.6.2 Pan genome analysis of 568 NTHi isolates

A pan genome of 12,249 genes was generated from the 568 draft NTHi genomes, including 40 newly sequenced isolates in this study (Figure 5.2). Here, the core genome was represented by 853 genes that were present in at least 539 isolates. The core genes accounted for approximately 47% of the number of genes (1,821) present in the reference NTHi genome (86-028NP) (Harrison et al., 2005). The accessory pan-genome which comprised of the shell and cloud genes, accounted for a larger proportion (93%) of the pan genome (Figure 5.2). The large accessory pan-genome confers diversity and high levels of genomic plasticity to NTHi. The majority of accessory pan-genes (81% of pan genes) were rare and strain-specific, found in less than15% of the NTHi collection. The distribution of genes in the population had a characteristic U-shape as reported in previous studies in *H. influenzae* and other bacterial species (Donati et al., 2010, De Chiara et al., 2014, Touchon et al., 2009).

5.6.3 Collection of 568 NTHi isolates exhibits a population structure defined by eight clades

The core genome alignment generated by concatenation of individual core gene alignments was used to infer the population structure. We extracted 97,262 biallelic SNPs from the core genome alignment of 696,234 bp. To this dataset, we applied the discriminant analysis of principal components (DAPC) to infer the number of clusters of genetically related individuals. Bayesian Information Criterion (BIC) supported the partitioning of this collection of isolates

into eight clusters or clades that were clearly separated from each other, except a slight overlap between clusters I and II (Figure 5.3A).

We also tested whether the presence/absence profile of accessory genes supported the partitioning of the collection into the predefined eight clades. For this, we performed a DAPC analysis on the dispensable genes using predefined grouping as identified by the cgSNPs DAPC. Isolates of clades III, IV, VII and VIII were clearly separated from each other, while isolates in clades I and II were found to be more closely related, which was consistent with the cgSNPs DAPC (Figure 5.3A and 5.3B). This suggests that the isolates in clades I and II are evolutionarily related and have a common ancestor. The composition of accessory genome DAPC further showed a close association between clades V and VI, which was not observed with the cgSNPs DAPC (Figure 5.3A and 5.3B). This indicates that a set of accessory genes that are shared between isolates of clades V and VI could have been either inherited from a common ancestor or acquired through horizontal gene transfer.

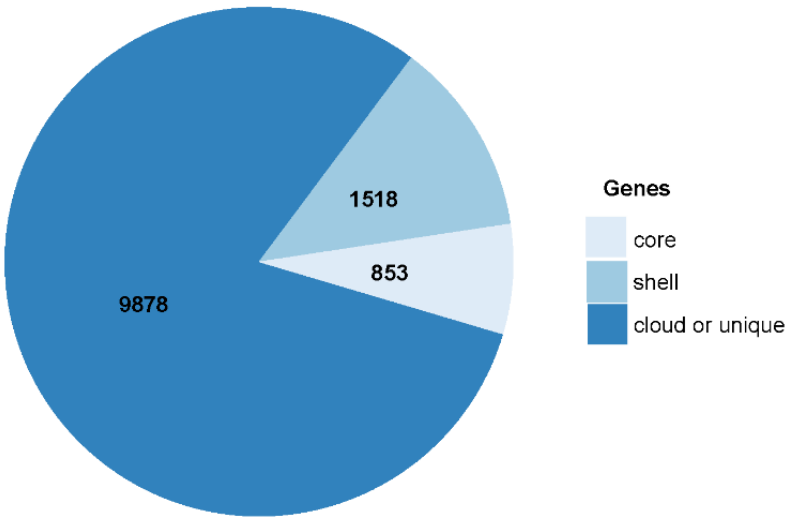


Figure 5.2 Distribution of genes present in the pan genome of NTHi collection (n=568), which was constructed using Roary version 3.12.

Of 12,249 pan genes, the NTHi core genome comprised 853 genes (present in at least 539 NTHi isolates). The remaining 11,396 genes of the accessory genome were further classified into the shell (*n*=1518 genes, present in less than 539 and at least 85 NTHi isolates) and cloud or unique (*n*=9,878 genes, present in less than 85 NTHi isolates). On average, 47% of each NTHi strain's gene set belonged to the core pan-genome. The remaining 53% of the strain's gene set belonged to the larger accessory pan-genome. This accessory pan-genome encompassing a large repertoire of genes confers diversity and high levels of plasticity to the NTHi genome.

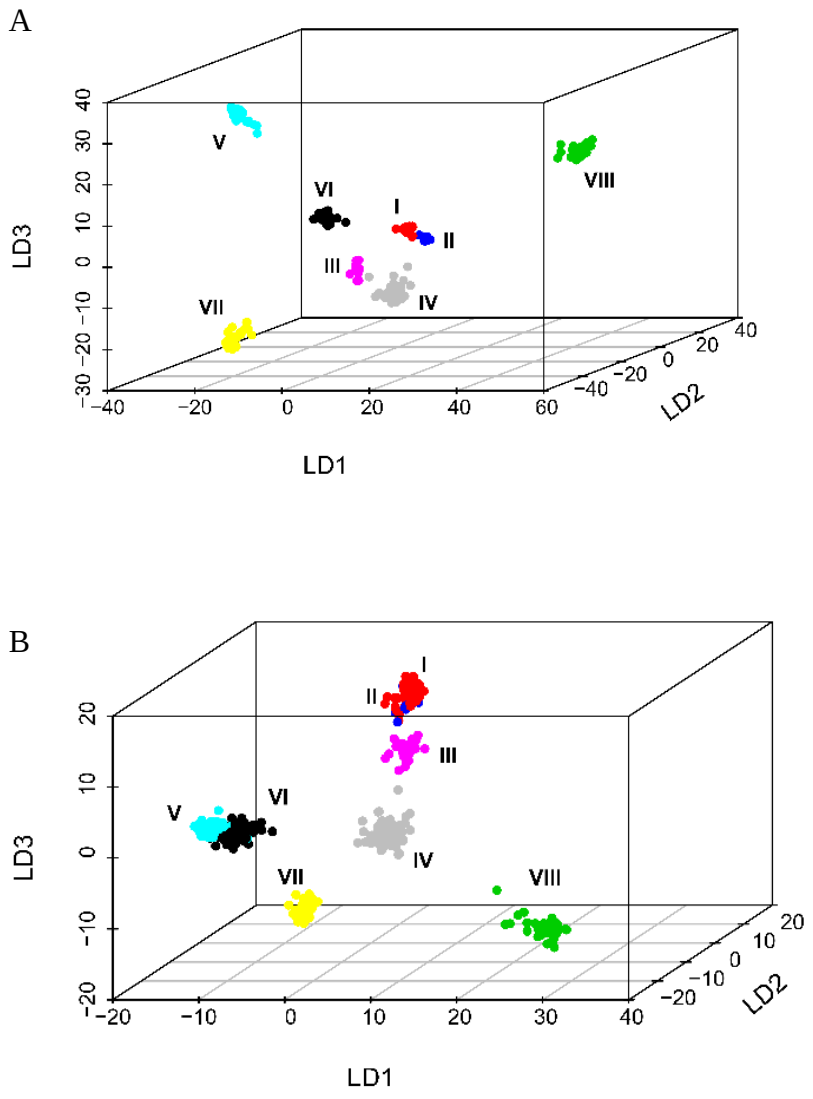


Figure 5.3 (A) 3D scatterplot of the core genome SNPs-based Discriminant Analysis of Principal Components (DAPC) of NTHi isolates (*n*=568).

DAPC resolved the NTHi isolates into eight clusters (clades). Clades I and II were closely related to each other, whereas all other clusters were distinctly separated. (B) The accessory genes based DAPC correlated with the cgSNPS-based DAPC with a distinct separation of clades III, IV, VIII and VIII and a close relationship between clades I and II. However, the only discrepancy was with the isolates of clades V and VI, which were clearly separated on cgSNPs DAPC whereas overlapped on the accessory gene content DAPC. Each dot is an isolate, coloured according to the classification into one of the eight clusters/clades as assigned by the core genome SNPs-based DAPC analysis.

5.6.4 Phylogeny separates isolates into groups that correlate with clades

We then compared the clade partitioning to standard phylogenetic analyses. We built a NJ phylogenetic tree using the cgSNPs (Figure 5.4). Molecular phylogenetic analysis correlated perfectly with the clade partitioning as defined by the population genetics, supporting the clonal nature of NTHi population characterised by eight distinct lineages. Isolates of clades I and II were evolutionarily related, as suggested by the cgSNPs and the accessory genes DAPC analyses. Consistent with the accessory genes DAPC, phylogenetic analysis showed a close relationship between clades V and VI. This suggests that a set of accessory genes shared by isolates of clades V and VI were vertically transferred from a common ancestor.

The clade partitioning, as defined by population genetics and evolutionary relationship based on the cgSNPs, partially correlated with the general clustering based on the MLST profile (Figure 5.1) as shown in the Figure 5.5. Clades I, II and III and Clades V and VI were clustered together in the MST (Figure 5.5) which correlated with the phylogenetic tree topology (Figure 5.4). However, isolates of Clades IV, VII and VIII did not form a distinct cluster and were uniformly dispersed over the MST tree (Figure 5.5).

5.6.5 Composition of the accessory genome but not the distribution of cgSNPs separates COPD from non-COPD strains

We then tested whether the 373 COPD strains could be associated with a particular clade(s). We overlaid the information on the clinical source of isolates on the NJ phylogenetic tree as shown in Figure 5.6. We found that COPD strains were distributed among all eight clades with no absolute separation of the strains according to clinical source (Figure 5.6). Expanding on our analysis, we next investigated whether the information on core genome-wide polymorphic sites could be used to classify the collection into groups based on their clinical phenotypes. Discrimination analysis on cgSNPs showed significant overlap between COPD and non-COPD isolates, which is consistent with the previous results, suggesting poor resolution of cgSNPs for separating isolates according to their clinical source (Figure 5.7A).

We then applied DAPC analysis on the composition of the accessory genome, and found that, based on the first discriminant function, the presence/absence profile of dispensable genes could separate the COPD strains from the rest (Figure 5.7B). The mean discriminant function separating the two populations was significantly different ($p < 0.0001$, unpaired t-test with Welch's correction). This result indicates that there is a difference in the composition of the accessory genome in NTHi that could potentially separate COPD strains from those associated with other clinical illnesses.

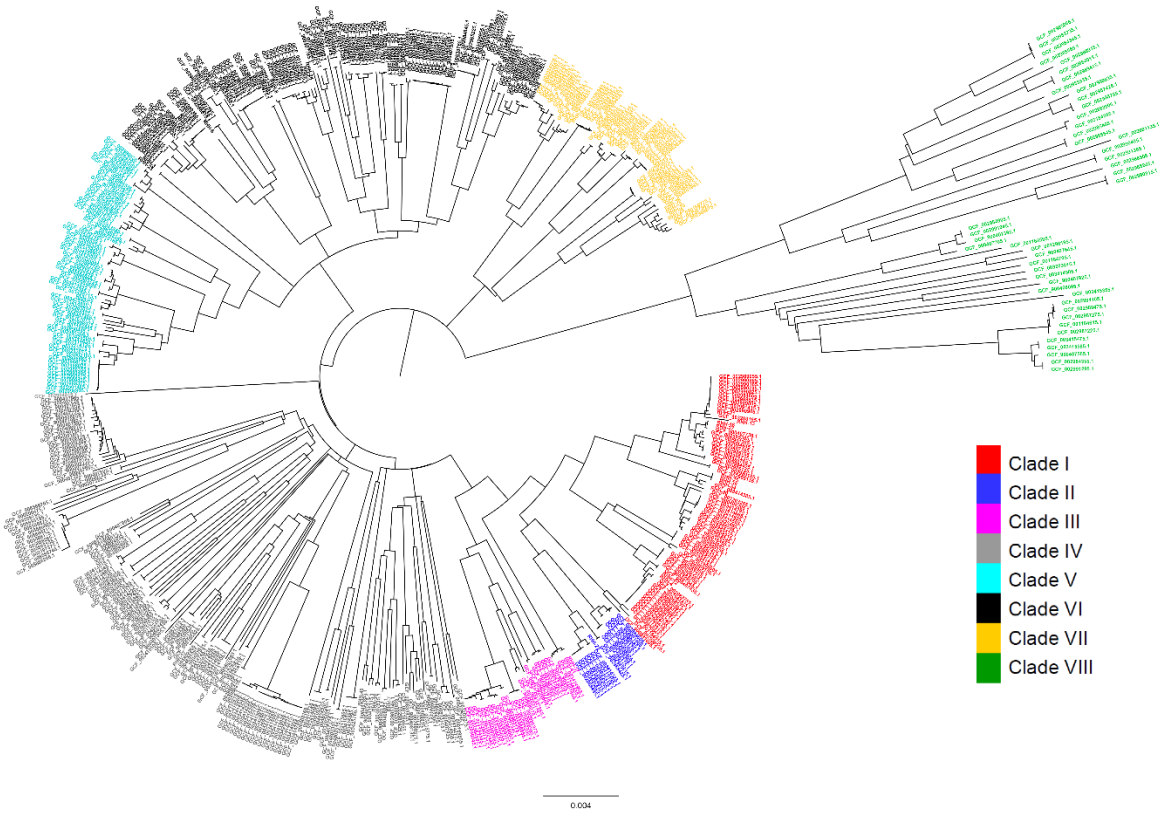


Figure 5.4 (A) Neighbour-joining phylogenetic tree of 568 NTHi isolates, based on the core genome SNPs.

The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. There were a total of 664,470 positions in the final dataset. Distinct sub-structuring of the NTHi population was evident with the core genome SNPs-based phylogeny. The phylogenetic analysis perfectly correlated with the DAPC-based classification that identified eight monophyletic clades. The close association between clades I and II, and between clades V and VI, as observed in accessory genome-based DAPC, is consistent with the evolutionary relationship between them observed in the core gene sequence-based phylogeny obtained using neighbour-joining maximum composite likelihood method. The tip labels are coloured according to the clades assigned by the core genome SNPs-based DAPC. SNP, single nucleotide polymorphism; DAPC, discriminant analysis of principal components.

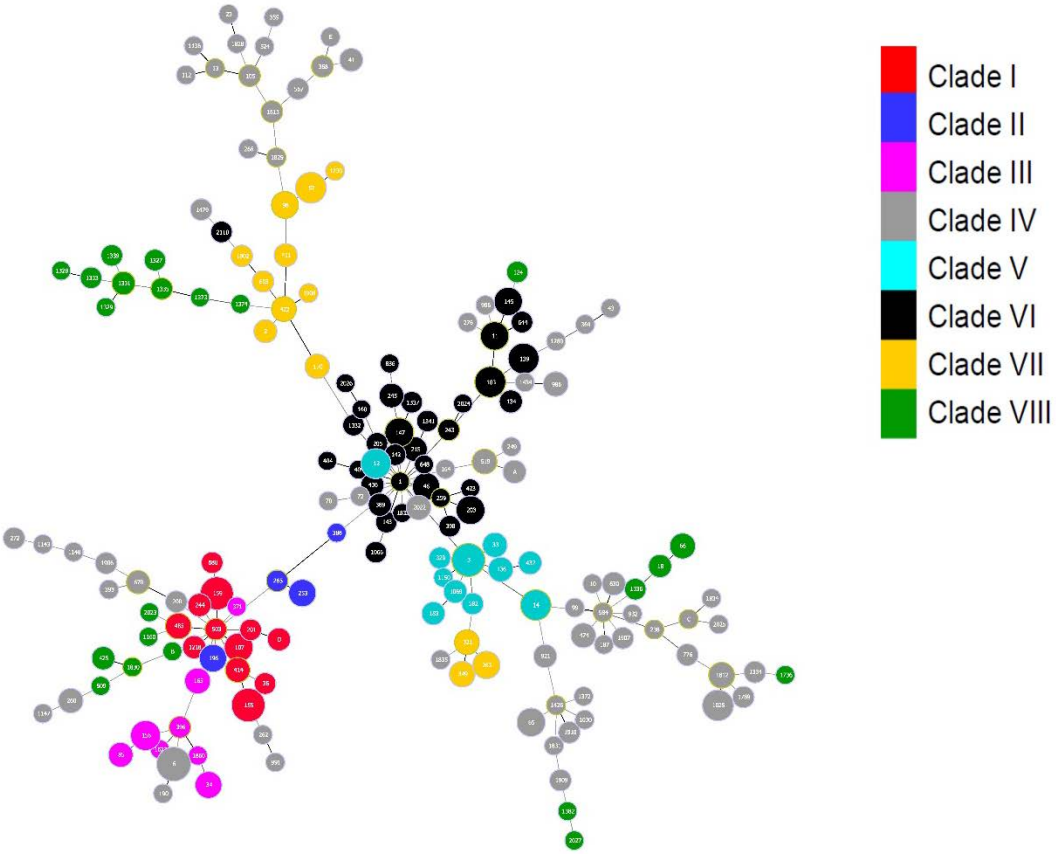


Figure 5.5 Classification of the isolates into Clades I-VIII highlighted on the minimum spanning tree (MST) based on the multi-locus sequence typing (MLST) allelic profile.

This was generated using the goeBurst full MST algorithm and was visualised with Phyloviz 2. Each node is a sequence type (ST). The isolates classified in clades I-III, V and VI appear to be correlated also in the MST. Moreover, clades I-III and V-VI, which are more closely related to each other according to the phylogenetic relationship based on the cgSNPs and the DAPC analysis of accessory genome composition, are also related on the MST. However, isolates of clades IV, VII and VIII are dispersed over the entire MST.

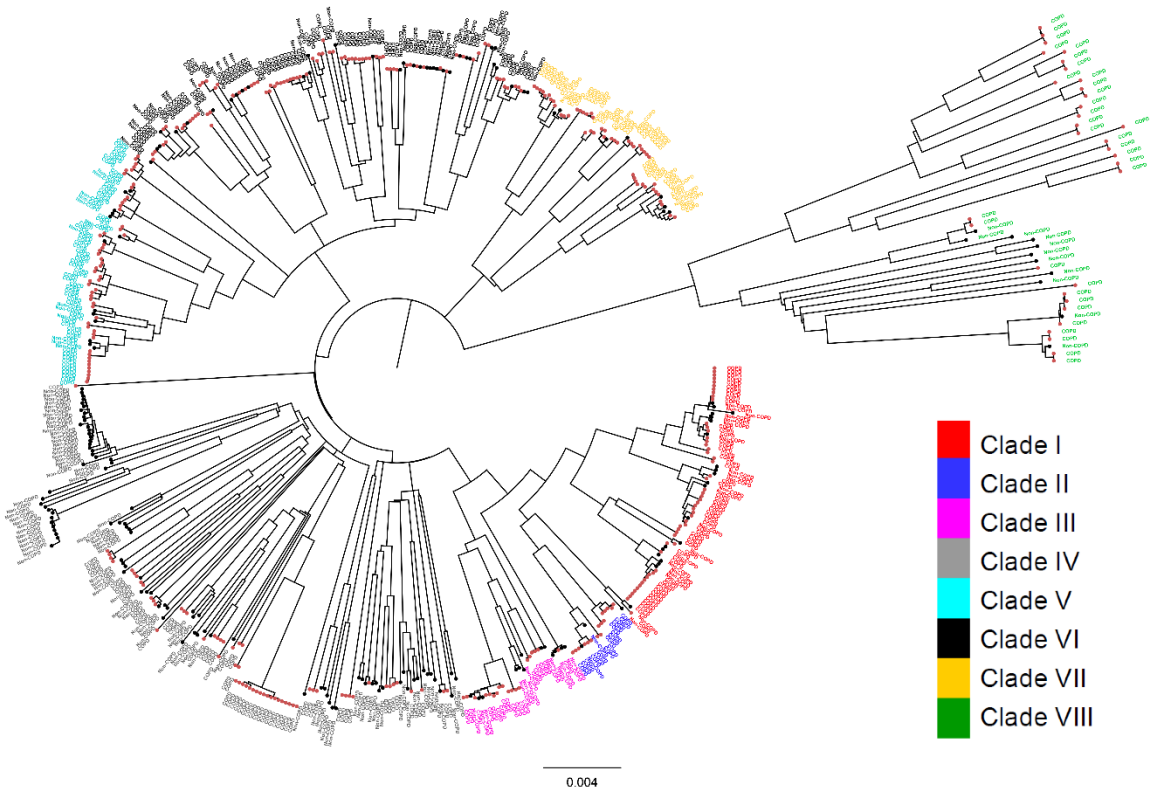
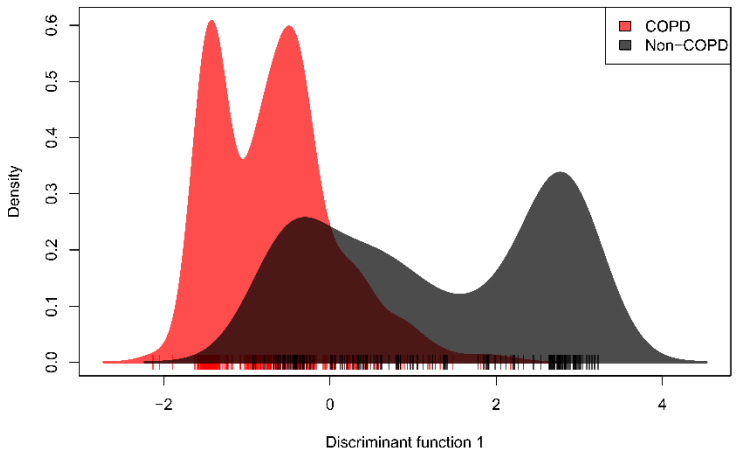


Figure 5.6 Distribution of 568 NTHi isolates over the core-genome SNPs-based Neighbour-joining phylogenetic tree, which is annotated with the clinical source of isolation of the samples as COPD or non-COPD.

Each evolutionary clade includes both COPD and non-COPD strains. Branch tips representing COPD strains are highlighted in red and the tip labels are coloured according to the clades (I to VIII) as assigned by the core genome SNPs-based DAPC analysis. COPD, chronic obstructive pulmonary disease; DAPC, discriminant analysis of principal components.

A



B

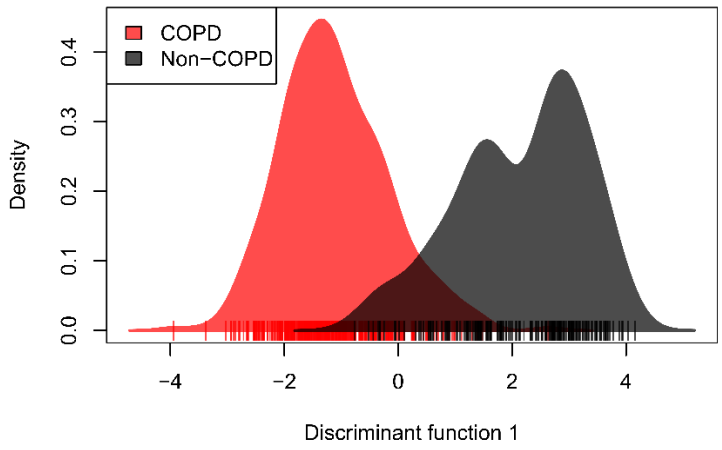


Figure 5.7 Discriminant Analysis of Principal Components (DAPC) of 568 NTHi genomes from COPD and other disease isolates.

(A) The first discriminant function of the retained PCs based on core genome SNPs leaves substantial overlap between COPD and non-COPD strains, 356 NTHi isolates (231 COPD and 125 non-COPD strains) were in the overlapping region. (B) DAPC analysis on presence/absence profile of accessory genes clearly provides a higher level of separation of COPD from non-COPD strains with only 226 NTHi isolates (119 COPD and 107 non-COPD strains) in the overlapping region. Composition of accessory genes, but not the distribution of polymorphic sites in the core genome, discriminates COPD strains from other clinical

phenotype strains of NTHi. Each line is an isolate. COPD and non-COPD isolates are coloured in red and black, respectively.

5.6.6 Genome wide association studies identified a set of accessory genes that are significantly associated with COPD strains

We determined the accessory genes in the bacterial dataset that were associated with a COPD phenotype. Scoary, a pan genome-wide association studies (pan-GWAS) tool, was used to screen the genes for association with COPD (Brynildsrud et al., 2016). After the FDR correction, Scoary predicted 680 genes present in the NTHi pan genome to be significantly enriched in NTHi collection from the COPD airways (Dataset 3). Out of the 680 genes, 226 genes were found to be significantly overrepresented in the COPD strains as compared to non-COPD strains (odds ratio ≥ 2, $p < 0.05$) (Dataset 3).

Additionally, Scoary identified 145 (out of 680) genes to be significantly associated with the COPD phenotype ($p < 0.05$) after pairwise comparison (Dataset 3). The other remaining 535 genes identified as significant prior to population-aware analysis were found to be lineage-specific effects upon inspection of the population distribution by the pairwise comparisons test. Furthermore, 1000 random permutations of the phenotype data were implemented, and the associated test statistic was calculated for each permutation. After the permutations, only 122 genes were found to have a significant association with the COPD phenotype (Dataset 3). Out of 122 significant hits, 86 hits were found to be different alleles (variants) of the same gene, one positively and one negatively associated with the COPD phenotype. The two alleles of these 43 genes were different enough to not be clustered as the same by Roary. Finally, there were 79 unique genes likely to play a role in COPD (Dataset 3).

5.6.7 Functional classification of candidate COPD-associated genes

We next predicted the biological functions of the 79 unique COPD candidate genes. Seventy out of 79 genes mapped to UniProt proteins with a minimum identity threshold of 90% (Dataset 4). Furthermore, the candidate genes were functionally annotated by assigning their encoded proteins to Clusters of Orthologous Groups of proteins (COGs) (Tatusov et al., 2001). The COG analysis, however, did not result in an increase in confident functional prediction for the candidate genes as compared to the original annotations, *i.e.* 70 genes were assigned to known orthologous groups when identity threshold was maintained at 90% (Dataset 4).

Additionally, the functionally annotated genes that confer COPD phenotype were classified based on their molecular functions and the biological processes they are involved in. Fifty-four and fifty-two of the 70 functionally annotated COPD-linked genes were assigned to GO molecular functions and biological processes, respectively (Dataset 4). A large number of these candidate genes were found to be associated with transmembrane transporter activity (n=10), regulating the transport of inorganic phosphates, cations (Na^+ and K^+), metal ions such as copper and iron, lactate, dicarboxylate, carbohydrate, amino acids and proteins in and out of the bacterial cell (Table 5.1). Others were genes involved in cell redox homeostasis and cellular carbohydrate and protein metabolic processes, including the biosynthesis of aromatic and branched-chain amino acids. Consistent with previous studies (Fernaays et al., 2006, Murphy et al., 2011), a variant form of immunoglobulin A peptidase gene has been found to be strongly associated with the COPD phenotype (odds ratio = 4.4, p = 0.0012). In addition, variants/alleles of genes encoding glycosyltransferases, such as *lex1* and *isgE*, and cytidylyltransferase encoding *licC* that are involved in lipooligosaccharide (LOS) biosynthesis were also found to be associated with the COPD strains. With regard to other virulence genes such as *Haemophilus* adhesion and penetration protein (Hap), higher molecular weight proteins 1 and 2 (HMW1/2),

and *Haemophilus influenzae* adhesin (Hia), they were found to be uniformly distributed among COPD and non-COPD strains.

H. influenzae is known to be naturally competent for transformation (Lorenz and Wackernagel, 1994, Sisco and Smith, 1979). *tfoX* or *sxy* is a regulatory gene required to positively regulate DNA uptake and transformation (Redfield, 1991). *H. influenzae* has a *tfoX* gene that encodes TfoX, which interacts with cyclic-AMP receptor protein (CRP) and promotes the expression of genes of the competence regulon (Karudapuram and Barcak, 1997, Redfield et al., 2005). *H. influenzae* utilises Type IV pili for the transport of DNA across the membrane into the cytoplasm (Chen and Dubnau, 2004). Variant forms of both *pilA* (encodes PilA, a major pilin subunit of type IV pili) (Carruthers et al., 2012) and *tfoX* are found to be associated with the COPD strains of NTHi (Table 1). RecJ is an exonuclease with 5'-3' single-stranded-DNA-specific exonuclease activity that plays a crucial role in DNA repair and recombination pathways (Sharma and Majumdar, 2009). RecJ is associated with mismatch repair, and in addition, in conjunction with RecQ helicase, initiates recombination from a double-stranded break (Burdett et al., 2001). A variant of *recJ* has also been found to be associated with the COPD phenotype (odds ratio = 2.5, $p = 0.019$).

Table 5.1 List of virulence and transformation competency associated genes that significantly correlated with COPD strains of NTHi.

Gene	Gene Name	OR	p	Pairwise p	Emp p	Function
igA	Immunoglobulin A1 protease autotransporter	4.5	0.0012	1.4×10^{-7}-15.3×10^{-6}	0.0009	Virulence
isgE	N-acetyl-glucosamine-transferase	5.5	0.0008	1.4×10^{-6}-6.6×10^{-5}	0.0009	LOS synthesis
lex1	Lipooligosaccharide biosynthesis protein Lex1	2.8	0.0274	3.6×10^{-5}-0.0227	0.0009	LOS synthesis
licC	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	6.4	4.9×10^{-9}	4.0×10^{-8}-2.3×10^{-7}	0.0019	LOS synthesis
pilA	Type IV pilin subunit protein PilA	5.1	0.0002	3.1×10^{-6}-3.1×10^{-6}	0.003	Adhesion; Transformation
tfoX	DNA transformation protein	8.6	0.0003	3.7×10^{-8}-4.2×10^{-7}	0.001	Transformation

OR, odds ratio; p, false rate discovery adjusted p value; Pairwise p, range of p values from the pairwise comparisons; Emp p, empirical p values after 1000 permutations.

5.7 Discussion

NTHi is associated with a wide range of diseases, including otitis media, meningitis and conjunctivitis, and is a major bacterial cause of exacerbations in COPD patients (Murphy et al., 2004, Thanavala and Lugade, 2011, van Wessel et al., 2011, Bodor et al., 1985). The ability to predict a disease phenotype based on the genotype of a pathogen would be valuable in informing an appropriate prevention and treatment response. In some cases, traditional phenotyping methods such as virulence factor profiling demonstrates clustering of the bacterial isolates into specific serotypes as in the case of *Streptococcus pneumoniae* (Hausdorff et al., 2005). While in others, MLST profiling correlate well with the disease induced by a pathogen, as in the case of *Enterococcus faecium;* for instance few particular MLST sequence types: ST796, ST1421, ST233 and ST80 are associated with vancomycin-resistance (Leong et al., 2018, Leong et al., 2019). In terms of infections with *H. influenzae*, capsulated strains have

been reported to form serotype-specific clusters, in particular serotypes c, d, e, and f that formed monophyletic clusters on a dendrogram constructed from the MLST allelic profile (Meats et al., 2003, De Chiara et al., 2014). However, the population of NTHi has been reported to be composed of heterogenous group of isolates that formed highly divergent clusters on the MLST-based MST (De Chiara et al., 2014). Among the diseases caused by NTHi, Brazilian purpuric fever, has been found to be caused by a well-defined NTHi clone (biogroup Aegyptius) (Brenner et al., 1988), however, any correlation between genotypes of NTHi and COPD has not been established yet (De Chiara et al., 2014). In this study, we examined the whole genome sequences of a large collection of NTHi strains that were isolated from different clinical sources, to reveal genetic basis of distinction between COPD and other phenotypic strains.

Our analyses indicate that conventional MLST typing exhibits low discriminatory power and is thus, unsuitable for identifying COPD specific clusters of NTHi. To increase the discriminatory power, we expanded the MLST scheme that comprises seven housekeeping genes and included 853 core genes in further analysis. We performed DAPC analysis on cgSNPs, which grouped 568 NTHi isolates into eight distinct clades (Figure 5.3A). In this larger panel of datasets analysis, we successfully identified discreet NTHi subpopulation structures that were less resolved in previous analysis by De Chiara and colleagues where fewer isolates were taken into account (De Chiara et al., 2014). NTHi isolates that fall within De Chiara's clade I and V were further resolved into two distinct clusters, however, the other clades (II, III, IV and VI) did not substructure with the analysis of 568 NTHi isolates Each clade comprised of NTHi strains from diverse disease phenotypes. Moreover, phylogeny derived from an analysis of cgSNPs did not provide sufficient resolution for grouping NTHi based on their clinical source. This result is consistent with the study conducted by De Chiara's group and later by Pettigrew and colleagues that analysed a large collection of 403 NTHi

genome sequences (De Chiara et al., 2014, Pettigrew et al., 2018). The finding that NTHi strains with highly similar core genome sequences caused a range of diseases suggests that differences in non-core accessory genes could contribute to the distinction between disease phenotypes (Lewis et al., 2017).

The eight distinct clades supported by population genetics perfectly correlated with molecular phylogenetic analysis based on the cgSNPs (Figure 5.3A and 5.4). Analysis of the composition of the accessory genome further correlated with the phylogeny (Figure 5.3B and 5.4). This consistency highlights the clonal nature of the NTHi population as described previously (De Chiara et al., 2014). We also identified the genes that were enriched (over-represented) in each clade (Dataset 5) and found a set of clade enriched genes distributed over the evolutionarily distant clades. For example, out of 456, 511, and 417 genes that were specifically overrepresented in Clades I, VII and VIII (FDR adjusted p value < 0.001), 11 genes were found to be common in these phylogenetically distinct clusters. This underscores the role of horizontal gene transfer prevalent in the NTHi population that accounts for genetic diversity among the species (Power et al., 2012)

We then applied discriminant analysis on the composition of accessory genes, which demonstrated a clear separation between strains associated with COPD and other clinical illnesses. Moreover, using a pan-GWAS approach, we identified a subset of NTHi accessory genes associated with the COPD phenotype. Some key COPD-associated genes likely to be involved in virulence or pathogenesis are listed in Table 5.1. NTHi utilises cells surface structures, such as LOS, fimbriae and outer membrane proteins to interact with the host cells and pave the way for colonisation and invasion (Foxwell et al., 1998). We identified a variant form of *pilA*, encoding type IV pilin subunit protein and LOS biosynthetic genes (*licC, lex1* and *isgE*) to be associated with the COPD strains. PilA and LOS have been demonstrated to

play an important role in biofilm formation and colonisation of the respiratory tract (Jurcisek et al., 2007, Swords et al., 2004). Evasion of host immune defence is probably the other mechanism how NTHi strains thrive in the COPD airways. A variant form of *iga*, encoding IgA protease that cleaves the immunoglobulin A has been identified to be specifically associated with the COPD isolates of NTHi (Lomholt et al., 1993). This association was not confirmed by the previous study that analysed 97 NTHi isolates, including 20 COPD strains. The low number of COPD strains in the analysis might have undermined the statistical power for correlation which is now resolved by our study that included a large number of COPD strains and established a strong association between them.

Pettigrew's and Molere's groups recently investigated large prospectively collected NTHi genomes to give insight into molecular changes during persistence in the COPD lung. (Moleres et al., 2018, Pettigrew et al., 2018). They found genetic changes in multiple genes that regulate expression of virulence functions, such as adherence, nutrient uptake and immune evasion that are likely to be involved in NTHi to the COPD lung. This suggests that in comparison to other ecological niches, such as the middle ear, sinuses, eye, meninges, and the upper respiratory tract, NTHi in COPD airways are exposed to different microenvironments defined by distinct nutrient availability, pH, oxidising potential and/or immune response. NTHi therefore, appears to undergo genomic changes to survive and adapt in the hostile environment of COPD airways. Consistent with these findings, we found that COPD isolates of NTHi encoded different metabolic activities compared to strains associated with other clinical phenotypes. This suggests that metabolic capacity, in part, plays an important role in enabling NTHi to contribute to COPD pathogenesis and further supports the concept of nutritional virulence as an important determinant of pathogenic capability in NTHi (Schoen et al., 2014).

In conclusion, our study indicates that the virulence and survival of NTHi in COPD is influenced by genes outside of the core genome. The set of accessory genes associated with COPD strains might impact the strains for establishing a successful niche in COPD airways by acquiring nutrient sources, evading immune response and by enhancing adhesion and colonisation of airways. In addition, variation in competence and recombination association genes may enable NTHi strains to acquire genes that confer a competitiveness advantage in the COPD airways. Further work will examine how the finding that COPD strains of NTHi possess a distinct gene content could be translated into improvements in the management of NTHi infections in COPD.

6 Chapter 6: Summary

COPD is a major public health problem affecting an estimated 300 million people globally. Besides tobacco smoking, exposure to biomass smoke is an important risk factor for COPD development and its progression. Despite the high burden of biomass smoke exposure, particularly in the low- and middle-income countries, only a small number of experimental studies have been conducted on the impact of biomass smoke on airway cells. In Chapter 1, I hypothesised that the generation of quantifiable and preservable smoke extracts could be used in multiple exposure experiments for comparative assessment of cellular responses to different types of smoke. To generate smoke extracts in-house, I developed a simple, reproducible and cost-effective method, which was successfully used to prepare quantifiable cigarette and biomass smoke extracts as discussed in Chapter 2. This procedure has advantages over the commonly practised method of bubbling combustion particles through cell culture medium in terms of reducing loss of yield of combustion products during preparation and generating extract for long-term use and storage. I tested the comparative effects of cigarette, cow dung and wood smoke extracts on inflammatory response using human bronchial epithelial cell line *in vitro*. I demonstrated both cigarette and biomass smoke extracts induced inflammatory responses, which were marked by increases in expression of PAFR and other pro-inflammatory mediators IL-1β, IL-6 and IL-8 on bronchial epithelial cells upon exposure to smoke extracts. Furthermore, I assessed the stability of smoke extracts over a period of 3 months and showed that the PAFR-inducing activity of smoke extracts was preserved for at least the 3-month period tested (Chapter 2). Thus, the smoke extracts prepared in quantifiable form in DMSO was preservable and was successfully used in multiple exposure experiments for comparative assessment, which proves our hypothesis to be correct. However, in terms of cellular response to smoke extracts, the assessment was limited to selected cellular receptor (PAFR) and cytokines. Although the chosen cellular receptor and cytokines were amongst the most

important inflammatory mediators, further investigation into a wider range of cellular pathways would provide a greater insight. Nevertheless, the ability to produce biomass smoke extracts in a reusable form, by using a cost-effective procedure will enable further experimental research in understanding the mechanistic role of biomass smoke exposure in the pathogenesis of respiratory disease including COPD.

We also tested the hypothesis that even with the removal of a core diazepine structure from the well-known PAFR antagonist WEB-2086, its potency towards inhibiting NTHi adherence to bronchial epithelial cells would be maintained. We synthesised a series of 2-aminothiphenes and 2-pyrrolothiophenes that were structurally analogous to WEB-2086 but lacked the core diazepine structure and tested their effects *in vitro* in reducing the adherence of NTHi to cigarette smoke extract-exposed bronchial epithelial cells (Chapter 3). In Chapter 3, I showed that the structural analogue of WEB-2086, compound C17, exhibited inhibitory activity against NTHi adherence on cigarette exposed cells that was comparable with the activity of WEB-2086. This proves our hypothesis to be correct that the core diazepine ring in WEB-2086 is not essential in for its activity to inhibit NTHi adhesion.

We then hypothesised that the exposure of bronchial epithelial cells to biomass smoke extracts (prepared by the method described in Chapter 2) would enhance adherence of NTHi. To test this hypothesis, I exposed bronchial epithelial cells to cigarette, cow dung and wood smoke extracts and assessed their effects on the adherence of NTHi as described in Chapter 4. I found cow dung but not wood smoke extract to enhance the adhesion of NTHi to bronchial epithelial cells. The results did not fully support our hypothesis as wood smoke extract did not exhibit any effect on NTHi adherence. Inducing effect of cow dung smoke exposure on NTHi adherence to BEAS-2B cells was found to be dose-dependent, which was similar to that of cigarette smoke exposure. Furthermore, I assessed if the compound C17 was effective in

reducing the cow dung smoke induced attachment of NTHi to the bronchial epithelial cells. I found the compound C17 to be as effective as WEB-2086 in alleviating the effect of cow dung smoke exposure towards NTHi adherence. This result further strengthens our previous hypothesis as the compound C17, a structural analogue of WEB-2086 that lacked a core diazepine ring, was demonstrated to have anti-NTHi adhesion activity against a wider range of smoke extracts.

Finally, we hypothesised that NTHi associated with COPD might exhibit genetic distinction from the strains isolated from non-COPD phenotypes. To test this hypothesis, I performed whole genome-based analyses on a larger set of 568 NTHi genomes collected from different disease phenotypes, including COPD (described in Chapter 5). This NTHi collection included 40 Australian isolates that were newly sequenced for my PhD work. I demonstrated that genotyping based on sequence profiles of both seven housekeeping genes and 500 core genes did not discriminate COPD strains from those isolated from other disease phenotypes. However, gene content analysis of accessory genome clearly segregated the NTHi population based on their clinical phenotypes. I further applied pan-GWAS approach and identified a set of accessory genes that were significantly associated with COPD strains. These results strongly support our hypothesis that there exists genetic difference between COPD and non-COPD strains. Functional analysis of the COPD candidate genes further revealed a set of biological processes that were significantly enriched in the COPD strains. These included metabolic functions that have key roles in cellular respiration and maintenance of redox homeostasis. Thus, in Chapter 5, we showed that specific metabolic genes may be required by NTHi for its survival or virulence in the COPD lung.

In conclusion, all the aims of my thesis have been fulfilled. This thesis has demonstrated the potential of biomass smoke to induce inflammation in the airways and promote bacterial

infection of respiratory epithelial cells. This provides insights into the mechanism how exposure to biomass smoke contributes to development and progression of COPD and provides experimental data to support the call for measures that reduce habitual exposure to biomass fuel smoke. In addition, the thesis further found that NTHi isolated from COPD is genetically distinct from other clinical phenotypes. This finding is important in the design of diagnostics and therapeutics that better target strains of NTHi that have the potential to induce acute exacerbations of COPD in patients.

Strengths and Limitations of the Thesis

In terms of strengths of the thesis, the new methodology described in this work generates biomass smoke extract in storable and reusable form for application in experiments with human bronchial epithelial cells. In addition, the ability to quantify the smoke extract with this method enables standardisation of assays and minimisation of technical variation between experiments. In addition, the work in this thesis has demonstrated that PAFR antagonist, and its analogues, has therapeutic potential in blocking attachment of NTHi to human bronchial epithelial cells exposed to cigarette and biomass smoke. This establishes PAFR as a potential drug target for alleviating the impact of tobacco and biomass smoke exposure. Furthermore, the thesis has produced a bioinformatics pipeline to elucidate the core and accessory genomes of NTHi and to distinguish NTHi isolates from COPD cases with respect to other clinical phenotypes. This pipeline can be used as tool for comparative genomics studies of a wide spectrum of pathogens, including NTHi In addition, experimental design and execution practiced in this thesis have produced data that have met the standards of international peer-review and acceptance for publication.

I acknowledge that the thesis still has several limitations. Firstly, the exposure experiments were entirely carried out *in vitro* using commercial bronchial epithelial cell lines. Further work is needed to evaluate the effect of smoke extracts on primary respiratory epithelial cells and also in the context of a fully functioning lung or immune system. Secondly, investigation of the impact of biomass smoke exposure on bronchial epithelial cells was limited to the adherence of a single respiratory pathogen, NTHi. This needs to be expanded to other respiratory pathogens, such as *Streptococcus pneumonia*, *Moraxella catarrhalis* and *Pseudomonas aeruginosa*. Lastly, a set of accessory genes has been identified to be associated with COPD strains that might impact the strains for establishing a successful niche in COPD airways. Further validation for the role of individual genes in contributing to enhanced survival and virulence of NTHi in the COPD lung is needed.

Future Directions

The findings presented in this thesis highlight the need for immediate global action towards reducing biomass fuel use and associated smoke exposure. In addition, the thesis further highlighted the potential of PAFR as a therapeutic target for alleviating the impact of biomass smoke exposure by reducing the risk of NTHi-infection in COPD. In addition, the findings presented here indicate that higher resolution diagnostics are possible for detecting NTHi infections that may present an increased risk of exacerbations in COPD patients. Translation of the findings from this thesis will contribute to the development of new management strategies for biomass smoke-related obstructive airway disease including COPD.

7 References

Adamus, W. S., Heuer, H., Meade, C. J. & Brecht, H. M. 1989. Safety, tolerability, and pharmacologic activity of multiple doses of the new platelet activating factor antagonist WEB 2086 in human subjects. *Clin Pharmacol Ther,* 45**,** 270-6.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A. & Blankenberg, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res,* 46**,** W537-w544.

Agrawal, A. & Murphy, T. F. 2011. Haemophilus influenzae infections in the H. influenzae type b conjugate vaccine era. *Journal of clinical microbiology,* 49**,** 3728-3732.

Albelda, S. M., Smith, C. W. & Ward, P. A. 1994. Adhesion molecules and inflammatory injury. *Faseb j,* 8**,** 504-12.

Amaral, A. F. S., Patel, J., Kato, B. S., Obaseki, D. O., Lawin, H., Tan, W. C., Juvekar, S. K., Harrabi, I., Studnicka, M., Wouters, E. F. M., Loh, L. C., Bateman, E. D., Mortimer, K., Buist, A. S. & Burney, P. G. J. 2018. Airflow Obstruction and Use of Solid Fuels for Cooking or Heating: BOLD Results. *Am J Respir Crit Care Med,* 197**,** 595-610.

Aponte, M., Jiang, W., Lakkis, M., Li, M.-J., Edwards, D., Albitar, L., Vitonis, A., Mok, S. C., Cramer, D. W. & Ye, B. 2008. Activation of platelet-activating factor receptor and pleiotropic effects on tyrosine phospho-EGFR/Src/FAK/paxillin in ovarian cancer. *Cancer Research,* 68**,** 5839-5848.

Arnout, J., Van Hecken, A., De Lepeleire, I., Miyamoto, Y., Holmes, I., De Schepper, P. & Vermylen, J. 1988. Effectiveness and tolerability of CV-3988, a selective PAF antagonist, after intravenous administration to man. *British journal of clinical pharmacology,* 25**,** 445-451.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

Assad, N. A., Balmes, J., Mehta, S., Cheema, U. & Sood, A. 2015. Chronic obstructive pulmonary disease secondary to household air pollution. *Semin Respir Crit Care Med,* 36**,** 408-21.

Assad, N. A., Kapoor, V. & Sood, A. 2016. Biomass smoke exposure and chronic lung disease. *Current Opinion in Pulmonary Medicine,* 22**,** 150-157.

Atack, J. M., Murphy, T. F., Bakaletz, L. O., Seib, K. L. & Jennings, M. P. 2018. Closed Complete Genome Sequences of Two Nontypeable Haemophilus influenzae Strains Containing Novel modA Alleles from the Sputum of Patients with Chronic Obstructive Pulmonary Disease. *Microbiol Resour Announc,* 7.

Attfield, M. D. 1985. Longitudinal decline in FEV1 in United States coalminers. *Thorax,* 40**,** 132-7.

Aurelio, L., Valant, C., Flynn, B. L., Sexton, P. M., White, J. M., Christopoulos, A. & Scammells, P. J. 2010. Effects of conformational restriction of 2-amino-3-benzoylthiophenes on A(1) adenosine receptor modulation. *J Med Chem,* 53**,** 6550-9.

Avadhanula, V., Rodriguez, C. A., Ulett, G. C., Bakaletz, L. O. & Adderson, E. E. 2006. Nontypeable Haemophilus influenzae adheres to intercellular adhesion molecule 1 (ICAM-1) on respiratory epithelial cells and upregulates ICAM-1 expression. *Infect Immun,* 74**,** 830-8.

Aziz, A., Sarovich, D. S., Nosworthy, E., Beissbarth, J., Chang, A. B., Smith-Vaughan, H., Price, E. P. & Harris, T. M. 2019. Molecular Signatures of Non-typeable Haemophilus influenzae Lung Adaptation in Pediatric Chronic Lung Disease. *Front Microbiol,* 10**,** 1622.

Bakaletz, L. O. & Novotny, L. A. 2018. Nontypeable Haemophilus influenzae (NTHi). *Trends Microbiol,* 26**,** 727-728.

Balmes, J. R. & Eisen, E. A. 2018. Household Air Pollution and Chronic Obstructive Pulmonary Disease. "A Riddle, Wrapped in a Mystery, Inside an Enigma". *Am J Respir Crit Care Med,* 197**,** 547-549.

Banerjee, A., Mondal, N. K., Das, D. & Ray, M. R. 2012. Neutrophilic inflammatory response and oxidative stress in premenopausal women chronically exposed to indoor air pollution from biomass burning. *Inflammation,* 35**,** 671-83.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. & Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol,* 19**,** 455-77.

Barbier, M., Oliver, A., Rao, J., Hanna, S. L., Goldberg, J. B. & Alberti, S. 2008. Novel phosphorylcholine-containing protein of Pseudomonas aeruginosa chronic infection isolates interacts with airway epithelial cells. *J Infect Dis,* 197**,** 465-73.

Barczyk, M., Carracedo, S. & Gullberg, D. 2010. Integrins. *Cell Tissue Res,* 339**,** 269-80.

Barenkamp, S. J., Munson, R. S. & Granoff, D. 1982. Outer membrane protein and biotype analysis of pathogenic nontypable Haemophilus influenzae. *Infection and immunity,* 36**,** 535-540.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E. & Kim, J. F. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature,* 461**,** 1243-7.

Baylar, A. & Ozkan, F. 2006. Applications of venturi principle to water aeration systems. *Environmental Fluid Mechanics,* 6**,** 341-357.

Beasley, V., Joshi, P. V., Singanayagam, A., Molyneaux, P. L., Johnston, S. L. & Mallia, P. 2012. Lung microbiology and exacerbations in COPD. *International journal of chronic obstructive pulmonary disease,* 7**,** 555.

Beauchemin, N., Draber, P., Dveksler, G., Gold, P., Gray-Owen, S., Grunert, F., Hammarstrom, S., Holmes, K. V., Karlsson, A., Kuroki, M., Lin, S. H., Lucka, L., Najjar, S. M., Neumaier, M., Obrink, B., Shively, J. E., Skubitz, K. M., Stanners, C. P., Thomas, P., Thompson, J. A., Virji, M., Von Kleist, S., Wagener, C., Watt, S. & Zimmermann, W. 1999. Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res,* 252**,** 243-9.

Behera, D. & Jindal, S. K. 1991. Respiratory symptoms in Indian women using domestic cooking fuels. *Chest,* 100**,** 385-8.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological),* 57**,** 289-300.

Berglund, D. J., Abbey, D. E., Lebowitz, M. D., Knutsen, S. F. & Mc Donnell, W. F. 1999. Respiratory symptoms and pulmonary function in an elderly nonsmoking population. *Chest,* 115**,** 49-59.

Bodor, F. F., Marchant, C. D., Shurin, P. A. & Barenkamp, S. J. 1985. Bacterial etiology of conjunctivitis-otitis media syndrome. *Pediatrics,* 76**,** 26-28.

Boschetto, P., Quintavalle, S., Miotto, D., Lo Cascio, N., Zeni, E. & Mapp, C. E. 2006. Chronic obstructive pulmonary disease (COPD) and occupational exposures. *J Occup Med Toxicol,* 1**,** 11.

Bradshaw, L. M., Fishwick, D., Slater, T. & Pearce, N. 1998. Chronic bronchitis, work related respiratory symptoms, and pulmonary function in welders in New Zealand. *Occup Environ Med,* 55**,** 150-4.

Braquet, P. G. 1997. Platelet-activating factor and its antagonists: Scientific background and clinical applications of ginkgolides. *Ginkgo Biloba A Global Treasure.* Springer.

Brenner, D., Mayer, L., Carlone, G., Harrison, L., Bibb, W., Brandileone, M., Sottnek, F., Irino, K., Reeves, M. & Swenson, J. 1988. Biochemical, genetic, and epidemiologic characterization of Haemophilus influenzae biogroup aegyptius (Haemophilus aegyptius) strains associated with Brazilian purpuric fever. *Journal of clinical microbiology,* 26**,** 1524-1534.

Bresser, P., Out, T. A., Van Alphen, L., Jansen, H. M. & Lutter, R. 2000. Airway inflammation in nonobstructive and obstructive chronic bronchitis with chronic haemophilus influenzae airway infection. Comparison with noninfected patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 162**,** 947-52.

Bresser, P., Van Alphen, L., Habets, F. J., Hart, A. A., Dankert, J., Jansen, H. M. & Lutter, R. 1997. Persisting Haemophilus influenzae strains induce lower levels of interleukin-6 and interleukin-8 in H292 lung epithelial cells than nonpersisting strains. *Eur Respir J,* 10**,** 2319-26.

Brook, I. & Gober, A. E. 2005. Recovery of potential pathogens and interfering bacteria in the nasopharynx of smokers and nonsmokers. *Chest,* 127**,** 2072-5.

Brook, I. & Gober, A. E. 2007. Effect of smoking cessation on the microbial flora. *Arch Otolaryngol Head Neck Surg,* 133**,** 135-8.

Bruce, K. D. & Jordens, J. 1991. Characterization of noncapsulate Haemophilus influenzae by whole-cell polypeptide profiles, restriction endonuclease analysis, and rRNA gene restriction patterns. *Journal of clinical microbiology,* 29**,** 291-296.

Bruce, N., Dherani, M., Liu, R., Hosgood, H. D., 3rd, Sapkota, A., Smith, K. R., Straif, K., Lan, Q. & Pope, D. 2015. Does household use of biomass fuel cause lung cancer? A systematic review and evaluation of the evidence for the GBD 2010 study. *Thorax,* 70**,** 433-41.

Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology,* 17**,** 238.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T. & Modrich, P. 2001. In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences,* 98**,** 6765-6770.

Camp, P. G., Ramirez-Venegas, A., Sansores, R. H., Alva, L. F., Mcdougall, J. E., Sin, D. D., Pare, P. D., Muller, N. L., Silva, C. I., Rojas, C. E. & Coxson, H. O. 2014. COPD phenotypes in biomass smoke- versus tobacco smoke-exposed Mexican women. *Eur Respir J,* 43**,** 725-34.

Capistrano, S. J., Van Reyk, D., Chen, H. & Oliver, B. G. 2017. Evidence of Biomass Smoke Exposure as a Causative Factor for the Development of COPD. *Toxics,* 5.

Carruthers, M. D., Tracy, E. N., Dickson, A. C., Ganser, K. B., Munson, R. S., Jr. & Bakaletz, L. O. 2012. Biological roles of nontypeable Haemophilus influenzae type IV pilus proteins encoded by the pil and com operons. *J Bacteriol,* 194**,** 1927-33.

Casals-Stenzel, J. 1987. Triazolodiazepines are potent antagonists of platelet activating factor (PAF) in vitro and in vivo. *Naunyn Schmiedebergs Arch Pharmacol,* 335**,** 351-5.

Casals-Stenzel, J. 1991. Thieno-triazolo-1,4-diazepines as antagonists of platelet-activating factor: present status. *Lipids,* 26**,** 1157-61.

Castell, J. V., Gomez-Lechon, M. J., David, M., Andus, T., Geiger, T., Trullenque, R., Fabra, R. & Heinrich, P. C. 1989. Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. *FEBS Lett,* 242**,** 237-9.

Cellai, C., Laurenzana, A., Vannucchi, A., Caporale, R., Paglierani, M., Di Lollo, S., Pancrazzi, A. & Paoletti, F. 2006. Growth inhibition and differentiation of human breast cancer cells by the PAFR antagonist WEB-2086. *British journal of cancer,* 94**,** 1637-1642.

Cerveri, I., Accordini, S., Verlato, G., Corsico, A., Zoia, M. C., Casali, L., Burney, P. & De Marco, R. 2001. Variations in the prevalence across countries of chronic bronchitis and smoking habits in young adults. *Eur Respir J,* 18**,** 85-92.

Chagnot, C., Listrat, A., Astruc, T. & Desvaux, M. 2012. Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components. *Cell Microbiol,* 14**,** 1687-96.

Chan, K. Y., Li, X., Chen, W., Song, P., Wong, N. W. K., Poon, A. N., Jian, W., Soyiri, I. N., Cousens, S., Adeloye, D., Sheikh, A., Campbell, H. & Rudan, I. 2017. Prevalence of chronic obstructive pulmonary disease (COPD) in China in 1990 and 2010. *J Glob Health,* 7**,** 020704.

Chan, R. D., Greenstein, S. M., Sablay, L., Alfonso, F., Tellis, V., Spitzer, A., Greifer, I. & Corey, H. E. 1995. Analysis of adhesion molecule expression by tubular epithelial cells using urine immunocytology. *Acta Cytol,* 39**,** 435-42.

Chen, I. & Dubnau, D. 2004. DNA uptake during bacterial transformation. *Nat Rev Microbiol,* 2**,** 241-9.

Chen, J., Lan, T., Zhang, W., Dong, L., Kang, N., Zhang, S., Fu, M., Liu, B., Liu, K. & Zhan, Q. 2015a. Feed-Forward Reciprocal Activation of PAFR and STAT3 Regulates Epithelial-Mesenchymal Transition in Non-Small Cell Lung Cancer. *Cancer Res,* 75**,** 4198-210.

Chen, J., Lan, T., Zhang, W., Dong, L., Kang, N., Zhang, S., Fu, M., Liu, B., Liu, K. & Zhang, C. 2015b. Platelet-activating factor receptor-mediated PI3K/AKT activation contributes to the malignant development of esophageal squamous cell carcinoma. *Oncogene,* 34**,** 5114-5127.

Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., Denning, G. M., Murphy, T. F., Sethi, S. & Look, D. C. 2005. Haemophilus influenzae from patients with chronic obstructive pulmonary disease exacerbation induce more inflammation than colonizers. *Am J Respir Crit Care Med,* 172**,** 85-91.

Choi, H. & Koh, S.-H. 2018. Understanding the role of glycogen synthase kinase-3 in L-DOPA-induced dyskinesia in Parkinson's disease. *Expert opinion on drug metabolism & toxicology,* 14**,** 83-90.

Chung, K., Mccusker, M., Page, C., Dent, G., Guinot, P. & Barnes, P. 1987. Effect of a ginkgolide mixture (BN 52063) in antagonising skin and platelet responses to platelet activating factor in man. *The Lancet,* 329**,** 248-251.

Chung, K. F. & Barnes, P. J. 1991. Role for platelet-activating factor in asthma. *Lipids,* 26**,** 1277-9.

Chung, S., Ban, S., Kim, S., Kim, B., Woo, S., Summers, J. B. & Conway, R. G. 1995. Synthesis and bioactivities of heterocyclic lipids as PAF antagonists. 2. *Bioorganic & Medicinal Chemistry Letters,* 5**,** 1097-1102.

Churg, A., Hobson, J. & Wright, J. 1989. Functional and morphologic comparison of silica- and elastase-induced airflow obstruction. *Exp Lung Res,* 15**,** 813-22.

Clementi, C. F. & Murphy, T. F. 2011. Non-typeable Haemophilus influenzae invasion and persistence in the human respiratory tract. *Front Cell Infect Microbiol,* 1**,** 1.

Cohen, B. H., Diamond, E. L., Graves, C. G., Kreiss, P., Levy, D. A., Menkes, H. A., Permutt, S., Quaskey, S. & Tockman, M. S. 1977. A common familial component in lung cancer and chronic obstructive pulmonary disease. *Lancet,* 2**,** 523-6.

Craig, J. E., Cliffe, A., Garnett, K. & High, N. J. 2001. Survival of nontypeable Haemophilus influenzae in macrophages. *FEMS microbiology letters,* 203**,** 55-61.

Crystal, R. G., Randell, S. H., Engelhardt, J. F., Voynow, J. & Sunday, M. E. 2008. Airway epithelial cells: current concepts and challenges. *Proc Am Thorac Soc,* 5**,** 772-7.

Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkila, I. & Tuomanen, E. I. 1995. Streptococcus pneumoniae anchor to activated human cells by the receptor for platelet-activating factor. *Nature,* 377**,** 435-8.

Da Silva Junior, I. A., De Sousa Andrade, L. N., Jancar, S. & Chammas, R. 2018. Platelet activating factor receptor antagonists improve the efficacy of experimental chemo- and radiotherapy. *Clinics (Sao Paulo),* 73**,** e792s.

Da Silva, L. F., Saldiva, S. R., Saldiva, P. H. & Dolhnikoff, M. 2012. Impaired lung function in individuals chronically exposed to biomass combustion. *Environ Res,* 112**,** 111-7.

De Chiara, M., Hood, D., Muzzi, A., Pickard, D. J., Perkins, T., Pizza, M., Dougan, G., Rappuoli, R., Moxon, E. R., Soriani, M. & Donati, C. 2014. Genome sequencing of disease and carriage isolates of

nontypeable Haemophilus influenzae identifies discrete population structure. *Proc Natl Acad Sci U S A,* 111**,** 5439-44.

Demaria, T. F., Lim, D. J., Barnishan, J., Ayers, L. W. & Birck, H. G. 1984. Biotypes of serologically nontypable Haemophilus influenzae isolated from the middle ears and nasopharynges of patients with otitis media with effusion. *J Clin Microbiol,* 20**,** 1102-4.

Demeo, D. L., Mariani, T. J., Lange, C., Srisuma, S., Litonjua, A. A., Celedon, J. C., Lake, S. L., Reilly, J. J., Chapman, H. A., Mecham, B. H., Haley, K. J., Sylvia, J. S., Sparrow, D., Spira, A. E., Beane, J., Pinto-Plata, V., Speizer, F. E., Shapiro, S. D., Weiss, S. T. & Silverman, E. K. 2006. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum Genet,* 78**,** 253-64.

Dewan, A., Liu, M., Hartman, S., Zhang, S. S., Liu, D. T., Zhao, C., Tam, P. O., Chan, W. M., Lam, D. S., Snyder, M., Barnstable, C., Pang, C. P. & Hoh, J. 2006. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science,* 314**,** 989-92.

Didelot, X., Meric, G., Falush, D. & Darling, A. E. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. *BMC Genomics,* 13**,** 256.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. 2016. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol,* 14**,** 150-62.

Donaldson, K., Brown, D., Clouter, A., Duffin, R., Macnee, W., Renwick, L., Tran, L. & Stone, V. 2002. The pulmonary toxicology of ultrafine particles. *J Aerosol Med,* 15**,** 213-20.

Donati, C., Hiller, N. L., Tettelin, H., Muzzi, A., Croucher, N. J., Angiuoli, S. V., Oggioni, M., Dunning Hotopp, J. C., Hu, F. Z., Riley, D. R., Covacci, A., Mitchell, T. J., Bentley, S. D., Kilian, M., Ehrlich, G. D., Rappuoli, R., Moxon, E. R. & Masignani, V. 2010. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome Biol,* 11**,** R107.

Duell, B. L., Su, Y. C. & Riesbeck, K. 2016. Host-pathogen interactions of nontypeable Haemophilus influenzae: from commensal to pathogen. *FEBS Lett,* 590**,** 3840-3853.

Dutta, A., Ray, M. R. & Banerjee, A. 2012. Systemic inflammatory changes and increased oxidative stress in rural Indian women cooking with biomass fuels. *Toxicol Appl Pharmacol,* 261**,** 255-62.

Dutta, A., Roychoudhury, S., Chowdhury, S. & Ray, M. R. 2013. Changes in sputum cytology, airway inflammation and oxidative stress due to chronic inhalation of biomass smoke during cooking in premenopausal rural Indian women. *Int J Hyg Environ Health,* 216**,** 301-8.

Dutta, C. & Pan, A. 2002. Horizontal gene transfer and bacterial diversity. *Journal of biosciences,* 27**,** 27-33.

Eberhard, T., Virkola, R., Korhonen, T., Kronvall, G. & Ullberg, M. 1998. Binding to human extracellular matrix by Neisseria meningitidis. *Infect Immun,* 66**,** 1791-4.

Ecevit, I. Z., Mccrea, K. W., Pettigrew, M. M., Sen, A., Marrs, C. F. & Gilsdorf, J. R. 2004. Prevalence of the hifBC, hmw1A, hmw2A, hmwC, and hia Genes in Haemophilus influenzae Isolates. *J Clin Microbiol,* 42**,** 3065-72.

Ehrlich, R. I., White, N., Norman, R., Laubscher, R., Steyn, K., Lombard, C. & Bradshaw, D. 2004. Predictors of chronic bronchitis in South African adults. *Int J Tuberc Lung Dis,* 8**,** 369-76.

Eisner, M. D., Anthonisen, N., Coultas, D., Kuenzli, N., Perez-Padilla, R., Postma, D., Romieu, I., Silverman, E. K. & Balmes, J. R. 2010. An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 182**,** 693-718.

Elhenawy, W., Tsai, C. N. & Coombes, B. K. 2019. Host-Specific Adaptive Diversification of Crohn's Disease-Associated Adherent-Invasive Escherichia coli. *Cell Host Microbe,* 25**,** 301-312.e5.

Erwin, A. L., Nelson, K. L., Mhlanga-Mutangadura, T., Bonthuis, P. J., Geelhood, J. L., Morlin, G., Unrath, W. C., Campos, J., Crook, D. W., Farley, M. M., Henderson, F. W., Jacobs, R. F., Muhlemann, K., Satola, S. W., Van Alphen, L., Golomb, M. & Smith, A. L. 2005. Characterization of genetic and phenotypic diversity of invasive nontypeable Haemophilus influenzae. *Infect Immun,* 73**,** 5853-63.

Erwin, A. L., Sandstedt, S. A., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., Unrath, W. C., Diggle, M. A., Theodore, M. J., Pleatman, C. R., Mothershed, E. A., Sacchi, C. T., Mayer, L. W., Gilsdorf, J. R. & Smith, A. L. 2008. Analysis of genetic relatedness of Haemophilus influenzae isolates by multilocus sequence typing. *J Bacteriol,* 190**,** 1473-83.

Erwin, A. L. & Smith, A. L. 2007. Nontypeable Haemophilus influenzae: understanding virulence and commensal behavior. *Trends in microbiology,* 15**,** 355-362.

Evans, D. J., Barnes, P. J., Cluzel, M. & O'connor, B. J. 1997. Effects of a potent platelet-activating factor antagonist, SR27417A, on allergen-induced asthmatic responses. *Am J Respir Crit Care Med,* 156**,** 11-6.

Ezzati, M. & Kammen, D. 2001. Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. *Lancet,* 358**,** 619-24.

Faizan, M. A. & Thakur, R. 2019. Association Between Solid Cooking Fuels and Respiratory Disease Across Socio-Demographic Groups in India. *J Health Pollut,* 9**,** 190911.

Farjo, R. S., Foxman, B., Patel, M. J., Zhang, L., Pettigrew, M. M., Mccoy, S. I., Marrs, C. F. & Gilsdorf, J. R. 2004. Diversity and sharing of Haemophilus influenzae strains colonizing healthy children attending day-care centers. *The Pediatric infectious disease journal,* 23**,** 41-46.

Farr, R. S., Cox, C. P., Wardlow, M. L. & Jorgensen, R. 1980. Preliminary studies of an acid-labile factor (ALF) in human sera that inactivates platelet-activating factor (PAF). *Clinical immunology and immunopathology,* 15**,** 318-330.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol,* 186**,** 1518-30.

Feldman, C. & Anderson, R. 2013. Cigarette smoking and mechanisms of susceptibility to infections of the respiratory tract and other organ systems. *J Infect,* 67**,** 169-84.

Fernaays, M. M., Lesse, A. J., Sethi, S., Cai, X. & Murphy, T. F. 2006. Differential genome contents of nontypeable Haemophilus influenzae strains from adults with chronic obstructive pulmonary disease. *Infect Immun,* 74**,** 3366-74.

Fernandes, L., Gulati, N., Fernandes, Y., Mesquita, A. M., Sardessai, M., Lammers, J. J., Mohamed Hoesein, F. A., Ten Hacken, N. H. T., Van Den Berge, M., Galban, C. J. & Siddiqui, S. 2017. Small airway imaging phenotypes in biomass- and tobacco smoke-exposed patients with COPD. *ERJ Open Res,* 3.

Fink, D. L., Green, B. A. & Geme Iii, J. W. S. 2002. The Haemophilus influenzae Hap autotransporter binds to fibronectin, laminin, and collagen IV. *Infect Immun,* 70**,** 4902-4907.

Fischer, M., Hedberg, K., Cardosi, P., Plikaytis, B. D., Hoesly, F. C., Steingart, K. R., Bell, T. A., Fleming, D. W., Wenger, J. D. & Perkins, B. A. 1997. Tobacco smoke as a risk factor for meningococcal disease. *The Pediatric infectious disease journal,* 16**,** 979-983.

Fletcher, C. 1960. Standardised questionnaire on respiratory symptoms: a statement prepared and approved by the MRC Committee on the Aetiology of Chronic Bronchitis (MRC breathlessness score). *Bmj,* 2**,** 1665.

Foxwell, A. R., Kyd, J. M. & Cripps, A. W. 1998. Nontypeable Haemophilus influenzae: pathogenesis and prevention. *Microbiol Mol Biol Rev,* 62**,** 294-308.

Francisco, A. P., Bugalho, M., Ramirez, M. & Carrico, J. A. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics,* 10**,** 152.

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M. & Carrico, J. A. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics,* 13**,** 87.

Frick, A. G., Joseph, T. D., Pang, L., Rabe, A. M., St Geme, J. W., 3rd & Look, D. C. 2000. Haemophilus influenzae stimulates ICAM-1 expression on respiratory epithelial cells. *J Immunol,* 164**,** 4185-96.

Gabay, C. 2006. Interleukin-6 and chronic inflammation. *Arthritis Res Ther,* 8 Suppl 2**,** S3.

Galardini, M. 2013. Roary_plots. Github.

Gantt, R. C. & Lincoln, J. E. 1988. Determinants of chronic bronchitis and lung dysfunction in Western Australian gold miners. *Br J Ind Med,* 45**,** 503.

Gauderman, W. J., Vora, H., Mcconnell, R., Berhane, K., Gilliland, F., Thomas, D., Lurmann, F., Avol, E., Kunzli, N., Jerrett, M. & Peters, J. 2007. Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. *Lancet,* 369**,** 571-7.

Gautam, S. S., Rajendra, K., Leong, K. W., Mac Aogáin, M. & O'toole, R. F. 2019. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *Journal of Biological Methods,* 6**,** e110.

Geelen, S., Bhattacharyya, C. & Tuomanen, E. 1993. The cell wall mediates pneumococcal attachment to and cytopathology in human endothelial cells. *Infect Immun,* 61**,** 1538-43.

Global Initiative for Chronic Obstructive Lung Disease. 2018. *Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease 2018 Report* [Online]. Fontana, WI, USA: Global Initiative for Chronic Obstructive Lung Disease. Available: https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf [Accessed 23/03/2020].

Gomez, F. P., Roca, J., Barbera, J. A., Chung, K. F., Peinado, V. I. & Rodriguez-Roisin, R. 1998. Effect of a platelet-activating factor (PAF) antagonist, SR 27417A, on PAF-induced gas exchange abnormalities in mild asthma. *Eur Respir J,* 11**,** 835-9.

Goodwin, S., Mcpherson, J. D. & Mccombie, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-51.

Gordon, S. B., Bruce, N. G., Grigg, J., Hibberd, P. L., Kurmi, O. P., Lam, K. B., Mortimer, K., Asante, K. P., Balakrishnan, K., Balmes, J., Bar-Zeev, N., Bates, M. N., Breysse, P. N., Buist, S., Chen, Z., Havens, D., Jack, D., Jindal, S., Kan, H., Mehta, S., Moschovis, P., Naeher, L., Patel, A., Perez-Padilla, R., Pope, D., Rylance, J., Semple, S. & Martin, W. J., 2nd 2014. Respiratory risks from household air pollution in low and middle income countries. *Lancet Respir Med,* 2**,** 823-60.

Granstrom, B. W., Xu, C. B., Nilsson, E., Vikman, P. & Edvinsson, L. 2006. Smoking particles enhance endothelin A and endothelin B receptor-mediated contractions by enhancing translation in rat bronchi. *BMC Pulm Med,* 6**,** 6.

Grigg, J. 2012. The platelet activating factor receptor: a new anti-infective target in respiratory disease? *Thorax,* 67**,** 840-1.

Grigg, J., Miyashita, L. & Suri, R. 2017. Pneumococcal infection of respiratory cells exposed to welding fumes; Role of oxidative stress and HIF-1 alpha. *PLoS One,* 12**,** e0173569.

Grigg, J., Walters, H., Sohal, S. S., Wood-Baker, R., Reid, D. W., Xu, C. B., Edvinsson, L., Morissette, M. C., Stampfli, M. R., Kirwan, M., Koh, L., Suri, R. & Mushtaq, N. 2012. Cigarette smoke and platelet-activating factor receptor dependent adhesion of Streptococcus pneumoniae to lower airway cells. *Thorax,* 67**,** 908-13.

Gruenert, D. C., Finkbeiner, W. E. & Widdicombe, J. H. 1995. Culture and transformation of human airway epithelial cells. *Am J Physiol,* 268**,** L347-60.

Gunen, H., Hacievliyagil, S. S., Yetkin, O., Gulbas, G., Mutlu, L. C. & Pehlivan, E. 2008. Prevalence of COPD: first epidemiological study of a large region in Turkey. *Eur J Intern Med,* 19**,** 499-504.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-5.

Hagstad, S., Backman, H., Bjerg, A., Ekerljung, L., Ye, X., Hedman, L., Lindberg, A., Toren, K., Lotvall, J., Ronmark, E. & Lundback, B. 2015. Prevalence and risk factors of COPD among never-smokers in two areas of Sweden - Occupational exposure to gas, dust or fumes is an important risk factor. *Respir Med,* 109**,** 1439-45.

Hall, D. O., Barnard, G. W. & Moss, P. 2013. *Biomass for energy in the developing countries: current role, potential, problems, prospects*, Elsevier.

Hamzah, N. A., Mohd Tamrin, S. B. & Ismail, N. H. 2016. Metal dust exposure and lung function deterioration among steel workers: an exposure-response relationship. *Int J Occup Environ Health,* 22**,** 224-232.

Han, M. K., Muellerova, H., Curran-Everett, D., Dransfield, M. T., Washko, G. R., Regan, E. A., Bowler, R. P., Beaty, T. H., Hokanson, J. E., Lynch, D. A., Jones, P. W., Anzueto, A., Martinez, F. J., Crapo, J. D., Silverman, E. K. & Make, B. J. 2013. GOLD 2011 disease severity classification in COPDGene: a prospective cohort study. *Lancet Respir Med,* 1**,** 43-50.

Hancock, D. B., Eijgelsheim, M., Wilk, J. B., Gharib, S. A., Loehr, L. R., Marciante, K. D., Franceschini, N., Van Durme, Y. M., Chen, T. H., Barr, R. G., Schabath, M. B., Couper, D. J., Brusselle, G. G., Psaty, B. M., Van Duijn, C. M., Rotter, J. I., Uitterlinden, A. G., Hofman, A., Punjabi, N. M., Rivadeneira, F., Morrison, A. C., Enright, P. L., North, K. E., Heckbert, S. R., Lumley, T., Stricker, B. H., O'connor, G. T. & London, S. J. 2010. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet,* 42**,** 45-52.

Hardy, G. G., Tudor, S. M. & St Geme, J. W., 3rd 2003. The pathogenesis of disease due to nontypeable Haemophilus influenzae. *Methods Mol Med,* 71**,** 1-28.

Harper, E. 2013. Lung foundation Australia: promoting lung health and supporting those with lung disease. *J Thorac Dis,* 5**,** 572-7.

Harrison, A., Dyer, D. W., Gillaspy, A., Ray, W. C., Mungur, R., Carson, M. B., Zhong, H., Gipson, J., Gipson, M., Johnson, L. S., Lewis, L., Bakaletz, L. O. & Munson, R. S., Jr. 2005. Genomic sequence of an otitis media isolate of nontypeable Haemophilus influenzae: comparative study with H. influenzae serotype d, strain KW20. *J Bacteriol,* 187**,** 4627-36.

Harrison, A., Hardison, R. L., Fullen, A. R., Wallace, R. M., Gordon, D. M., White, P., Jennings, R. N., Justice, S. S. & Mason, K. M. 2020. Continuous Microevolution Accelerates Disease Progression during Sequential Episodes of Infection. *Cell Rep,* 30**,** 2978-2988.e3.

Hart, J. E., Laden, F., Schenker, M. B. & Garshick, E. 2006. Chronic obstructive pulmonary disease mortality in diesel-exposed railroad workers. *Environ Health Perspect,* 114**,** 1013-7.

Hauck, C. R. 2002. Cell adhesion receptors - signaling capacity and exploitation by bacterial pathogens. *Med Microbiol Immunol,* 191**,** 55-62.

Hausdorff, W. P., Feikin, D. R. & Klugman, K. P. 2005. Epidemiological differences among pneumococcal serotypes. *Lancet Infect Dis,* 5**,** 83-93.

Hergott, C. B., Roche, A. M., Naidu, N. A., Mesaros, C., Blair, I. A. & Weiser, J. N. 2015. Bacterial exploitation of phosphorylcholine mimicry suppresses inflammation to promote airway infection. *J Clin Invest,* 125**,** 3878-90.

Herr, C., Beisswenger, C., Hess, C., Kandler, K., Suttorp, N., Welte, T., Schroeder, J. M. & Vogelmeier, C. 2009. Suppression of pulmonary innate host defence in smokers. *Thorax,* 64**,** 144-9.

Hewitt, R., Farne, H., Ritchie, A., Luke, E., Johnston, S. L. & Mallia, P. 2016. The role of viral infections in exacerbations of chronic obstructive pulmonary disease and asthma. *Therapeutic advances in respiratory disease,* 10**,** 158-174.

Higgins, M. & Keller, J. 1975. Familial occurrence of chronic respiratory disease and familial resemblance in ventilatory capacity. *J Chronic Dis,* 28**,** 239-51.

Hill, A. T., Campbell, E. J., Hill, S. L., Bayley, D. L. & Stockley, R. A. 2000. Association between airway bacterial load and markers of airway inflammation in patients with stable chronic bronchitis. *Am J Med,* 109**,** 288-95.

Hirouchi, M., Mizutani, H., Kohno, Y. & Kuriyama, K. 1992. Characteristics of the association of brotizolam, a thieno-triazolo diazepine derivative, with the benzodiazepine receptor: a selective and high affinity ligand of the central type I benzodiazepine receptor. *Jpn J Pharmacol,* 59**,** 387-91.

Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J. C. & Ehrlich, G. D. 2007. Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol,* 8**,** R103.

Holguin, F., Flores, S., Ross, Z., Cortez, M., Molina, M., Molina, L., Rincon, C., Jerrett, M., Berhane, K., Granados, A. & Romieu, I. 2007. Traffic-related exposures, airway function, inflammation, and respiratory symptoms in children. *Am J Respir Crit Care Med,* 176**,** 1236-42.

Honda, Z.-I., Nakamura, M., Miki, I., Minami, M., Watanabe, T., Seyama, Y., Okado, H., Toh, H., Ito, K. & Miyamoto, T. 1991. Cloning by functional expression of platelet-activating factor receptor from guinea-pig lung. *Nature,* 349**,** 342-346.

Hozawa, S., Haruta, Y., Ishioka, S. & Yamakido, M. 1995. Effects of a PAF antagonist, Y-24180, on bronchial hyperresponsiveness in patients with asthma. *Am J Respir Crit Care Med,* 152**,** 1198-202.

Hsieh, K. H. 1991. Effects of PAF antagonist, BN52021, on the PAF-, methacholine-, and allergen-induced bronchoconstriction in asthmatic children. *Chest,* 99**,** 877-82.

Hu, G., Zhou, Y., Hong, W., Tian, J., Hu, J., Peng, G., Cui, J., Li, B. & Ran, P. 2013. Development and systematic oxidative stress of a rat model of chronic bronchitis and emphysema induced by biomass smoke. *Exp Lung Res,* 39**,** 229-40.

Hu, G., Zhou, Y., Tian, J., Yao, W., Li, J., Li, B. & Ran, P. 2010. Risk of COPD from exposure to biomass smoke: a metaanalysis. *Chest,* 138**,** 20-31.

Huang, G. T., Eckmann, L., Savidge, T. C. & Kagnoff, M. F. 1996. Infection of human intestinal epithelial cells with invasive bacteria upregulates apical intercellular adhesion molecule-1 (ICAM)-1) expression and neutrophil adhesion. *J Clin Invest,* 98**,** 572-83.

Huang, L., Pu, J., He, F., Liao, B., Hao, B., Hong, W., Ye, X., Chen, J., Zhao, J., Liu, S., Xu, J., Li, B. & Ran, P. 2017. Positive feedback of the amphiregulin-EGFR-ERK pathway mediates PM2.5 from wood smoke-induced MUC5AC expression in epithelial cells. *Sci Rep,* 7**,** 11084.

Huffnagle, G. B., Dickson, R. P. & Lukacs, N. W. 2017. The respiratory tract microbiome and lung inflammation: a two-way street. *Mucosal Immunol,* 10**,** 299-306.

Hyland, I. K., O'toole, R. F., Smith, J. A. & Bissember, A. C. 2018. Progress in the Development of Platelet-Activating Factor Receptor (PAFr) Antagonists and Applications in the Treatment of Inflammatory Diseases. *ChemMedChem,* 13**,** 1873-1884.

Iovino, F., Brouwer, M. C., Van De Beek, D., Molema, G. & Bijlsma, J. J. 2013. Signalling or binding: the role of the platelet-activating factor receptor in invasive pneumococcal disease. *Cell Microbiol,* 15**,** 870-81.

Iovino, F., Seinen, J., Henriques-Normark, B. & Van Dijl, J. M. 2016. How Does Streptococcus pneumoniae Invade the Brain? *Trends Microbiol,* 24**,** 307-15.

Ishii, S., Nagase, T. & Shimizu, T. 2002. Platelet-activating factor receptor. *Prostaglandins Other Lipid Mediat,* 68-69**,** 599-609.

Ishii, S., Nagase, T., Shindou, H., Takizawa, H., Ouchi, Y. & Shimizu, T. 2004. Platelet-activating factor receptor develops airway hyperresponsiveness independently of airway inflammation in a murine asthma model. *J Immunol,* 172**,** 7095-102.

Ishizuka, S., Yamaya, M., Suzuki, T., Nakayama, K., Kamanaka, M., Ida, S., Sekizawa, K. & Sasaki, H. 2001. Acid exposure stimulates the adherence of Streptococcus pneumoniae to cultured human airway epithelial cells: effects on platelet-activating factor receptor expression. *Am J Respir Cell Mol Biol,* 24**,** 459-68.

Jackson, M. D., Wong, S. M. & Akerley, B. J. 2019. Underlying Glycans Determine the Ability of Sialylated Lipooligosaccharide To Protect Nontypeable Haemophilus influenzae from Serum IgM and Complement. *Infect Immun,* 87.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J. & Abdelalim, A. 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet,* 392**,** 1789-1858.

Janssen, N. A., Brunekreef, B., Van Vliet, P., Aarts, F., Meliefste, K., Harssema, H. & Fischer, P. 2003. The relationship between air pollution from heavy traffic and allergic sensitization, bronchial hyperresponsiveness, and respiratory symptoms in Dutch schoolchildren. *Environ Health Perspect,* 111**,** 1512-8.

Jayes, L., Haslam, P. L., Gratziou, C. G., Powell, P., Britton, J., Vardavas, C., Jimenez-Ruiz, C. & Leonardi-Bee, J. 2016. SmokeHaz: Systematic Reviews and Meta-analyses of the Effects of Smoking on Respiratory Health. *Chest,* 150**,** 164-79.

Ji, W., Chen, J., Mi, Y., Wang, G., Xu, X. & Wang, W. 2016. Platelet-activating factor receptor activation promotes prostate cancer cell growth, invasion and metastasis via ERK1/2 pathway. *Int J Oncol,* 49**,** 181-8.

Jolley, K. A., Bray, J. E. & Maiden, M. C. J. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res,* 3**,** 124.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24**,** 1403-5.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet,* 11**,** 94.

Jones, L. L., Hashim, A., Mckeever, T., Cook, D. G., Britton, J. & Leonardi-Bee, J. 2011. Parental and household smoking and the increased risk of bronchitis, bronchiolitis and other lower respiratory infections in infancy: systematic review and meta-analysis. *Respiratory research,* 12**,** 5.

Jones, P. W., Harding, G., Berry, P., Wiklund, I., Chen, W. H. & Kline Leidy, N. 2009. Development and first validation of the COPD Assessment Test. *Eur Respir J,* 34**,** 648-54.

Jurcisek, J. A., Bookwalter, J. E., Baker, B. D., Fernandez, S., Novotny, L. A., Munson, R. S., Jr. & Bakaletz, L. O. 2007. The PilA protein of non-typeable Haemophilus influenzae plays a role in biofilm formation, adherence to epithelial cells and colonization of the mammalian upper respiratory tract. *Mol Microbiol,* 65**,** 1288-99.

Kan, H., Heiss, G., Rose, K. M., Whitsel, E., Lurmann, F. & London, S. J. 2007. Traffic exposure and lung function in adults: the Atherosclerosis Risk in Communities study. *Thorax,* 62**,** 873-9.

Kappler, U., Dhouib, R., Nair, R. P. & Mcewan, A. G. 2017. Draft Genome Sequences of Three Nontypeable Strains of Haemophilus influenzae, C188, R535, and 1200, Isolated from Different Types of Disease. *Genome Announc,* 5.

Karudapuram, S. & Barcak, G. J. 1997. The Haemophilus influenzae dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. *J Bacteriol,* 179**,** 4815-20.

Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research,* 30**,** 3059-3066.

Kc, R., Hyland, I. K., Smith, J. A., Shukla, S. D., Hansbro, P. M., Zosky, G. R., Karupiah, G. & O'toole, R. F. 2020a. Cow Dung Biomass Smoke Exposure Increases Adherence of Respiratory Pathogen Nontypeable Haemophilus influenzae to Human Bronchial Epithelial Cells. *Exposure and Health***,** 1-13.

Kc, R., Leong, K. W. C., Harkness, N. M., Lachowicz, J., Gautam, S. S., Cooley, L. A., Mcewan, B., Petrovski, S., Karupiah, G. & O'toole, R. F. 2020b. Whole-genome analyses reveal gene content differences between nontypeable Haemophilus influenzae isolates from chronic obstructive pulmonary disease compared to other clinical phenotypes. *Microb Genom*.

Kc, R., Leong, K. W. C., Mcewan, B., Lachowicz, J., Harkness, N. M., Petrovski, S., Karupiah, G. & O'toole, R. F. 2020c. Draft Genome Sequence of an Isolate of Nontypeable Haemophilus influenzae from an Acute Exacerbation of Chronic Obstructive Pulmonary Disease in Tasmania. *Microbiol Resour Announc,* 9.

Kc, R., Shukla, S. D., Gautam, S. S., Hansbro, P. M. & O'toole, R. F. 2018a. The role of environmental exposure to non-cigarette smoke in lung disease. *Clin Transl Med,* 7**,** 39.

Kc, R., Shukla, S. D., Walters, E. H. & O'toole, R. F. 2017. Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease. *Microbiology,* 163**,** 421-430.

Kc, R., Zosky, G. R., Shukla, S. D. & O'toole, R. F. 2018b. A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells. *Biology Methods and Protocols,* 3**,** bpy010.

Keely, S., Glover, L. E., Weissmueller, T., Macmanus, C. F., Fillon, S., Fennimore, B. & Colgan, S. P. 2010. Hypoxia-inducible factor-dependent regulation of platelet-activating factor receptor as a route for gram-positive bacterial translocation across epithelia. *Mol Biol Cell,* 21**,** 538-46.

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

Kimura, A. & Kishimoto, T. 2010. IL-6: regulator of Treg/Th17 balance. *Eur J Immunol,* 40**,** 1830-5.

Klaile, E., Klassert, T. E., Scheffrahn, I., Muller, M. M., Heinrich, A., Heyl, K. A., Dienemann, H., Grunewald, C., Bals, R., Singer, B. B. & Slevogt, H. 2013. Carcinoembryonic antigen (CEA)-related cell adhesion molecules are co-expressed in the human lung and their expression can be modulated in bronchial epithelial cells by non-typable Haemophilus influenzae, Moraxella catarrhalis, TLR3, and type I and II interferons. *Respir Res,* 14**,** 85.

Kocbach, A., Herseth, J. I., Lag, M., Refsnes, M. & Schwarze, P. E. 2008. Particles from wood smoke and traffic induce differential pro-inflammatory response patterns in co-cultures. *Toxicol Appl Pharmacol,* 232**,** 317-26.

Krzyzanowski, M., Jedrychowski, W. & Wysocki, M. 1986. Factors associated with the change in ventilatory function and the development of chronic obstructive pulmonary disease in a 13-year follow-up of the Cracow Study: risk of chronic obstructive pulmonary disease. *American Review of Respiratory Disease,* 134**,** 1011-1019.

Kuespert, K., Pils, S. & Hauck, C. R. 2006. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol,* 18**,** 565-71.

Kulkarni, N., Pierse, N., Rushton, L. & Grigg, J. 2006. Carbon in airway macrophages and lung function in children. *N Engl J Med,* 355**,** 21-30.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution,* 35**,** 1547-1549.

Kurmi, O. P., Dunster, C., Ayres, J. G. & Kelly, F. J. 2013. Oxidative potential of smoke from burning wood and mixed biomass fuels. *Free Radic Res,* 47**,** 829-35.

Kurmi, O. P., Semple, S., Simkhada, P., Smith, W. C. & Ayres, J. G. 2010. COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax,* 65**,** 221-8.

Lacross, N. C., Marrs, C. F. & Gilsdorf, J. R. 2013. Population structure in nontypeable Haemophilus influenzae. *Infect Genet Evol,* 14**,** 125-36.

Landis, S. H., Muellerova, H., Mannino, D. M., Menezes, A. M., Han, M. K., Van Der Molen, T., Ichinose, M., Aisanov, Z., Oh, Y. M. & Davis, K. J. 2014. Continuing to Confront COPD International Patient Survey: methods, COPD prevalence, and disease burden in 2012-2013. *Int J Chron Obstruct Pulmon Dis,* 9**,** 597-611.

Lange, P. 2009. Chronic obstructive pulmonary disease and risk of infection. *Pneumonol Alergol Pol,* 77**,** 284-8.

Larson, T. V. & Koenig, J. Q. 1994. Wood smoke: emissions and noncancer respiratory effects. *Annu Rev Public Health,* 15**,** 133-56.

Leavey, A., Patel, S., Martinez, R., Mitroo, D., Fortenberry, C., Walker, M., Williams, B. & Biswas, P. 2017. Organic and inorganic speciation of particulate matter formed during different combustion phases in an improved cookstove. *Environ Res,* 158**,** 33-42.

Leclerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. 1996. High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science,* 274**,** 1208-11.

Lee, A., Kinney, P., Chillrud, S. & Jack, D. 2015. A Systematic Review of Innate Immunomodulatory Effects of Household Air Pollution Secondary to the Burning of Biomass Fuels. *Ann Glob Health,* 81**,** 368-74.

Lee, M. S., Lee, Y. S. & Chiou, C. S. 2006. The suitable restriction enzymes for pulsed-field gel electrophoresis analysis of Bordetella pertussis. *Diagn Microbiol Infect Dis,* 56**,** 217-9.

Leimbach, A., Hacker, J. & Dobrindt, U. 2013. E. coli as an all-rounder: the thin line between commensalism and pathogenicity. *Curr Top Microbiol Immunol,* 358**,** 3-32.

Leong, K. W., Cooley, L. A., Anderson, T. L., Gautam, S. S., Mcewan, B., Wells, A., Wilson, F., Hughson, L. & O'toole, R. F. 2018. Emergence of vancomycin-resistant Enterococcus faecium at an Australian hospital: a whole genome sequencing analysis. *Scientific reports,* 8**,** 1-11.

Leong, K. W. C., Kalukottege, R., Cooley, L. A., Anderson, T. L., Wells, A., Langford, E. & O'toole, R. F. 2019. State-Wide Genomic and Epidemiological Analyses of Vancomycin-Resistant Enterococcus faecium in Tasmania's Public Hospitals. *Front Microbiol,* 10**,** 2940.

Leung, J. M., Tiew, P. Y., Mac Aogain, M., Budden, K. F., Yong, V. F., Thomas, S. S., Pethe, K., Hansbro, P. M. & Chotirmall, S. H. 2017. The role of acute and chronic respiratory colonization and infections in the pathogenesis of COPD. *Respirology,* 22**,** 634-650.

Lewis, B. B., Carter, R. A., Ling, L., Leiner, I., Taur, Y., Kamboj, M., Dubberke, E. R., Xavier, J. & Pamer, E. G. 2017. Pathogenicity locus, core genome, and accessory gene contributions to Clostridium difficile virulence. *MBio,* 8**,** e00885-17.

Li, C., Zhihong, H., Wenlong, L., Xiaoyan, L., Qing, C., Wenzhi, L., Siming, X. & Shengming, L. 2016. The nucleotide-binding oligomerization domain-like receptor family pyrindomain-containing 3 inflammasome regulates bronchial epithelial cell injury and proapoptosis after exposure to biomass fuel smoke. *American Journal of Respiratory Cell and Molecular Biology,* 55**,** 815-824.

Li, N., Ren, A., Wang, X., Fan, X., Zhao, Y., Gao, G. F., Cleary, P. & Wang, B. 2015. Influenza viral neuraminidase primes bacterial coinfection through TGF-beta-mediated expression of host cell receptors. *Proc Natl Acad Sci U S A,* 112**,** 238-43.

Lieberman, T. D., Michel, J. B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Jr., Skurnik, D., Leiby, N., Lipuma, J. J., Goldberg, J. B., Mcadam, A. J., Priebe, G. P. & Kishony, R. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet,* 43**,** 1275-80.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., et al. 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2224-60.

Loeb, M., Neupane, B., Walter, S. D., Hanning, R., Carusone, S. C., Lewis, D., Krueger, P., Simor, A. E., Nicolle, L. & Marrie, T. J. 2009. Environmental risk factors for community-acquired pneumonia hospitalization in older adults. *J Am Geriatr Soc,* 57**,** 1036-40.

Lomholt, H., Van Alphen, L. & Kilian, M. 1993. Antigenic variation of immunoglobulin A1 proteases among sequential isolates of Haemophilus influenzae from healthy children and patients with chronic obstructive pulmonary disease. *Infect Immun,* 61**,** 4575-81.

Long, S. S., Teter, M. J. & Gilligan, P. H. 1983. Biotype of Haemophilus influenzae: correlation with virulence and ampicillin resistance. *J Infect Dis,* 147**,** 800-6.

Lopez-Campos, J. L., Calero, C., Arellano-Orden, E., Marquez-Martin, E., Cejudo-Ramos, P., Ortega Ruiz, F. & Montes-Worboys, A. 2012. Increased levels of soluble ICAM-1 in chronic obstructive pulmonary disease and resistant smokers are related to active smoking. *Biomark Med,* 6**,** 805-11.

Lorenz, M. G. & Wackernagel, W. 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev,* 58**,** 563-602.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature,* 489**,** 220-30.

Lv, X. X., Liu, S. S., Li, K., Cui, B., Liu, C. & Hu, Z. W. 2017. Cigarette smoke promotes COPD by activating platelet-activating factor receptor and inducing neutrophil autophagic death in mice. *Oncotarget,* 8**,** 74720-74735.

Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. *J Theor Biol,* 202**,** 195-204.

Magitta, N. F., Walker, R. W., Apte, K. K., Shimwela, M. D., Mwaiselage, J. D., Sanga, A. A., Namdeo, A. K., Madas, S. J. & Salvi, S. S. 2018. Prevalence, risk factors and clinical correlates of COPD in a rural setting in Tanzania. *Eur Respir J,* 51.

Mccarthy, C. E., Duffney, P. F., Gelein, R., Thatcher, T. H., Elder, A., Phipps, R. P. & Sime, P. J. 2016a. Dung biomass smoke activates inflammatory signaling pathways in human small airway epithelial cells. *Am J Physiol Lung Cell Mol Physiol,* 311**,** L1222-l1233.

Mccarthy, C. E., Duffney, P. F., Gelein, R., Thatcher, T. H., Elder, A., Phipps, R. P. & Sime, P. J. 2016b. Dung biomass smoke activates inflammatory signaling pathways in human small airway epithelial cells. *American Journal of Physiology - Lung Cellular and Molecular Physiology,* 311**,** L1222-L1233.

Meats, E., Feil, E. J., Stringer, S., Cody, A. J., Goldstein, R., Kroll, J. S., Popovic, T. & Spratt, B. G. 2003. Characterization of encapsulated and noncapsulated Haemophilus influenzae and determination of phylogenetic relationships by multilocus sequence typing. *Journal of clinical microbiology,* 41**,** 1623-1636.

Mehra, D., Geraghty, P. M., Hardigan, A. A. & Foronjy, R. 2012a. A comparison of the inflammatory and proteolytic effects of dung biomass and cigarette smoke exposure in the lung. *PLoS One,* 7**,** e52889.

Mehra, D., Mazumder, S., Dube, B., Taneja, S., Hardigan, A., Geraghty, P., Foronjy, R. F. & D'armiento, J. 2012b. Respiratory health effects of dung biomass smoke exposure. *American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS,* 185.

Meijer, E., Kromhout, H. & Heederik, D. 2001. Respiratory effects of exposure to low levels of concrete dust containing crystalline silica. *Am J Ind Med,* 40**,** 133-40.

Messer, J. S., Liechty, E. R., Vogel, O. A. & Chang, E. B. 2017. Evolutionary and ecological forces that shape the bacterial communities of the human gut. *Mucosal Immunol,* 10**,** 567-579.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res,* 47**,** D419-d426.

Miyashita, L., Padovan, M., Suri, R., Whitehouse, A. & Grigg, J. 2017. Expression of Nasal Epithelial Platelet Activating Factor Receptor (PAFR) and in vivo Exposure to Air Pollution. *Journal of Allergy and Clinical Immunology,* 139**,** AB264.

Miyashita, L., Suri, R., Dearing, E., Mudway, I., Dove, R. E., Neill, D. R., Van Zyl-Smit, R., Kadioglu, A. & Grigg, J. 2018. E-cigarette vapour enhances pneumococcal adherence to airway epithelial cells. *Eur Respir J,* 51.

Moghaddam, S. J., Ochoa, C. E., Sethi, S. & Dickey, B. F. 2011. Nontypeable Haemophilus influenzae in chronic obstructive pulmonary disease and lung cancer. *International journal of chronic obstructive pulmonary disease,* 6**,** 113.

Moleres, J., Fernandez-Calvet, A., Ehrlich, R. L., Marti, S., Perez-Regidor, L., Euba, B., Rodriguez-Arce, I., Balashov, S., Cuevas, E., Linares, J., Ardanuy, C., Martin-Santamaria, S., Ehrlich, G. D., Mell, J. C. & Garmendia, J. 2018. Antagonistic Pleiotropy in the Bifunctional Surface Protein FadL (OmpP1) during Adaptation of Haemophilus influenzae to Chronic Lung Infection Associated with Chronic Obstructive Pulmonary Disease. *mBio,* 9.

Molin, S. & Tolker-Nielsen, T. 2003. Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Current opinion in biotechnology,* 14**,** 255-261.

Montes De Oca, M., Zabert, G., Moreno, D., Laucho-Contreras, M. E., Lopez Varela, M. V. & Surmont, F. 2017. Smoke, Biomass Exposure, and COPD Risk in the Primary Care Setting: The PUMA Study. *Respir Care,* 62**,** 1058-1066.

Moxon, E. R. & Vaughn, K. A. 1981. The type b capsular polysaccharide as a virulence determinant of Haemophilus influenzae: studies using clinical isolates and laboratory transformants. *J Infect Dis,* 143**,** 517-24.

Mudway, I. S., Duggan, S. T., Venkataraman, C., Habib, G., Kelly, F. J. & Grigg, J. 2005. Combustion of dried animal dung as biofuel results in the generation of highly redox active fine particulates. *Part Fibre Toxicol,* 2**,** 6.

Mukaida, N. 2000. Interleukin-8: an expanding universe beyond neutrophil chemotaxis and activation. *Int J Hematol,* 72**,** 391-398.

Mukaida, N. 2003. Pathophysiological roles of interleukin-8/CXCL8 in pulmonary diseases. *Am J Physiol Lung Cell Mol Physiol,* 284**,** L566-77.

Mukherjee, S., Roychoudhury, S., Siddique, S., Banerjee, M., Bhattacharya, P., Lahiri, T. & Ray, M. R. 2014. Respiratory symptoms, lung function decrement and chronic obstructive pulmonary disease in pre-menopausal Indian women exposed to biomass smoke. *Inhal Toxicol,* 26**,** 866-72.

Murphy, T. F. 2006. Otitis media, bacterial colonization, and the smoking parent. The University of Chicago Press.

Murphy, T. F., Bernstein, J. M., Dryja, D. M., Campagnari, A. A. & Apicella, M. A. 1987. Outer membrane protein and lipooligosaccharide analysis of paired nasopharyngeal and middle ear isolates in otitis media due to nontypable Haemophilus influenzae: pathogenetic and epidemiological observations. *Journal of Infectious Diseases,* 156**,** 723-731.

Murphy, T. F., Brauer, A. L., Schiffmacher, A. T. & Sethi, S. 2004. Persistent colonization by Haemophilus influenzae in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine,* 170**,** 266-272.

Murphy, T. F., Lesse, A. J., Kirkham, C., Zhong, H., Sethi, S. & Munson, R. S., Jr. 2011. A clonal group of nontypeable Haemophilus influenzae with two IgA proteases is adapted to infection in chronic obstructive pulmonary disease. *PLoS One,* 6**,** e25923.

Murphy, T. F., Sethi, S., Klingman, K. L., Brueggemann, A. B. & Doern, G. V. 1999. Simultaneous respiratory tract colonization by multiple strains of nontypeable Haemophilus influenzae in chronic obstructive pulmonary disease: implications for antibiotic therapy. *The journal of infectious diseases,* 180**,** 404-409.

Mushtaq, N., Ezzati, M., Hall, L., Dickson, I., Kirwan, M., Png, K. M., Mudway, I. S. & Grigg, J. 2011. Adhesion of Streptococcus pneumoniae to human airway epithelial cells exposed to urban particulate matter. *J Allergy Clin Immunol,* 127**,** 1236-42.e2.

Nadel, J. A. 2000. Role of neutrophil elastase in hypersecretion during COPD exacerbations, and proposed therapies. *Chest,* 117**,** 386s-9s.

Naeher, L. P., Brauer, M., Lipsett, M., Zelikoff, J. T., Simpson, C. D., Koenig, J. Q. & Smith, K. R. 2007. Woodsmoke health effects: a review. *Inhal Toxicol,* 19**,** 67-106.

Negro Alvarez, J. M., Miralles Lopez, J. C., Ortiz Martinez, J. L., Abellan Aleman, A. & Rubio Del Barrio, R. 1997. Platelet-activating factor antagonists. *Allergol Immunopathol (Madr),* 25**,** 249-58.

Neurath, M. 1993. [Expression of tenascin, laminin and fibronectin following traumatic rupture of the anterior cruciate ligament]. *Z Orthop Ihre Grenzgeb,* 131**,** 168-72.

Niven, R. M., Fletcher, A. M., Pickering, C. A., Fishwick, D., Warburton, C. J., Simpson, J. C., Francis, H. & Oldham, L. A. 1997. Chronic bronchitis in textile workers. *Thorax,* 52**,** 22-7.

North, B. V., Curtis, D. & Sham, P. C. 2003. A note on calculation of empirical P values from Monte Carlo procedure. *Am J Hum Genet,* 72**,** 498-9.

Novotny, L. A. & Bakaletz, L. O. 2016. Intercellular adhesion molecule 1 serves as a primary cognate receptor for the Type IV pilus of nontypeable Haemophilus influenzae. *Cell Microbiol,* 18**,** 1043-55.

Nuorti, J. P., Butler, J. C., Farley, M. M., Harrison, L. H., Mcgeer, A., Kolczak, M. S., Breiman, R. F. & Team, A. B. C. S. 2000. Cigarette smoking and invasive pneumococcal disease. *New England Journal of Medicine,* 342**,** 681-689.

O'toole, R. F., Shukla, S. D. & Walters, E. H. 2016. Does upregulated host cell receptor expression provide a link between bacterial adhesion and chronic respiratory disease? *J Transl Med,* 14**,** 304.

Ochman, H., Lawrence, J. G. & Groisman, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature,* 405**,** 299-304.

Oishi, K., Sonoda, F., Kobayashi, S., Iwagaki, A., Nagatake, T., Matsushima, K. & Matsumoto, K. 1994. Role of interleukin-8 (IL-8) and an inhibitory effect of erythromycin on IL-8 release in the airways of patients with chronic airway diseases. *Infect Immun,* 62**,** 4145-52.

Olloquequi, J. & Silva, O. R. 2016. Biomass smoke as a risk factor for chronic obstructive pulmonary disease: effects on innate immunity. *Innate Immun,* 22**,** 373-81.

Osman, K. L., Jefferies, J. M., Woelk, C. H., Cleary, D. W. & Clarke, S. C. 2018. The adhesins of non-typeable Haemophilus influenzae. *Expert Rev Anti Infect Ther,* 16**,** 187-196.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A. & Parkhill, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3.

Pahla, G., Mamvura, T., Ntuli, F. & Muzenda, E. 2017. Energy densification of animal waste lignocellulose biomass and raw biomass. *south african journal of chemical engineering,* 24**,** 168-175.

Pant, K. P. 2012. Cheaper fuel and higher health costs among the poor in rural Nepal. *Ambio,* 41**,** 271-83.

Park, E. & Lee, K. 2003. Particulate exposure and size distribution from wood burning stoves in Costa Rica. *Indoor Air,* 13**,** 253-9.

Patel, P. S. & Kearney, J. F. 2017. CD36 and Platelet-Activating Factor Receptor Promote House Dust Mite Allergy Development. *J Immunol,* 199**,** 1184-1195.

Pauwels, R. A., Buist, A. S., Calverley, P. M., Jenkins, C. R. & Hurd, S. S. 2001. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med,* 163**,** 1256-76.

Pettigrew, M. M., Ahearn, C. P., Gent, J. F., Kong, Y., Gallo, M. C., Munro, J. B., D'mello, A., Sethi, S., Tettelin, H. & Murphy, T. F. 2018. Haemophilus influenzae genome evolution during

persistence in the human airways in chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A,* 115**,** E3256-e3265.

Phillips, Z. N., Brizuela, C., Jennison, A. V., Staples, M., Grimwood, K., Seib, K. L., Jennings, M. P. & Atack, J. M. 2019. Analysis of Invasive Nontypeable Haemophilus influenzae Isolates Reveals Selection for the Expression State of Particular Phase-Variable Lipooligosaccharide Biosynthetic Genes. *Infect Immun,* 87.

Plow, E. F., Haas, T. A., Zhang, L., Loftus, J. & Smith, J. W. 2000. Ligand binding to integrins. *J Biol Chem,* 275**,** 21785-8.

Po, J. Y., Fitzgerald, J. M. & Carlsten, C. 2011. Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. *Thorax,* 66**,** 232-9.

Pope, D., Diaz, E., Smith-Sivertsen, T., Lie, R. T., Bakke, P., Balmes, J. R., Smith, K. R. & Bruce, N. G. 2015. Exposure to household air pollution from wood combustion and association with respiratory symptoms and lung function in nonsmoking women: results from the RESPIRE trial, Guatemala. *Environ Health Perspect,* 123**,** 285-92.

Post, W., Heederik, D. & Houba, R. 1998. Decline in lung function related to exposure and selection processes among workers in the grain processing and animal feed industry. *Occup Environ Med,* 55**,** 349-55.

Power, P. M., Bentley, S. D., Parkhill, J., Moxon, E. R. & Hood, D. W. 2012. Investigations into genome diversity of Haemophilus influenzae using whole genome sequencing of clinical isolates and laboratory transformants. *BMC Microbiol,* 12**,** 273.

Pracht, D., Elm, C., Gerber, J., Bergmann, S., Rohde, M., Seiler, M., Kim, K. S., Jenkinson, H. F., Nau, R. & Hammerschmidt, S. 2005. PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. *Infect Immun,* 73**,** 2680-9.

Ramos, D., Proenca, M., Leite, M. R., Ferreira, A. D., Trevisan, I. B., Brigida, G. F. S., Tacao, G. Y. & Ramos, E. M. C. 2017. Effects of exposure to biomass burning on pulmonary inflammatory markers and pulmonary function in individuals with COPD. *Rev Port Pneumol (2006),* 23**,** 273-279.

Redfield, R. J. 1991. sxy-1, a Haemophilus influenzae mutation causing greatly enhanced spontaneous competence. *J Bacteriol,* 173**,** 5612-8.

Redfield, R. J., Cameron, A. D., Qian, Q., Hinds, J., Ali, T. R., Kroll, J. S. & Langford, P. R. 2005. A novel CRP-dependent regulon controls expression of competence genes in Haemophilus influenzae. *J Mol Biol,* 347**,** 735-47.

Regalado, J., Perez-Padilla, R., Sansores, R., Paramo Ramirez, J. I., Brauer, M., Pare, P. & Vedal, S. 2006. The effect of biomass burning on respiratory symptoms and lung function in rural Mexican women. *Am J Respir Crit Care Med,* 174**,** 901-5.

Reisen, F. & Brown, S. K. 2009. Australian firefighters' exposure to air toxics during bushfire burns of autumn 2005 and 2006. *Environ Int,* 35**,** 342-52.

Rennard, S. I. 2004. Cigarette smoke in research. *Am J Respir Cell Mol Biol,* 31**,** 479-80.

Repapi, E., Sayers, I., Wain, L. V., Burton, P. R., Johnson, T., Obeidat, M., Zhao, J. H., Ramasamy, A., Zhai, G., Vitart, V., Huffman, J. E., Igl, W., Albrecht, E., Deloukas, P., Henderson, J., Granell, R., Mcardle, W. L., Rudnicka, A. R., Barroso, I., Loos, R. J., Wareham, N. J., Mustelin, L., Rantanen, T., Surakka, I., Imboden, M., Wichmann, H. E., Grkovic, I., Jankovic, S., Zgaga, L., Hartikainen, A. L., Peltonen, L., Gyllensten, U., Johansson, A., Zaboli, G., Campbell, H., Wild, S. H., Wilson, J. F., Gläser, S., Homuth, G., Völzke, H., Mangino, M., Soranzo, N., Spector, T. D., Polasek, O., Rudan, I., Wright, A. F., Heliövaara, M., Ripatti, S., Pouta, A., Naluai, A. T., Olin, A. C., Torén, K., Cooper, M. N., James, A. L., Palmer, L. J., Hingorani, A. D., Wannamethee, S. G., Whincup, P. H., Smith, G. D., Ebrahim, S., Mckeever, T. M., Pavord, I. D., Macleod, A. K., Morris, A. D., Porteous, D. J., Cooper, C., Dennison, E., Shaheen, S., Karrasch, S., Schnabel, E., Schulz, H., Grallert, H., Bouatia-Naji, N., Delplanque, J., Froguel, P., Blakey, J. D., Britton, J. R., Morris, R. W., Holloway, J. W., Lawlor, D. A., Hui, J., Nyberg, F., Jarvelin, M. R., Jackson, C., Kähönen, M., Kaprio, J., Probst-Hensch, N. M., Koch, B., Hayward, C., Evans, D. M., Elliott, P., Strachan, D. P., Hall, I. P. & Tobin, M. D. 2010. Genome-wide association study identifies five loci associated with lung function. *Nat Genet,* 42**,** 36-44.

Rivera, R. M., Cosio, M. G., Ghezzo, H., Salazar, M. & Perez-Padilla, R. 2008. Comparison of lung morphology in COPD secondary to cigarette and biomass smoke. *Int J Tuberc Lung Dis,* 12**,** 972-7.

Robbins, A. S., Abbey, D. E. & Lebowitx, M. D. 1993. Passive smoking and chronic respiratory disease symptoms in non-smoking adults. *International Journal of Epidemiology,* 22**,** 809-817.

Rodgman, A. & Perfetti, T. A. 2016. *The chemical components of tobacco and tobacco smoke*, CRC press.

Roebuck, K. A., Rahman, A., Lakshminarayanan, V., Janakidevi, K. & Malik, A. B. 1995. H2O2 and tumor necrosis factor-alpha activate intercellular adhesion molecule 1 (ICAM-1) gene transcription through distinct cis-regulatory elements within the ICAM-1 promoter. *J Biol Chem,* 270**,** 18966-74.

Rogalsky, D. K., Mendola, P., Metts, T. A. & Martin, W. J., 2nd 2014. Estimating the number of low-income americans exposed to household air pollution from burning solid fuels. *Environ Health Perspect,* 122**,** 806-10.

Romieu, I., Riojas-Rodriguez, H., Marron-Mares, A. T., Schilmann, A., Perez-Padilla, R. & Masera, O. 2009. Improved biomass stove intervention in rural Mexico: impact on the respiratory health of women. *Am J Respir Crit Care Med,* 180**,** 649-656.

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J. & Abdelalim, A. 2018. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet,* 392**,** 1736-1788.

Rylance, J., Fullerton, D. G., Scriven, J., Aljurayyan, A. N., Mzinza, D., Barrett, S., Wright, A. K., Wootton, D. G., Glennie, S. J., Baple, K., Knott, A., Mortimer, K., Russell, D. G., Heyderman, R. S. & Gordon, S. B. 2015. Household air pollution causes dose-dependent inflammation and altered phagocytosis in human macrophages. *Am J Respir Cell Mol Biol,* 52**,** 584-93.

Saito, M., Umeda, A. & Yoshida, S. 1999. Subtyping of Haemophilus influenzae strains by pulsed-field gel electrophoresis. *J Clin Microbiol,* 37**,** 2142-7.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol,* 4**,** 406-25.

World Health Organization. 2003. *Database of Household Air Pollution Studies in Developing Countries, Protection of the Human Environment* [Online]. Geneva: World Health Organization. [Accessed 11 Aug 2018].

Salipante, S. J., Sengupta, D. J., Cummings, L. A., Land, T. A., Hoogestraat, D. R. & Cookson, B. T. 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J Clin Microbiol,* 53**,** 1072-9.

Salvi, S. S. & Barnes, P. J. 2009. Chronic obstructive pulmonary disease in non-smokers. *Lancet,* 374**,** 733-43.

Samet, J. M. 2014. The Surgeon Generals' reports and respiratory diseases. From 1964 to 2014. *Ann Am Thorac Soc,* 11**,** 141-8.

Sana, A., Somda, S. M. A., Meda, N. & Bouland, C. 2018. Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. *BMJ Open Respir Res,* 5**,** e000246.

Schikowski, T., Mills, I. C., Anderson, H. R., Cohen, A., Hansell, A., Kauffmann, F., Krämer, U., Marcon, A., Perez, L. & Sunyer, J. 2014. Ambient air pollution: a cause of COPD? *European Respiratory Journal,* 43**,** 250-263.

Schikowski, T., Sugiri, D., Ranft, U., Gehring, U., Heinrich, J., Wichmann, H. E. & Kramer, U. 2005. Long-term air pollution exposure and living close to busy roads are associated with COPD in women. *Respir Res,* 6**,** 152.

Schmidt, S. a. J., Johansen, M. B., Olsen, M., Xu, X., Parker, J. M., Molfino, N. A., Lash, T. L., Sørensen, H. T. & Christiansen, C. F. 2014. The impact of exacerbation frequency on mortality following acute exacerbations of COPD: a registry-based cohort study. *BMJ open,* 4**,** e006720.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

Schoen, C., Kischkies, L., Elias, J. & Ampattu, B. J. 2014. Metabolism and virulence in Neisseria meningitidis. *Frontiers in cellular and infection microbiology,* 4**,** 114.

Schröder, J. 1989. The monocyte-derived neutrophil activating peptide (NAP/interleukin 8) stimulates human neutrophil arachidonate-5-lipoxygenase, but not the release of cellular arachidonate. *J Exp Med,* 170**,** 847-863.

Schwartz, D. A., Thorne, P. S., Yagla, S. J., Burmeister, L. F., Olenchock, S. A., Watt, J. L. & Quinn, T. J. 1995. The role of endotoxin in grain dust-induced lung disease. *Am J Respir Crit Care Med,* 152**,** 603-8.

Seeman, T. 2014. Mlst. Github.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

Sethi, S. 1999. Etiology and management of infections in chronic obstructive pulmonary disease. *Clinical Pulmonary Medicine,* 6**,** 327-332.

Sethi, S. 2010. Infection as a comorbidity of COPD. *Eur Respir J,* 35**,** 1209-15.

Sethi, S., Evans, N., Grant, B. J. & Murphy, T. F. 2002. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med,* 347**,** 465-471.

Sethi, S., Maloney, J., Grove, L., Wrona, C. & Berenson, C. S. 2006. Airway inflammation and bronchial bacterial colonization in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 173**,** 991-8.

Sethi, S. & Murphy, T. F. 2001. Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review. *Clin Microbiol Rev,* 14**,** 336-63.

Shao, J., Wheeler, A. J., Chen, L., Strandberg, B., Hinwood, A., Johnston, F. H. & Zosky, G. R. 2018. The pro-inflammatory effects of particulate matter on epithelial cells are associated with elemental composition. *Chemosphere,* 202**,** 530-537.

Sharma, M. & Majumdar, P. K. 2009. Occupational lifestyle diseases: An emerging issue. *Indian J Occup Environ Med,* 13**,** 109-12.

Shen, K., Antalis, P., Gladitz, J., Sayeed, S., Ahmed, A., Yu, S., Hayes, J., Johnson, S., Dice, B., Dopico, R., Keefe, R., Janto, B., Chong, W., Goodwin, J., Wadowsky, R. M., Erdos, G., Post, J. C., Ehrlich, G. D. & Hu, F. Z. 2005. Identification, distribution, and expression of novel genes in 10 clinical isolates of nontypeable Haemophilus influenzae. *Infect Immun,* 73**,** 3479-91.

Shimizu, K., Yoshii, Y., Morozumi, M., Chiba, N., Ubukata, K., Uruga, H., Hanada, S., Saito, N., Kadota, T., Ito, S., Wakui, H., Takasaka, N., Minagawa, S., Kojima, J., Hara, H., Numata, T., Kawaishi, M., Saito, K., Araya, J., Kaneko, Y., Nakayama, K., Kishi, K. & Kuwano, K. 2015. Pathogens in COPD exacerbations identified by comprehensive real-time PCR plus older methods. *Int J Chron Obstruct Pulmon Dis,* 10**,** 2009-16.

Shirasaki, H., Nishikawa, M., Adcock, I. M., Mak, J. C., Sakamoto, T., Shimizu, T. & Barnes, P. J. 1994. Expression of platelet-activating factor receptor mRNA in human and guinea pig lung. *Am J Respir Cell Mol Biol,* 10**,** 533-7.

Shukla, S. D. 1992. Platelet-activating factor receptor and signal transduction mechanisms. *Faseb j,* 6**,** 2296-301.

Shukla, S. D., Fairbairn, R. L., Gell, D. A., Latham, R. D., Sohal, S. S., Walters, E. H. & O'toole, R. F. 2016a. An antagonist of the platelet-activating factor receptor inhibits adherence of both nontypeable Haemophilus influenzae and Streptococcus pneumoniae to cultured human bronchial epithelial cells exposed to cigarette smoke. *Int J Chron Obstruct Pulmon Dis,* 11**,** 1647-55.

Shukla, S. D., Mahmood, M. Q., Weston, S., Latham, R., Muller, H. K., Sohal, S. S. & Walters, E. H. 2017. The main rhinovirus respiratory tract adhesion site (ICAM-1) is upregulated in smokers and patients with chronic airflow limitation (CAL). *Respir Res,* 18**,** 6.

Shukla, S. D., Muller, H. K., Latham, R., Sohal, S. S. & Walters, E. H. 2016b. Platelet-activating factor receptor (PAFr) is upregulated in small airways and alveoli of smokers and COPD patients. *Respirology,* 21**,** 504-10.

Shukla, S. D., Sohal, S. S., Mahmood, M. Q., Reid, D., Muller, H. K. & Walters, E. H. 2014. Airway epithelial platelet-activating factor receptor expression is markedly upregulated in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis,* 9**,** 853-61.

Shukla, S. D., Sohal, S. S., O'toole, R. F., Eapen, M. S. & Walters, E. H. 2015. Platelet activating factor receptor: gateway for bacterial chronic airway infection in chronic obstructive pulmonary disease and potential therapeutic target. *Expert Rev Respir Med,* 9**,** 473-85.

Siddharthan, T., Grigsby, M. R., Goodman, D., Chowdhury, M., Rubinstein, A., Irazola, V., Gutierrez, L., Miranda, J. J., Bernabe-Ortiz, A., Alam, D., Kirenga, B., Jones, R., Van Gemert, F., Wise, R. A. & Checkley, W. 2018. Association between Household Air Pollution Exposure and Chronic Obstructive Pulmonary Disease Outcomes in 13 Low- and Middle-Income Country Settings. *Am J Respir Crit Care Med,* 197**,** 611-620.

Singh, B., Fleury, C., Jalalvand, F. & Riesbeck, K. 2012. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. *FEMS Microbiol Rev,* 36**,** 1122-80.

Sisco, K. L. & Smith, H. O. 1979. Sequence-specific DNA uptake in Haemophilus transformation. *Proc Natl Acad Sci U S A,* 76**,** 972-6.

Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol,* 13**,** 605-11.

Smith, M., Li, L., Augustyn, M., Kurmi, O., Chen, J., Collins, R., Guo, Y., Han, Y., Qin, J., Xu, G., Wang, J., Bian, Z., Zhou, G., Peto, R. & Chen, Z. 2014. Prevalence and correlates of airflow obstruction in approximately 317,000 never-smokers in China. *Eur Respir J,* 44**,** 66-77.

Smoking, U. S. S. G. S. a. C. O. 1964. *Smoking and health: report of the advisory committee to the Surgeon General of the Public Health Service*, US Department of Health, Education, and Welfare, Public Health Service.

Soler, N., Ewig, S., Torres, A., Filella, X., Gonzalez, J. & Zaubet, A. 1999. Airway inflammation and bronchial microbial patterns in patients with stable chronic obstructive pulmonary disease. *Eur Respir J,* 14**,** 1015-22.

Sonego, M., Pellegrin, M. C., Becker, G. & Lazzerini, M. 2015. Risk factors for mortality from acute lower respiratory infections (ALRI) in children under five years of age in low and middle-income countries: a systematic review and meta-analysis of observational studies. *PLoS One,* 10**,** e0116380.

Spence, D. P., Johnston, S. L., Calverley, P. M., Dhillon, P., Higgins, C., Ramhamadany, E., Turner, S., Winning, A., Winter, J. & Holgate, S. T. 1994. The effect of the orally active platelet-activating factor antagonist WEB 2086 in the treatment of asthma. *Am J Respir Crit Care Med,* 149**,** 1142-8.

Sriram, K. B., Cox, A. J., Clancy, R. L., Slack, M. P. E. & Cripps, A. W. 2018. Nontypeable Haemophilus influenzae and chronic obstructive pulmonary disease: a review for clinicians. *Crit Rev Microbiol,* 44**,** 125-142.

Australian Bureau of Statistics. 2014. *Environmental Issues: Energy Use and Conservation* [Online]. Canberra: Australian Bureau of Statistics. Available: http://www.abs.gov.au/ausstats/abs@.nsf/mf/4602.0.55.001 [Accessed].

Staunton, D. E., Marlin, S. D., Stratowa, C., Dustin, M. L. & Springer, T. A. 1988. Primary structure of ICAM-1 demonstrates interaction between members of the immunoglobulin and integrin supergene families. *Cell,* 52**,** 925-33.

Su, Y. C., Mukherjee, O., Singh, B., Hallgren, O., Westergren-Thorsson, G., Hood, D. & Riesbeck, K. 2016. Haemophilus influenzae P4 Interacts With Extracellular Matrix Proteins Promoting Adhesion and Serum Resistance. *J Infect Dis,* 213**,** 314-23.

Subbanagounder, G., Leitinger, N., Shih, P. T., Faull, K. F. & Berliner, J. A. 1999. Evidence that phospholipid oxidation products and/or platelet-activating factor play an important role in early atherogenesis : in vitro and In vivo inhibition by WEB 2086. *Circ Res,* 85**,** 311-8.

Sumagin, R., Robin, A. Z., Nusrat, A. & Parkos, C. A. 2014. Transmigrated neutrophils in the intestinal lumen engage ICAM-1 to regulate the epithelial barrier and neutrophil recruitment. *Mucosal Immunol,* 7**,** 905-15.

Sunyer, J. 2001. Urban air pollution and chronic obstructive pulmonary disease: a review. *Eur Respir J,* 17**,** 1024-33.

Suri, R., Periselneris, J., Lanone, S., Zeidler-Erdely, P. C., Melton, G., Palmer, K. T., Andujar, P., Antonini, J. M., Cohignac, V., Erdely, A., Jose, R. J., Mudway, I., Brown, J. & Grigg, J. 2016. Exposure to welding fumes and lower airway infection with Streptococcus pneumoniae. *J Allergy Clin Immunol,* 137**,** 527-534.e7.

Sussan, T. E., Ingole, V., Kim, J. H., Mccormick, S., Negherbon, J., Fallica, J., Akulian, J., Yarmus, L., Feller-Kopman, D., Wills-Karp, M., Horton, M. R., Breysse, P. N., Agrawal, A., Juvekar, S., Salvi, S. & Biswal, S. 2014. Source of biomass cooking fuel determines pulmonary response to household air pollution. *Am J Respir Cell Mol Biol,* 50**,** 538-48.

Swanson, K. J., Kado, N. Y., Funk, W. E., Pleil, J. D., Madden, M. C. & Ghio, A. J. 2009. Release of the pro-inflammatory markers by BEAS-2B cells following in vitro exposure to biodiesel extracts. *Open Toxicol J,* 3**,** 8-15.

Swords, W. E., Buscher, B. A., Ver Steeg Ii, K., Preston, A., Nichols, W. A., Weiser, J. N., Gibson, B. W. & Apicella, M. A. 2000. Non-typeable Haemophilus influenzae adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol,* 37**,** 13-27.

Swords, W. E., Jones, P. A. & Apicella, M. A. 2003. The lipo-oligosaccharides of Haemophilus influenzae: an interesting array of characters. *Journal of endotoxin research,* 9**,** 131-144.

Swords, W. E., Moore, M. L., Godzicki, L., Bukofzer, G., Mitten, M. J. & Voncannon, J. 2004. Sialylation of lipooligosaccharides promotes biofilm formation by nontypeable Haemophilus influenzae. *Infect Immun,* 72**,** 106-13.

Tager, I., Tishler, P. V., Rosner, B., Speizer, F. E. & Litt, M. 1978. Studies of the familial aggregation of chronic bronchitis and obstructive airways disease. *Int J Epidemiol,* 7**,** 55-62.

Tager, I. B., Rosner, B., Tishler, P. V., Speizer, F. E. & Kass, E. H. 1976. Household aggregation of pulmonary function and chronic bronchitis. *Am Rev Respir Dis,* 114**,** 485-92.

Tamimi, A., Serdarevic, D. & Hanania, N. A. 2012. The effects of cigarette smoke on airway inflammation in asthma and COPD: therapeutic implications. *Respir Med,* 106**,** 319-28.

Tamura, G., Takishima, T., Mue, S., Makino, S., Itoh, K., Miyamoto, T., Shida, T. & Nakajima, S. 1996. Effect of a potent platelet-activating factor antagonist, WEB-2086, on asthma. A multicenter, double-blind placebo-controlled study in Japan. *Adv Exp Med Biol,* 416**,** 371-80.

Tamura, K., Nei, M. & Kumar, S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A,* 101**,** 11030-5.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res,* 29**,** 22-8.

Terashita, Z.-I., Tsushima, S., Yoshioka, Y., Nomura, H., Inada, Y. & Nishikawa, K. 1983. CV-3988-a specific antagonist of platelet activating factor (PAF). *Life Sciences,* 32**,** 1975-1982.

Thanavala, Y. & Lugade, A. A. 2011. Role of nontypeable Haemophilus influenzae in otitis media and chronic obstructive pulmonary disease. *Adv Otorhinolaryngol,* 72**,** 170-5.

Thomsen, M., Nordestgaard, B. G., Vestbo, J. & Lange, P. 2013. Characteristics and outcomes of chronic obstructive pulmonary disease in never smokers in Denmark: a prospective population study. *The Lancet Respiratory medicine,* 1**,** 543-550.

Torres-Cruz, J. & Van Der Woude, M. W. 2003. Slipped-strand mispairing can function as a phase variation mechanism in Escherichia coli. *J Bacteriol,* 185**,** 6990-4.

Tosi, M. F., Stark, J. M., Smith, C. W., Hamedani, A., Gruenert, D. C. & Infeld, M. D. 1992. Induction of ICAM-1 expression on human airway epithelial cells by inflammatory cytokines: effects on neutrophil-epithelial cell adhesion. *Am J Respir Cell Mol Biol,* 7**,** 214-21.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M. E., Frapy, E., Garry, L., Ghigo, J. M., Gilles, A. M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M. A., Pichon, C., Rouy, Z., Ruf, C. S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E. P. & Denamur, E. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet,* 5**,** e1000344.

Trosini-Desert, V., Germaud, P. & Dautzenberg, B. 2004. Tobacco smoke and risk of bacterial infection. *Revue des maladies respiratoires,* 21**,** 539-547.

Tsoupras, A. B., Iatrou, C., Frangia, C. & Demopoulos, C. A. 2009. The implication of platelet activating factor in cancer growth and metastasis: potent beneficial role of PAF-inhibitors and antioxidants. *Infect Disord Drug Targets,* 9**,** 390-9.

Ulvestad, B., Bakke, B., Eduard, W., Kongerud, J. & Lund, M. B. 2001. Cumulative exposure to dust causes accelerated decline in lung function in tunnel workers. *Occup Environ Med,* 58**,** 663-9.

Van Alphen, L., Caugant, D. A., Duim, B., O'rourke, M. & Bowler, L. D. 1997. Differences in genetic diversity of nonecapsulated Haemophilus influenzae from various diseases. *Microbiology,* 143**,** 1423-1431.

Van Wessel, K., Rodenburg, G. D., Veenhoven, R. H., Spanjaard, L., Van Der Ende, A. & Sanders, E. A. 2011. Nontypeable Haemophilus influenzae invasive disease in The Netherlands: a retrospective surveillance study 2001-2008. *Clin Infect Dis,* 53**,** e1-7.

Vanker, A., Barnett, W., Workman, L., Nduru, P. M., Sly, P. D., Gie, R. P. & Zar, H. J. 2017. Early-life exposure to indoor air pollution or tobacco smoke and lower respiratory tract illness and wheezing in African infants: a longitudinal birth cohort study. *Lancet Planet Health,* 1**,** e328-e336.

Ventola, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *P t,* 40**,** 277-83.

Viniol, C. & Vogelmeier, C. F. 2018. Exacerbations of COPD. *European Respiratory Review,* 27**,** 170103.

Wain, L. V., Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., et al. 2017. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet,* 49**,** 416-425.

Wallace Jr, R., Musher, D., Septimus, E., Mcgowan Jr, J., Quinones, F., Wiss, K., Vance, P. & Trier, P. 1981. Haemophilus influenzae infections in adults: characterization of strains by serotypes, biotypes, and β-lactamase production. *Journal of Infectious Diseases,* 144**,** 101-106.

Walser, A., Flynn, T., Mason, C., Crowley, H., Maresca, C. & O'donnell, M. 1991. Thienotriazolodiazepines as platelet-activating factor antagonists. Steric limitations for the substituent in position 2. *J Med Chem,* 34**,** 1440-6.

Wang, C., Xu, J., Yang, L., Xu, Y., Zhang, X., Bai, C., Kang, J., Ran, P., Shen, H., Wen, F., Huang, K., Yao, W., Sun, T., Shan, G., Yang, T., Lin, Y., Wu, S., Zhu, J., Wang, R., Shi, Z., Zhao, J., Ye, X., Song, Y., Wang, Q., Zhou, Y., Ding, L., Yang, T., Chen, Y., Guo, Y., Xiao, F., Lu, Y., Peng, X., Zhang, B., Xiao, D., Chen, C. S., Wang, Z., Zhang, H., Bu, X., Zhang, X., An, L., Zhang, S., Cao, Z., Zhan, Q., Yang, Y., Cao, B., Dai, H., Liang, L. & He, J. 2018. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. *Lancet,* 391**,** 1706-1717.

Weber, K. H. & Heuer, H. O. 1989. Hetrazepines as antagonists of platelet activating factor. *Med Res Rev,* 9**,** 181-218.

Weinmann, S., Vollmer, W. M., Breen, V., Heumann, M., Hnizdo, E., Villnave, J., Doney, B., Graziani, M., Mcburnie, M. A. & Buist, A. S. 2008. COPD and occupational exposures: a case-control study. *J Occup Environ Med,* 50**,** 561-9.

World Health Organisation. 2006. *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide; Global update 2005* [Online]. Geneva: World Health Organisation. Available: http://apps.who.int/iris/bitstream/handle/10665/69477/WHO_SDE_PHE_OEH_06.02_eng.pdf?sequence=1 [Accessed 06/08/2018].

World Health Organisation. 2014. *Global Health Observatory Map Gallery; Map on "World : Proportion of population with primary reliance on clean fuels (%), 2014"* [Online]. Geneva: World Health Organisation. Available: http://gamapserver.who.int/mapLibrary/Files/Maps/Global_CleanFuels_2014.png [Accessed 31/08/2018].

World Health Organisation. 2016. *Household air pollution and health Fact sheet number 292 updated in February 2016* [Online]. World Health Organisation. Available: http://www.who.int/mediacentre/factsheets/fs292/en/ [Accessed 06/08/2018].

World Health Organisation. 2018a. *Ambient (outdoor) air quality and health* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health [Accessed 28/07/2018].

World Health Organisation. 2018b. *Disease burden and mortality estimates. Cause-specific mortality, 2000–2016* [Online]. Geneva World Health Organisation. Available: http://www.who.int/healthinfo/global_burden_disease/estimates/en/ [Accessed 11/06/2018].

World Health Organisation. 2018c. *Household air pollution and health* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health [Accessed 06/08/2018].

World Health Organisation. 2018d. *The top 10 causes of death* [Online]. Geneva: World Health Organisation. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Accessed 28/07/2018].

World Health Organisation. 2019. *Tobacco* [Online]. Geneva: World Health Organisation. Available: https://www.who.int/news-room/fact-sheets/detail/tobacco [Accessed 04/11/2019].

Williams, B. J., Morlin, G., Valentine, N. & Smith, A. L. 2001. Serum Resistance in an Invasive, NontypeableHaemophilus influenzae Strain. *Infection and immunity,* 69**,** 695-705.

Wilson, B. A., Garud, N. R., Feder, A. F., Assaf, Z. J. & Pennings, P. S. 2016. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. *Mol Ecol,* 25**,** 42-66.

Wilson, M. T. & Hamilos, D. L. 2014. The nasal and sinus microbiome in health and disease. *Current allergy and asthma reports,* 14**,** 485.

Wong, S. & Akerley, B. J. 2012. Genome-scale approaches to identify genes essential for Haemophilus influenzae pathogenesis. *Frontiers in cellular and infection microbiology,* 2**,** 23.

Wordley, J., Walters, S. & Ayres, J. G. 1997. Short term variations in hospital admissions and mortality and particulate air pollution. *Occup Environ Med,* 54**,** 108-16.

Wright, J. L., Harrison, N., Wiggs, B. & Churg, A. 1988. Quartz but not iron oxide causes air-flow obstruction, emphysema, and small airways lesions in the rat. *Am Rev Respir Dis,* 138**,** 129-35.

Xing, Y. F., Xu, Y. H., Shi, M. H. & Lian, Y. X. 2016. The impact of PM2.5 on the human respiratory system. *J Thorac Dis,* 8**,** E69-74.

Yang, Y., Mao, J., Ye, Z., Li, J., Zhao, H. & Liu, Y. 2017. Risk factors of chronic obstructive pulmonary disease among adults in Chinese mainland: A systematic review and meta-analysis. *Respir Med,* 131**,** 158-165.

Yu, Y., Zhang, M., Zhang, X., Cai, Q., Hong, S., Jiang, W. & Xu, C. 2014a. Synergistic effects of combined platelet-activating factor receptor and epidermal growth factor receptor targeting in ovarian cancer cells. *Journal of hematology & oncology,* 7**,** 39.

Yu, Y., Zhang, X., Hong, S., Zhang, M., Cai, Q., Zhang, M., Jiang, W. & Xu, C. 2014b. The expression of platelet-activating factor receptor modulates the cisplatin sensitivity of ovarian cancer cells: a novel target for combination therapy. *Br J Cancer,* 111**,** 515-24.

Zalacain, R., Sobradillo, V., Amilibia, J., Barron, J., Achotegui, V., Pijoan, J. I. & Llorente, J. L. 1999. Predisposing factors to bacterial colonization in chronic obstructive pulmonary disease. *Eur Respir J,* 13**,** 343-8.

Zanobetti, A., Bind, M. A. & Schwartz, J. 2008. Particulate air pollution and survival in a COPD cohort. *Environ Health,* 7**,** 48.

Zar, H. J. & Ferkol, T. W. 2014. The global burden of respiratory disease-impact on child health. *Pediatr Pulmonol,* 49**,** 430-4.

Zhou, Y., Zou, Y., Li, X., Chen, S., Zhao, Z., He, F., Zou, W., Luo, Q., Li, W., Pan, Y., Deng, X., Wang, X., Qiu, R., Liu, S., Zheng, J., Zhong, N. & Ran, P. 2014. Lung function and incidence of chronic obstructive pulmonary disease after improved cooking fuels and kitchen ventilation: a 9-year prospective cohort study. *PLoS Med,* 11**,** e1001621.

Zou, Y., Li, S., Zou, W., Hu, G., Zhou, Y., Peng, G., He, F., Li, B. & Ran, P. 2014. Upregulation of gelatinases and epithelial-mesenchymal transition in small airway remodeling associated with chronic exposure to wood smoke. *PLoS One,* 9**,** e96708.

8 Appendix: Publications

It is the following published article: Kc R., Shukla, S. D., Walters, E. H., O'Toole, R. F. 2017. Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease, Microbiology, 163(4), 421-30

KC et al. Clin Trans Med (2018) 7:39
https://doi.org/10.1186/s40169-018-0217-2

Clinical and Translational Medicine

REVIEW

Open Access

The role of environmental exposure to non-cigarette smoke in lung disease

Rajendra KC[1], Shakti D. Shukla[3], Sanjay S. Gautam[1], Philip M. Hansbro[3,4] and Ronan F. O'Toole[1,2*]

Abstract

Chronic exposure to household indoor smoke and outdoor air pollution is a major contributor to global morbidity and mortality. The majority of these deaths occur in low and middle-income countries. Children, women, the elderly and people with underlying chronic conditions are most affected. In addition to reduced lung function, children exposed to biomass smoke have an increased risk of developing lower respiratory tract infections and asthma-related symptoms. In adults, chronic exposure to biomass smoke, ambient air pollution, and opportunistic exposure to fumes and dust are associated with an increased risk of developing chronic bronchitis, chronic obstructive pulmonary disease (COPD), lung cancer and respiratory infections, including tuberculosis. Here, we review the evidence of prevalence of COPD in people exposed to non-cigarette smoke. We highlight mechanisms that are likely involved in biomass-smoke exposure-related COPD and other lung diseases. Finally, we summarize the potential preventive and therapeutic strategies for management of COPD induced by non-cigarette smoke exposure.

Keywords: Non-cigarette smoke, Biomass smoke, Occupational exposure, Air pollution, Lung disease, Chronic obstructive pulmonary disease (COPD)

Introduction

Worldwide, in 2016, approximately 9 million deaths were attributed to lung diseases, including COPD, lung cancer and lower respiratory tract infections, among which 82.4% occurred in low and middle-income countries [1]. Tobacco smoking is one of the most well-characterized risk factors for lung disease development, killing more than 7 million people each year [2]. However, compared to 1.1 billion tobacco smokers globally, nearly 3 billion people are exposed to biomass smoke [3, 4]. Moreover, data from epidemiologic studies conducted in Asia, Europe, South America and Africa have consistently shown associations between biomass smoke exposure and lung diseases, even after controlling for the primary risk factor, tobacco smoking [5, 6]. Here, we review the association of lung diseases, including COPD, with exposure to biomass smoke, fumes, dust, gases and outdoor air pollution.

Methods

We conducted a qualitative examination of the published peer-reviewed literature from 1980 to June 2018 using the Medline database (https://www.nlm.nih.gov/bsd/medline.html). We used the search terms "biomass fuel", "biomass smoke", "wood smoke", "solid fuel", "coal smoke", "indoor smoke", household pollutants", "outdoor air pollution", "occupational exposure" combined using the Boolean logic 'OR'. Similarly, the search terms "lung disease", "pulmonary disease", "airway disease", "bronchitis", "emphysema", "asthma", "lung cancer", "acute respiratory tract infection" were also combined using the Boolean logic 'OR'. The two subsets were combined using the Boolean logic 'AND', and the results were restricted to meta-analyses, systematic reviews, observational studies, comparative studies and original articles published in English. The search resulted in 1953 articles that were screened from which 168 studies were selected as being relevant to the review.

*Correspondence: ronan.otoole@utas.edu.au
[1] School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia
Full list of author information is available at the end of the article

Prevalence and burden of biomass smoke exposure

Annually, nearly 4 million people die prematurely from illness attributable to household biomass smoke exposure, among which an estimated 55% die due to respiratory diseases, including pneumonia (lower respiratory tract infections), COPD and lung cancer [7]. Nearly half of the world's population rely on biomass, such as wood, animal dung, and crop residues as a primary fuel source for cooking and heating purposes [7]. The proportion of households using clean fuels (liquid petroleum gas, biogas and electricity) varies considerably across the globe (and even in the same continent) [8]. The access to clean fuels is especially limited to relatively smaller populations in the low- and middle-income countries (Fig. 1). Consequently, populations residing in the regional and rural households in developing countries are heavily reliant on biomass fuel for cooking or heating purposes (Fig. 1) [9].

The particular concern with energy from biomass fuel is the use of inefficient stoves for combustion, which generates toxic gases like carbon monoxide and nitrogen oxides; suspended particulate matter containing volatile organic compounds (VOCs) such as methane, aldehydes, benzene and its derivatives; and polycyclic aromatic hydrocarbons (PAHs) like benzo[*a*]pyrene and anthracene [10]. Particulate matter (PM) with an aerodynamic diameter of < 2.5 microns ($PM_{2.5}$) is light and can remain suspended in the air for longer periods [11]. These particles can be inhaled deep into the lungs, and have been linked to oxidative stress and inflammation induced damage of the respiratory system [11]. Moreover, in developing countries, the cultural practice of indoor cooking in housing with poor air ventilation potentially exposes women and children to $PM_{2.5}$ up to the levels that are 1000 times higher than the threshold recommended by WHO (25 μg/m^3) [12, 13].

Wood is the most common biomass fuel, however, use of animal dung, such as cow, sheep and horse, as a source of fuel is also widespread, especially in rural areas of low- and middle-income countries, including

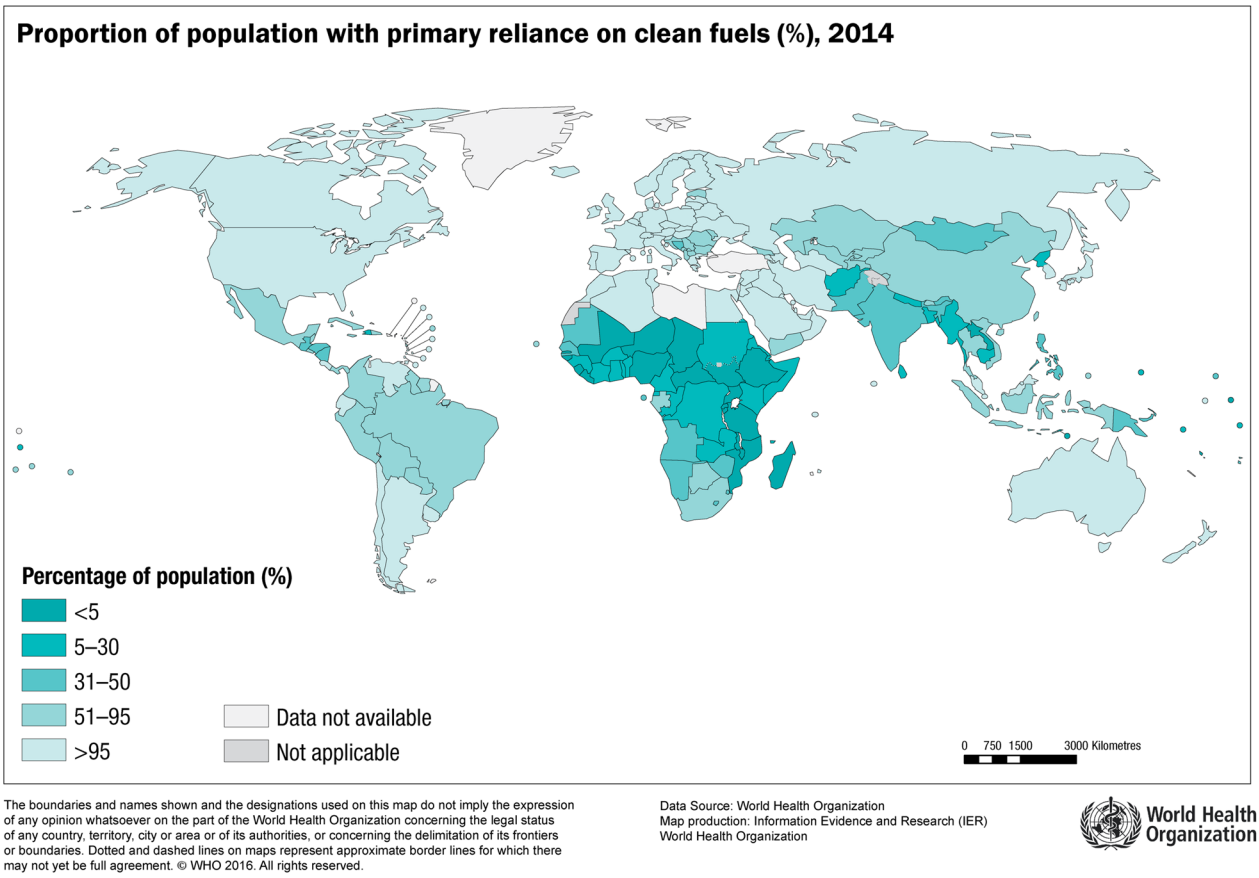


Fig. 1 Global use of clean fuels in 2014, by the World Health Organization. Countries with the lowest (< 5%) and the highest (> 95%) proportion of people using clean fuels as the primary domestic source of energy are shaded dark and light blue, respectively. Reproduced with permission from the World Health Organization [8], Copyright (2016)

India, Nepal, and sub-Saharan Africa due to its availability in areas with limited vegetation and its lower cost [14]. Despite this, animal dung is the least efficient biofuel and burns faster as compared to wood [15]. Relative to wood smoke, combustion of animal dung produces more particulate matter (23% more $PM_{2.5}$ per kilogram), toxic byproducts, such as PAHs and oxidizing species, such as redox active metals (copper and iron) and quinones [16, 17].

Nonetheless, exposure to biomass smoke is not exclusively an issue in low- and middle-income countries. Use of indoor wood fires for heating purposes and for imparting flavor during cooking processes, such as barbecuing and wood-smoking of food, is becoming more popular in high-income countries, thus, increasing biomass smoke exposure [18]. In 2014, in a survey carried out in Australia, approximately 10% of households used wood as the main source of heating [19]. However, in developed countries, use of biomass fuel is primarily seasonal and exposure is largely limited by better ventilation [20]. Besides indoor biomass smoke exposure, people in developed countries such as Australia, Canada and the USA are also exposed to outdoor biomass smoke from frequent bushfires [21]. Importantly, individuals exposed to biomass smoke are more likely to have respiratory symptoms and reduced lung function [22–24].

Association of biomass smoke exposure with respiratory diseases

Chronic airway disease

Worldwide, COPD claimed more than 3 million deaths in 2016, among which 81.8% occurred in low- and middle-income countries [1]. Tobacco smoking is still the main causative factor of COPD in high-income countries where only 15–25% of COPD cases are never smokers (Table 1). However, in low and middle countries more than 45% of COPD patients have been found to be never smokers (Table 1). The existing literature provides strong evidence that smoke from biomass fuels is an independent risk factor for the development of COPD, particularly in low- and middle-income countries where the reliance on biomass fuel is still very high. The noxious particles in biomass smoke induce an inflammatory response through: upregulation of pro-inflammatory cytokines such as interleukin 6 (IL-6), tumor necrosis factor alpha (TNFα), and granulocyte colony stimulatory factor (G-CSF); recruitment of immune cells, such as macrophages and neutrophils; upregulation of gelatinases (matrix metalloproteinase 2 and 9); and epithelial-mesenchymal transition (EMT), thereby reducing lung function and contributing to the onset/progression of COPD [25–27]. Da Silva et al. in a case–control study in Brazil, evaluated the effect of exposure to biomass combustion $PM_{2.5}$ on lung function [22]. There was a significant loss in pulmonary function in non-smoker biomass users compared to non-smoker liquefied petroleum gas (LPG) users (forced expiratory volume in one second to

Table 1 Proportion of never smokers among COPD patients and associated risk factors

Study center and design	Participants	Age (years)	Proportion of never smokers among COPD patients (%)			Risk factors for COPD in never-smokers	References
			Overall	Male	Female		
Multinational (35 centres, 16 countries; ECRHS)	17,966	20–44	17.0	13.4	21.6	Occupational exposure to vapours, gas, dust, or fumes	Cerveri et al. [33]
Malataya, Turkey (CS)	1160	> 18	22.5			Exposure to biomass smoke	Gunen et al. [34]
China (CPH; nationwide CS)	50,991	> 20	50.5	10.1	91.7	Exposure to biomass fuel smoke and $PM_{2.5}$, parental history of respiratory disease	Wang et al. [28]
South Africa (nationwide survey)	13,826	> 18	47·6	24.8	61.0	Biomass fuel, occupational exposure, history of pulmonary tuberculosis	Ehrlich et al. [35]
Maswa, Tanzania (CS)	869	> 35	62.1	–	–		Magitta et al. [36]
Västra Götaland and Norrbotten, Sweden (CS)	1839	21–78	21	–	–	Occupational exposure to gas, dust and fumes	Hagstad et al. [37]
Copenhegen, Denmark (CS)	68,501	20–100	22.3	18.9	25.5		Thomsen et al. [38]
Multinational 12 countries; population-based survey)	73,745	> 40	36 (24 USA-64 Mexico)	21	49		Landis et al. [39]

ECRHS European Community Respiratory Health Survey; *CS* cross-sectional study; *CPH* China Pulmonary Health; *COPD* chronic obstructive pulmonary disease

forced vital capacity ratio (FEV_1/FVC) 0.79 versus 0.85, $p<0.05$, respectively). In addition, pulmonary function was negatively correlated with the level and duration of $PM_{2.5}$ exposure (FEV_1/FVC: $r=-0.63$, $p<0.05$ and -0.52, $p<0.05$, respectively) [22]. A minor reduction in FEV_1/FVC was observed in women in Mexico using biomass fuel compared to clean gas (0.80 vs 0.83, $p=0.03$, respectively) [24].

In a recent nationwide cross-sectional study conducted in 50,991 individuals from ten provinces of China, more than half of the COPD patients were never smokers [28]. Importantly, the proportion of never smokers in female COPD patients was markedly high (91.7%). Additionally, individuals exposed to indoor biomass smoke were 62% more likely to develop COPD than unexposed individuals who used clean fuel sources (age and sex adjusted OR = 1.6, $p=0.003$) [28].

A meta-analysis of 11 cross-sectional and four case–control studies covering a wide range of countries identified household biomass smoke exposure as an independent risk factor for developing COPD in both men (OR 4.30, 95% CI 1.85–10.01) and women (OR 2.73, 95% CI 2.28–3.28), and in both the Asian population (OR 2.31, 95% CI 1.41–3.78) and the non-Asian population (OR 2.56, 95% CI 1.71–3.83) [5]. This finding is supported by another recent cross-sectional study, the PUMA (Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America), which assessed 1740 individuals from multiple nations in South America, who were at greater risk of developing COPD [6]. The PUMA study reported that individuals exposed to household biomass smoke are twice as likely to develop COPD than unexposed people (adjusted OR 2.28, 95% CI 1.18–4.41) [6]. Moreover, a systematic review of 24 epidemiological studies revealed that household biomass smoke exposure was associated with COPD development in both urban (OR 1.6, 95% CI 1.2–2.0) and rural women (OR 2.0, 95% CI 1.5–2.8) [29].

In terms of COPD phenotypes, biomass smoke exposed individuals were found to have more lung fibrosis and bronchiolitis and less emphysema whereas tobacco smokers exhibited higher levels of emphysema (radiologist CT score—a measure of the extent of emphysema based on computed tomography (CT) at inspiration and expiration) 0.7 versus 2.3, $p=0.001$; emphysema on CT 19% versus 27%, $p=0.046$) [30, 31]. This finding is further supported by Fernandes et al. who reported a significantly lower level of emphysema in biomass exposed women than in women who were tobacco smokers (PRM^{Emph} (parametric response mapping, an image tool used to quantify emphysema based on paired CT images at inspiration and expiration) 1.84% (0.69 ± 3.72%) versus 9.85% (2.40–16.34%); $p=0.001$) [32].

Lung cancer

Globally, 1.7 million deaths in 2016 were attributed to lung cancer, among which 66% occurred in low- and middle-income countries [1]. The combustion products of biomass fuel emissions such as dibenz[*a,h*]anthracene, cyclopenta[*cd*]pyrene and 1,3-butadiene have been grouped as a probable human carcinogen (group 2A) by the International Agency for Research on Cancer [40]. Saldana and colleagues, in a hospital based case–control study with 136 cases of primary lung cancer and 137 controls, reported a significant association between the magnitude of biomass smoke exposure in hour-years (years of exposure multiplied by average hours of exposure per day) and the risk of lung cancer after adjusting for sex, smoking, socioeconomic status and housing with asbestos sheet roof (OR for an exposure > 300 h-years 3.01, 95% CI 1.12–8.36) [41]. A systematic review including 13 case control studies summarized an increased risk (OR 1.17, 95% CI 1.01–1.37) of lung cancer with biomass exposure associated with cooking/heating, in particular, among women in developing countries [42]. In a retrospective cohort study conducted in Xuanwei city of China, mortality from lung cancer was compared between lifelong users of smoky coal (27,310 individuals) and smokeless coal (9962 individuals). Individuals exposed to smoky coal were at greater risk of death from lung cancer when compared to those exposed to smokeless coal (for men, HR 36, 95% CI 20–65; for women, 99, 95% CI 37–266) [43]. A similar observation was reported in a meta-analysis from 28 epidemiological studies where a significantly higher risk of lung cancer was observed in females (OR 1.81, 95% CI 1.54–2.12; $p=0.034$) exposed to biomass/solid fuel smoke compared to males (OR 1.16, 95% CI 0.79–1.69) with related exposure [44]. The increased susceptibility of women to biomass/coal carcinogens compared to men may be due to differences in base-line exposure (women being exposed to biomass/coal smoke for extended periods of time) or due to differences in lung size (women having smaller lungs than men, hence more damage done for the similar amount of smoke inhaled) [45]. The association was even more prominent in individuals using coal (OR 1.82, 95% CI 1.60–2.06) compared to biomass (OR 1.50, 95% CI 1.17–1.94), which may be due to the production of predominant Group 1 carcinogenic PAHs during coal combustion [44, 46]. Although the studies included in this review have either been adjusted for smoking or studied a population of non-smokers, none of these studies commented on the presence of other co-morbid conditions, such as COPD, at the time of lung cancer diagnosis.

Lower respiratory tract infections

In addition to COPD and lung cancer, the other major respiratory cause of global mortality is lower respiratory tract infection (LRI) and tuberculosis, which was directly implicated in 2.96 and 1.29 million deaths in 2016, respectively [1]. Of the approximate 4 million deaths globally, around 89% of the mortality was reported in low- and middle-income countries alone [47]. An evaluation of 14 studies in developing countries indicates that young children living in households where biomass fuels are utilized, have a two to three times higher risk of respiratory infections than unexposed children after adjustment of potential confounding factors [48, 49]. A recent systematic review of 77 studies from 39 low- and middle-income countries evaluated the risk factors for mortality from acute lower respiratory infections in children under 5 years of age and found that biomass smoke exposure is significantly associated with an increased risk of death from LRI (OR 3.0, 95% CI 2.1–4.3) [50]. However, only a small number of studies have focused on the association between biomass smoke exposure and risk of lower respiratory tract infection in the adult population. Ezzati et al. evaluated the risk of LRI in 229 individuals between 5 and 49 years of age in Central Kenya upon exposure to biomass fuel derived PM_{10} [51]. The risk of LRI was positively correlated with the level of PM_{10} exposure, however, the odds ratio was significant only above 2000 μg/m^3 (adjusted OR 3.3, 95% CI 1.1–9.9) [51]. Exposure to household biomass smoke was associated with risk of hospitalization due to pneumonia in older adults aged 65 and above (adjusted OR 3.3, 95% CI 1.6–6.9) [52]. On the other hand, although the prevalence of acute respiratory infection was higher among biomass users (15–45%), after controlling for major confounding factors such as age and smoking, the risk of LRI was not significant when including adults aged less than 65 [53–55]. In addition, biomass smoke exposure is significantly associated with the risk of developing tuberculosis. For example, a close correlation between biomass fuel usage and tuberculosis (OR 3.14, 95% CI 1.15–8.56) has been reported from northern India compared to usage of liquid petroleum gas in kitchens after adjusting for confounding variables, such as tobacco smoking and close contact with TB cases (6).

Other respiratory conditions

Several studies have highlighted the effect of biomass smoke exposure in respiratory symptoms, such as cough, wheeze, mucus overproduction and dyspnoea [22–24]. A case–control study in Brazil reported an increased risk of developing chronic cough (OR 2.9, 95% CI 1.68–5.10), wheeze (OR 2.33, 95% CI 1.25–4.38) and dyspnoea (OR 2.59, 95% CI 1.32–5.09) in adults exposed to biomass smoke as compared to LPG users [22]. Similarly, a study in Eastern India assessed the impact of chronic exposure to indoor biomass smoke on respiratory health in 681 non-smoking women and 438 age matched women not exposed to biomass smoke [23]. Compared to LPG users, biomass users had a higher prevalence of cough, mucus production and wheeze (71.8% versus 30.8%, $p=0.001$), and dyspnoea (58.4% versus 19.9%, $p=0.001$) [23].

Besides lower respiratory tract symptoms, other lung diseases, such as asthma and pneumoconiosis have been correlated with exposure to indoor biomass smoke. Oluwole et al. reported an increased prevalence of asthma-related symptoms, such as wheezing or whistling in the chest, cough, dyspnoea and chest tightness in rural school children in Nigeria exposed to biomass smoke (OR 2.37, 95% CI 1.16–4.84) [56]. However, evidence of its association with asthma is inconclusive. A meta-analysis of 25 studies did not find any significant association of biomass smoke exposure and asthma in children (OR 0.5, 95% CI 0.1–2.0) or women (1.34, 95% CI 0.9–1.9) [57]. However, a significant difference has been reported between bronchiectasis (diagnosed by computed tomography scans) in women and exposure to biomass fuel, compared to exposure to tobacco smoke (14% versus 0%, $p=0.009$) [30]. Particularly, the "Hut lung disease" or domestically acquired pneumoconiosis has been noted as a particulate lung disease in women exposed to biomass fuel smoke or agricultural activities but not associated with mining [58–60].

Exposure to other types of air pollutants

Environmental exposure to noxious particles and gases

High levels of air pollution have also been implicated as risk factors for the development of COPD, although it is less potent than active smoking. Approximately 25–60% of COPD patients are never-smokers in developing countries [28, 36, 61]. Increasing population, urbanization, economic profile and pollution are several factors that contribute substantially to the COPD burden [62]. The role of outdoor air pollution (including traffic-related fine particulate matter) as a causative factor for airflow limitation is gaining attention in recent times, due to increased vehicular pollution, as well as industrialization of the two most populated countries, China and India [63].

$PM_{2.5}$ is considered to be the most health-damaging of the particulate matter, as it can penetrate deep into the lungs and initiate deleterious effects on the airway, including but not limited to airway oxidative stress, pulmonary and systemic inflammation, ciliary dysfunction, amplification of infections, and increases in bronchial reactivity [64]. The main components of PM_{10} are sulfate, nitrates, ammonia, sodium chloride, black carbon, mineral dust and water [4].

Several investigators have assessed the effect of air pollution as a potential risk factor for COPD. In children and young adults, cross-sectional studies have shown a relationship between higher outdoor pollutant levels (especially traffic related pollution) and lower lung function [65, 66]. Kulkarni et al. reported a likely causative dose-dependent inverse association between the carbon content (as a biomarker of particulate matter exposure, PM_{10}) of airway macrophages with lung function in children [67]. Furthermore, exposure to traffic-related pollution, indicated by residential distance from a highway, was associated with impaired lung growth and lung function deficits at 18 years of age [68]. Also, higher traffic density and proximity to highways was significantly associated with lower lung function (FEV_1) and FVC, but only in females [69], who are also at a greater risk of developing COPD than for those living farther away [70].

Long-term exposure to airborne particles and particulate matter is significantly associated with the risk of premature death and acute care hospitalizations, especially in patients with severe disease [71, 72]. Moreover, daily variation in exposure to outdoor air pollution (mainly the particulate matter) significantly correlates with acute exacerbations of COPD [73]. The mechanisms that underlie obstruction due to air pollution are likely to be similar to those caused by cigarette smoking, but we do not yet have conclusive evidence for this, despite the detrimental effects of air pollution on lung health.

Occupational exposure to dusts and fumes

Several longitudinal studies have shown an association between certain dusty occupational exposures and COPD, i.e., coal mining [74], gold mining [75], work related to tunnel-construction [76], low levels of concrete dust containing crystalline silica exposure (concrete production industries) [77], exposure to cotton in textile industries [78], workers exposed to welding fumes [79], grain handlers and postal workers (exposure to endotoxins) [80] and animal feed industry [81]. Moreover, chronic exposure to metallic dust (primarily, cobalt and chromium) was found to be associated with deterioration of lung function [82]. In addition, exposure to chemical vapors, irritants and fumes can also contribute to accelerated loss of airflow [83]. Another study involving railroad workers reported a positive association between COPD mortality and occupational exposure to diesel exhaust [84]. A population-based study found positive associations between several occupational exposure measures (mineral dusts, metal dusts and fumes, organic dusts, irritant gases or vapors, sensitizers, organic solvents, diesel exhaust, and environmental tobacco smoke) with COPD, among both ever-smokers and never-smokers [85].

Experimental studies in animal models have demonstrated that exposures to several agents, such as sulphur dioxide, mineral dusts, vanadium and endotoxin, are capable of inducing chronic obstructive bronchitis [83]. Intratracheally instilled silica (quartz) produces airflow obstruction (functional change), which correlates with the presence of both emphysema and small-airway lesions [86]. Inorganic dusts containing silica are also associated with neutrophil and macrophage accumulation and morphological changes in the rat lung [86]. These morphological changes in small airways and lung parenchyma were similar to those in rats treated with elastase, which represents a well-established model of experimental COPD [87].

Mechanisms of lung damage due to biomass smoke exposure

Despite the immense burden of chronic airway disease, only a limited number of mechanistic studies have examined the immunomodulatory effects of biomass smoke exposure. Sussan et al. explored the mechanisms of pulmonary responses in mice after acute (6 and 24 h) or subchronic (3 times a week for 8 weeks) exposure to wood or cow dung PM [17]. Acute exposure to wood smoke elicited a stronger pro-inflammatory response, as indicated by increased expression of G-CSF, Keratinocyte chemoattractant (KC), C-X-C motif chemokine 10 (CXCL10), IL-6, TNFα and interleukin 12 p70 subunit (IL12p70) [17]. The induction of pro-inflammatory cytokine expression was higher with cow dung than wood smoke PM [17]. On the other hand, sub-chronic exposures exhibited differences in pulmonary response, where wood smoke elicited an eosinophilic response in contrast to a predominantly neutrophilic response induced by cow dung smoke [17]. The inflammatory response elicited by both wood and cow dung smoke was mediated via nuclear factor kappa B (NF-κB) signaling [17]. In addition, in an in vitro study, c-Jun N-terminal kinase-activator protein-1 (JNK-AP-1) signaling, and not the NFκB regulatory pathway, was found to be involved in mediating inflammatory responses in human primary small airway epithelial cells upon exposure to dung biomass smoke [88].

Moreover, in a case–control study conducted in India with 142 biomass users and 126 age-matched LPG user women, a significant neutrophilic inflammatory response was observed in the biomass using group [89]. Also, compared to LPG users, reactive oxygen species (ROS) generation by leukocytes (in both blood and sputum) and the systemic level of antioxidant enzyme superoxide dismutase (SOD) were higher and lower, respectively in women using biomass fuel [89]. A similar increase in systemic oxidative stress was reported in rats exposed to

biomass smoke, as indicated by an increased plasma level of malondialdehyde and a reduced level of SOD [90]. Several in vitro studies reported depletion of antioxidants, such as ascorbate, urate and reduced glutathione in respiratory tract lining fluid when incubated with wood and animal dung smoke extracts [16, 91]. Therefore, biomass smoke exposure induces inflammation and oxidative stress mediated lung damage which potentially contributes to the development/progression of COPD.

In terms of lung cancer, biomass smoke induced ROS directly damages DNA and promotes lung cell proliferation and turnover, resulting in fibrosis and development of lung tumours in rats [92]. Furthermore, biomass exposure upregulates the production of extracellular matrix proteins, including perlecan and fibronectin with a capacity to induce fibrosis in cultured cells [93]. In addition, PAHs released by incomplete combustion of biomass fuels are evidently carcinogenic in both in vitro and in vivo studies [94–97]. Metabolism of PAHs leads to the formation of active carcinogens diol-epoxides, radical cations and o-quinones [94, 98]. These activated PAH metabolites can form adducts with DNA, resulting in mutations, alteration of gene expression profiles, and tumorigenesis [94]. Gene 'hotspots' for adduct formation by PAH metabolites include oncogenes such as p53, K-ras and the H-ras [99, 100]. PAH-DNA adducts are associated with a two-fold increased risk of lung cancer [101].

Biomass smoke exposure is also implicated in increased susceptibility to bacterial infection through several mechanisms, including alterations in alveolar macrophage phagocytosis and/or upregulation of host surface receptors on the respiratory epithelium [102, 103]. In an in vitro study, the phagosomal function of wood smoke particle exposed human alveolar macrophages was tested with respect to uptake of fluorescently-labelled beads, *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* [102]. Wood smoke exposed macrophages demonstrated reduced phagocytosis of fluorescent beads, *S. pneumoniae* and *M. tuberculosis* with a negative linear correlation between macrophage particulate content and phagocytosis [104]. Furthermore, oxidative stress is believed to upregulate expression of intercellular adhesion molecule-1 (ICAM-1), and platelet activating factor receptors (PAFR) allowing attachment and invasion of respiratory bacteria, including *S. pneumoniae*, *Haemophilus influenzae* and *Pseudomonas aeruginosa,* which are also important bacterial pathogens in COPD [105–107]. Although airway ICAM-1 and PAFR expression were markedly upregulated in tobacco smokers and COPD patients, further studies are warranted to demonstrate the effect of biomass smoke exposure on inducing the expression of these host surface receptors [108–110]. Recently, KC et al. developed a method to generate batches of biomass smoke extracts that can be preserved for longer periods for use in multiple exposure experiments, thus minimizing inter-assay variation [103]. This will facilitate further research on mechanistic role of smoke extracts in the inflammatory response and the pathogenesis of respiratory diseases including COPD, lung cancer and LRI.

Preventive and therapeutic measures

Biomass smoke is one of the major risk factors for the onset/progression of COPD in developing countries [61]. To reduce the exposure to smoke generated from combustion of biomass fuel, several preventive interventions have been recommended, with some already implemented.

Based on previous intervention studies, the use of efficient cooking stoves and the promotion of cleaner fuel use (e.g., biogas, liquefied petroleum gas and electricity) has been effective in preventing biomass smoke exposure, at least to some extent [111–113]. The use of improved cooking stoves (ICS) has been found to reduce the mean $PM_{2.5}$ levels by 63.2% in Nepal, 37.0% Senegal, 32.8% in the Gambia and 18–45% in Kenya [111, 112, 114]. Although the decrease in mean $PM_{2.5}$ levels was significantly high, none of the improved cooking stoves achieved the WHO guideline level of mean kitchen $PM_{2.5}$ of 25 μg/m^3 for 24-h. A recent 9-year prospective cohort study in China evaluated the effect of intervention with improved cooking fuels and kitchen intervention [113]. Use of clean fuels and improved ventilation were associated with a lower decline in FEV_1 (12 mL/year, 4–20 mL/year versus 13 mL/year, 4–23 mL/year) compared to those with neither intervention, after adjusting for confounders [113]. In addition, the duration of improved fuel use and ventilation was negatively correlated with the decline of FEV_1 ($p < 0.05$) [113]. Moreover, the fuel and ventilation improvement intervention were associated with a reduction in the risk of COPD, with an odds ratio of 0.28 (95% CI, 0.11–0.73) [113]. Despite the perception that biomass is a cheap fuel source, the annual health-related cost per household associated with biomass smoke exposure (16.94 USD) in Nepal is 61.3% higher than the annual cost of biogas usage (10.38 USD), an alternative cleaner fuel [14]. In addition, the findings of the study suggest that providing community education and creating rural employment and income generation opportunities are important measures in promoting the sustainable use of clean fuels [14].

In terms of therapeutic intervention, exposure to household and outdoor air pollutants are found to reduce antioxidant defense by decreasing the levels of ascorbate, urate, SOD and reduced glutathione resulting in inflammation of the airways [115]. Antioxidants and

anti-inflammatory drugs are potential therapeutics for the prevention of biomass smoke induced inflammation-mediated lung injury [116, 117]. Based on the evidence that wood and cow dung smoke upregulated the expression of PAFR on human bronchial epithelial cells, we believe that the induction of microbial adhesion receptors on respiratory epithelia could be targeted therapeutically to prevent LRIs. Several in vitro studies have shown that the use of PAFR antagonists such as CV-3988 and WEB-2086 abrogates the adhesion of *S. pneumoniae* and nontypeable *H. influenzae* to cigarette smoke exposed human bronchial and alveolar epithelial cells [118, 119]. In addition, ICAM-1, a major receptor for rhinoviruses and nontypeable *H. influenzae* could be another important target with therapeutic potential [120, 121]. However, we stress the point that further research is needed to elucidate the mechanisms of biomass smoke induced susceptibility to LRI, as well as development of COPD and lung cancer. But while new therapeutics may be possible with regard to reducing the progression or symptoms of biomass smoke-related lung disease, the major focus needs to be maintained on a global reduction in the exposure of human populations to biomass smoke.

Strengths and limitations

There are several strengths of our review. Our narrative review comes at a time when the United Nations and the World Health Organization have specifically focused on the urgent need to limit non-communicable diseases, which has also garnered support from various Heads-of-Countries and Governments [122]. In addition, we have examined the epidemiology of non-cigarette smoke etiologies (biomass smoke, air pollution and multiple occupational exposures) implicated in heightened risk of developing a range of lung diseases, most notably, COPD, lung cancer and respiratory infections. We have reviewed and summarized potential mechanisms involved in the pathogenesis of non-cigarette smoke exposure and lung disease. Finally, we have highlighted the potential preventive, as well as therapeutic strategies to limit lung diseases related to these exposures.

One of the limitations is that there may be variability across epidemiological studies in terms of data collection and analyses. In particular, factors such as geographical locations, sampling points within a city as well as multiple cities in a specific country, the type of sampling (longitudinal versus one time-point sampling) contributes significantly to heterogeneity of results obtained. In addition, participant recruitment criteria (i.e., from general population or community healthcare centers), reliability of self-reported duration of exposure are crucial factors that need to be standardized globally. Notably, variability of spirometry technique and the threshold cut-off for COPD (i.e., $FEV_1/FVC < 0.70$ or $FEV_1/FVC < LLN$) vary across studies. In addition, it is difficult to tease out contributions of individual exposure(s)-related causations and hence the results may represent disease associated with mixed exposures. The heterogeneity in future epidemiological studies could be addressed by utilizing standardized and validated research methodologies/strategies to ascertain burden of lung diseases associated with non-cigarette smoke exposures. We are not concluding direct causation of lung diseases due to non-cigarette smoke exposures in our review; however, our intent was to direct policymakers and researchers towards the possible mechanisms reported to date. Clearly, focused and detailed mechanistic studies are needed to establish direct causation of non-cigarette smoke exposures and lung pathologies, which should be further validated in large-scale epidemiological studies, as well as mechanistic models of experimental animal models. Finally, we limited our analysis to published literature and did not include media/analytical reports from recently initiated preventive strategies (such as Ujjwala yojana in India; [123]). However, this could be followed up once the government reports are released by various countries.

Summary and conclusions

The prevalence of COPD amongst non-smokers is a worrying concern globally, as almost half the World's population is chronically exposed to biomass smoke, air pollution and other exposures (chemicals, dusts, and fumes). Evidently, low socio-economic status drives large sections of societies living in low- and middle-income countries towards biomass fuels for cooking/heating purposes. In addition, poor awareness about respiratory health and lack of implementation of adequate safety measures (e.g., poorly ventilated homes) predispose these individuals to heightened risk of developing biomass smoke-induced lung diseases, including COPD, lung cancer and respiratory infections. Chronic biomass smoke (and other non-cigarette exposures) delivers deleterious compounds (PMs, PAHs etc.) that activate an inflammatory cascade in the lungs, which may potentially spill over systemically. This is coupled with marked immune cell dysfunction, activation of oncogenes and upregulation of microbial receptors in the respiratory tract, which contribute to onset/progression of COPD and other related diseases (lung cancer and infections). Preventive strategies should include awareness about the long-term health benefits of using cleaner fuels over biomass-fuels, use of more efficient cooking equipment, and the importance of adequate ventilation in low socioeconomic localities. Although research has only begun to unravel the underlying molecular and pathological mechanisms

involved in biomass smoke-induced COPD, anti-oxidants, anti-inflammatory and anti-microbial drugs may be promising new therapeutics to limit the effects of non-smoking related COPD.

Abbreviations

AP-1: activator protein 1; CI: confidence interval; COPD: chronic obstructive pulmonary disease; CS: cross-sectional study; CPH: China Pulmonary Health; CT: computed tomography; CXCL10: C-X-C motif chemokine 10; DNA: deoxyribonucleic acid; ECRHS: European Community Respiratory Health Survey; FEV_1: forced expiratory volume in one second; FVC: forced vital capacity; OR: odds ratio; G-CSF: granulocyte colony-stimulating factor; HR: hazard ratio; ICAM-1: intercellular adhesion molecule-1; ICS: improved cook stoves; IL-12p70: interleukin 12 p70 subunit; IL-6: interleukin 6; JNK: c-Jun N-terminal kinase; KC: keratinocyte chemoattractant; LPG: liquefied petroleum gas; LRI: lower respiratory tract infection; NFkB: nuclear factor kappa B; PAFR: platelet activating factor receptor; PAHs: polycyclic aromatic hydrocarbons; PM: particulate matter; PRM: parametric response mapping; PUMA: Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America; ROS: reactive oxygen species; SOD: superoxide dismutase; TB: tuberculosis; TNFα: tumour necrosis factor alpha; VOCS: volatile organic compounds; WHO: World Health Organization.

Authors' contributions

SDS and RFOT conceived the review. RKC, SDS, SSG, and RFOT drafted the review. RKC, SSG and RFOT revised the review for submission. All authors read and approved the final manuscript.

Author details

[1] School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia. [2] Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin, Ireland. [3] Priority Research Centre for Healthy Lungs, Hunter Medical Research Institute, School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, Australia. [4] Centenary Institute and University of Technology Sydney, Sydney, New South Wales, Australia.

Acknowledgements

RKC is a recipient of Tasmania Graduate Research Scholarship.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable to this manuscript.

Consent for publication

This declaration is not applicable to this manuscript.

Ethics approval and consent to participate

This declaration is not relevant to this manuscript.

Publisher's Note

Received: 3 October 2018 Accepted: 23 November 2018
Published online: 05 December 2018

References

1. World Health Organisation (2018) Disease burden and mortality estimates. Cause-specific mortality, 2000–2016. WHO, Geneva. http://www.who.int/healthinfo/global_burden_disease/estimates/en/. Accessed 11 Aug 2018
2. World Health Organisation (2018) Tobacco; key facts. WHO, Geneva. http://www.who.int/en/news-room/fact-sheets/detail/tobacco. Accessed 06 Aug 2018
3. World Health Organisation (2018) Prevalence of tobacco smoking. WHO, Geneva. http://www.who.int/gho/tobacco/use/en/. Accessed 29 July 2018
4. World Health Organisation (2018) Ambient (outdoor) air quality and health. WHO, Geneva. http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. Accessed 06 Aug 2018
5. Hu G, Zhou Y, Tian J, Yao W, Li J, Li B et al (2010) Risk of COPD from exposure to biomass smoke: a metaanalysis. Chest 138(1):20–31
6. Montes de Oca M, Zabert G, Moreno D, Laucho-Contreras ME, Lopez Varela MV, Smoke Surmont F (2017) Biomass exposure, and COPD risk in the primary care setting: The PUMA Study. Respiratory care. 62(8):1058–1066
7. World Health Organisation (2018) Household air pollution and health. WHO, Geneva. http://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health. Accessed 06 Aug 2018
8. Global Health Observatory Map Gallery; Map on "World: Proportion of population with primary reliance on clean fuels (%), 2014". Geneva: World Health Organisation; 2014. http://gamapserver.who.int/mapLibrary/Files/Maps/Global_CleanFuels_2014.png. Accessed 31 Aug 2018
9. Hall DO, Barnard GW, Moss P (2013) Biomass for energy in the developing countries: current role, potential, problems, prospects. Elsevier, Amsterdam
10. Leavey A, Patel S, Martinez R, Mitroo D, Fortenberry C, Walker M et al (2017) Organic and inorganic speciation of particulate matter formed during different combustion phases in an improved cookstove. Environ Res 158:33–42
11. Xing YF, Xu YH, Shi MH, Lian YX (2016) The impact of PM2.5 on the human respiratory system. J Thorac Dis. 8(1):E69–E74
12. World Health Organisation (2006) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide; Global update 2005. WHO, Geneva. http://apps.who.int/iris/bitstream/handle/10665/69477/WHO_SDE_PHE_OEH_06.02_eng.pdf?sequence=1. Accessed 11 Aug 2018
13. Saksena STL, Smith KR (2003) Database of household air pollution studies in developing countries, protection of the human environment. World Health Organization, Geneva
14. Pant KP (2012) Cheaper fuel and higher health costs among the poor in rural Nepal. Ambio 41(3):271–283
15. Pahla G, Mamvura T, Ntuli F, Muzenda E (2017) Energy densification of animal waste lignocellulose biomass and raw biomass. S Afr J Chem Eng 24:168–175
16. Mudway IS, Duggan ST, Venkataraman C, Habib G, Kelly FJ, Grigg J (2005) Combustion of dried animal dung as biofuel results in the generation of highly redox active fine particulates. Part Fibre Toxicol 2:6
17. Sussan TE, Ingole V, Kim JH, McCormick S, Negherbon J, Fallica J et al (2014) Source of biomass cooking fuel determines pulmonary response to household air pollution. Am J Respir Cell Mol Biol 50(3):538–548
18. Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ et al (2007) Woodsmoke health effects: a review. Inhalation Toxicol 19(1):67–106
19. Environmental Issues: Energy Use and Conservation Canberra: Australian Bureau of Statistics; 2014. http://www.abs.gov.au/ausstats/abs@.nsf/mf/4602.0.55.001. Accessed 03 Dec 2014
20. Rogalsky DK, Mendola P, Metts TA, Martin WJ 2nd (2014) Estimating the number of low-income americans exposed to household air pollution from burning solid fuels. Environ Health Perspect 122(8):806–810
21. Reisen F, Brown SK (2009) Australian firefighters' exposure to air toxics during bushfire burns of autumn 2005 and 2006. Environ Int 35(2):342–352
22. da Silva LF, Saldiva SR, Saldiva PH, Dolhnikoff M (2012) Impaired lung function in individuals chronically exposed to biomass combustion. Environ Res 112:111–117
23. Mukherjee S, Roychoudhury S, Siddique S, Banerjee M, Bhattacharya P, Lahiri T et al (2014) Respiratory symptoms, lung function decrement and chronic obstructive pulmonary disease in pre-menopausal Indian women exposed to biomass smoke. Inhalation Toxicol 26(14):866–872
24. Regalado J, Perez-Padilla R, Sansores R, Paramo Ramirez JI, Brauer M, Pare P et al (2006) The effect of biomass burning on respiratory

symptoms and lung function in rural Mexican women. Am J Respir Crit Care Med 174(8):901–905
25. Zou Y, Li S, Zou W, Hu G, Zhou Y, Peng G et al (2014) Upregulation of gelatinases and epithelial-mesenchymal transition in small airway remodeling associated with chronic exposure to wood smoke. PLoS ONE 9(5):e96708
26. Olloquequi J, Silva OR (2016) Biomass smoke as a risk factor for chronic obstructive pulmonary disease: effects on innate immunity. Innate Immunity. 22(5):373–381
27. Ramos D, Proenca M, Leite MR, Ferreira AD, Trevisan IB, Brigida GFS et al (2017) Effects of exposure to biomass burning on pulmonary inflammatory markers and pulmonary function in individuals with COPD. Rev Port de Pneumol 23(5):273–279
28. Wang C, Xu J, Yang L, Xu Y, Zhang X, Bai C et al (2018) Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 391(10131):1706–1717
29. Sana A, Somda SMA, Meda N, Bouland C (2018) Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. BMJ Open Respir Res 5(1):e000246
30. Camp PG, Ramirez-Venegas A, Sansores RH, Alva LF, McDougall JE, Sin DD et al (2014) COPD phenotypes in biomass smoke-versus tobacco smoke-exposed Mexican women. Eur Respir J 43(3):725–734
31. Rivera RM, Cosio MG, Ghezzo H, Salazar M, Perez-Padilla R (2008) Comparison of lung morphology in COPD secondary to cigarette and biomass smoke. Inte J Tuberc Lung Dis 12(8):972–977
32. Fernandes L, Gulati N, Fernandes Y, Mesquita AM, Sardessai M, Lammers JJ et al (2017) Small airway imaging phenotypes in biomass- and tobacco smoke-exposed patients with COPD. ERJ Open Res. 3(2):00124–2016
33. Cerveri I, Accordini S, Verlato G, Corsico A, Zoia MC, Casali L et al (2001) Variations in the prevalence across countries of chronic bronchitis and smoking habits in young adults. Eur Res J 18(1):85–92
34. Gunen H, Hacievliyagil SS, Yetkin O, Gulbas G, Mutlu LC, Pehlivan E (2008) Prevalence of COPD: first epidemiological study of a large region in Turkey. Eur J Intern Med 19(7):499–504
35. Ehrlich RI, White N, Norman R, Laubscher R, Steyn K, Lombard C et al (2004) Predictors of chronic bronchitis in South African adults. Int J Tuberc Lung Dis 8(3):369–376
36. Magitta NF, Walker RW, Apte KK, Shimwela MD, Mwaiselage JD, Sanga AA et al (2018) Prevalence, risk factors and clinical correlates of COPD in a rural setting in Tanzania. Eur Respir J. 51(2):1700182
37. Hagstad S, Backman H, Bjerg A, Ekerljung L, Ye X, Hedman L et al (2015) Prevalence and risk factors of COPD among never-smokers in two areas of Sweden–Occupational exposure to gas, dust or fumes is an important risk factor. Respir Med 109(11):1439–1445
38. Thomsen M, Nordestgaard BG, Vestbo J, Lange P (2013) Characteristics and outcomes of chronic obstructive pulmonary disease in never smokers in Denmark: a prospective population study. Lancet Respir Med. 1(7):543–550
39. Landis SH, Muellerova H, Mannino DM, Menezes AM, Han MK, van der Molen T et al (2014) Continuing to Confront COPD International Patient Survey: methods, COPD prevalence, and disease burden in 2012-2013. Int J Chronic Obstr Pulm Dis 9:597–611
40. International Agency for Research on Cancer (2010) Household use of solid fuels and high-temperature frying. International Agency for Research on Cancer, Lyon
41. Baez-Saldana R, Rumbo-Nava U, Canseco-Raymundo A, Castillo-Gonzalez P, Leon-Duenas S, Rivera-Rosales R et al (2018) 35PD Magnitude of exposure to biomass fuel smoke and risk of lung cancer in women: cases and controls study. J Thorac Oncol 13(4):S20
42. Bruce N, Dherani M, Liu R, Hosgood HD, Sapkota A, Smith KR et al (2015) Does household use of biomass fuel cause lung cancer? A systematic review and evaluation of the evidence for the GBD 2010 study. Thorax. 70:433–441
43. Barone-Adesi F, Chapman RS, Silverman DT, He X, Hu W, Vermeulen R et al (2012) Risk of lung cancer associated with domestic use of coal in Xuanwei, China: retrospective cohort study. BMJ 345:e5414
44. Kurmi OP, Arya PH, Lam K-BH, Sorahan T, Ayres JG (2012) Lung cancer risk and solid fuel smoke exposure: a systematic review and meta-analysis. Eur Respir J 40(5):1228–1237
45. Bellemare F, Jeanneret A, Couture J (2003) Sex differences in thoracic dimensions and configuration. Am J Respir Crit Care Med 168(3):305–312
46. Petrilli FL, Kanitz S (1971) Medico-social aspects of air pollution. Minerva Med 62(18):870–923
47. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, Aboyans V, Adetokunboh O, Afshin A, Agrawal A, Ahmadi A (2017) Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 390(10100):1151–1210
48. Smith KR, Samet JM, Romieu I, Bruce N (2000) Indoor air pollution in developing countries and acute lower respiratory infections in children. Thorax 55(6):518–532
49. Mishra V (2003) Indoor air pollution from biomass combustion and acute respiratory illness in preschool age children in Zimbabwe. Int J Epidemiol 32(5):847–853
50. Sonego M, Pellegrin MC, Becker G, Lazzerini M (2015) Risk factors for mortality from acute lower respiratory infections (ALRI) in children under five years of age in low and middle-income countries: a systematic review and meta-analysis of observational studies. PLoS ONE 10(1):e0116380
51. Ezzati M, Kammen D (2001) Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. Lancet 358(9282):619–624
52. Loeb M, Neupane B, Walter SD, Hanning R, Carusone SC, Lewis D et al (2009) Environmental risk factors for community-acquired pneumonia hospitalization in older adults. J Am Geriatr Soc 57(6):1036–1040
53. Shrestha IL, Shrestha SL (2005) Indoor air pollution from biomass fuels and respiratory health of the exposed population in Nepalese households. Int J Occup Environ Health 11(2):150–160
54. Taylor ET, Nakai S (2012) Prevalence of acute respiratory infections in women and children in Western Sierra Leone due to smoke from wood and charcoal stoves. Int J Environ Res Public Health 9(6):2252–2265
55. Kilabuko JH, Matsuki H, Nakai S (2007) Air quality and acute respiratory illness in biomass fuel using homes in Bagamoyo, Tanzania. Int J Environ Res Public Health 4(1):39–44
56. Oluwole O, Arinola GO, Huo D, Olopade CO (2017) Household biomass fuel use, asthma symptoms severity, and asthma underdiagnosis in rural schoolchildren in Nigeria: a cross-sectional observational study. BMC Pulm Med 17(1):3
57. Po JY, FitzGerald JM, Carlsten C (2011) Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. Thorax 66(3):232–239
58. Diaz JV, Koff J, Gotway MB, Nishimura S, Balmes JR (2006) Case report: a case of wood-smoke-related pulmonary Disease. Environ Health Perspect 114(5):759–762
59. Gold JA, Jagirdar J, Hay JG, Addrizzo-Harris DJ, Naidich DP, Rom WN (2000) Hut lung. A domestically acquired particulate lung disease. Medicine 79(5):310–317
60. Grobbelaar JP, Bateman ED (1991) Hut lung: a domestically acquired pneumoconiosis of mixed aetiology in rural women. Thorax 46(5):334–340
61. Salvi SS, Barnes PJ (2009) Chronic obstructive pulmonary disease in non-smokers. Lancet 374(9691):733–743
62. Chan KY, Li X, Chen W, Song P, Wong NWK, Poon AN et al (2017) Prevalence of chronic obstructive pulmonary disease (COPD) in China in 1990 and 2010. J Glob Health 7(2):020704
63. Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D et al (2010) An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 182(5):693–718
64. Donaldson K, Brown D, Clouter A, Duffin R, MacNee W, Renwick L et al (2002) The pulmonary toxicology of ultrafine particles. J Aerosol Med 15(2):213–220
65. Holguin F, Flores S, Ross Z, Cortez M, Molina M, Molina L et al (2007) Traffic-related exposures, airway function, inflammation, and respiratory symptoms in children. Am J Respir Crit Care Med 176(12):1236–1242
66. Janssen NA, Brunekreef B, van Vliet P, Aarts F, Meliefste K, Harssema H et al (2003) The relationship between air pollution from heavy traffic and allergic sensitization, bronchial hyperresponsiveness, and

BIOLOGY Methods & Protocols

Biology Methods and Protocols, 2018, 1–9

doi: 10.1093/biomethods/bpy010
Methods Manuscript

METHODS MANUSCRIPT

A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells

Rajendra K.C. ,[1] Graeme R. Zosky,[1] Shakti D. Shukla [2] and Ronan F. O'Toole [1,*]

[1]College of Health and Medicine, University of Tasmania, Hobart, Tasmania 7000, Australia and [2]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, New South Wales 2308, Australia

*Correspondence address: School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, TAS 7000, Australia. Tel: +61-3-62266974; E-mail: ronan.otoole@utas.edu.au

Abstract

Nearly half of the world's population uses biomass fuel for the purposes of cooking and heating. Smoke derived from biomass increases the risk of the development of lung diseases, including pneumonia, chronic obstructive pulmonary disease, airway tract infections, and lung cancer. Despite the evidence linking biomass smoke exposure to pulmonary disease, only a small number of experimental studies have been conducted on the impact of biomass smoke on airway epithelial cells. This is in part due to the lack of a standard and easily accessible procedure for the preparation of biomass smoke. Here, we describe a cost-effective and reproducible method for the generation of different smoke extracts, in particular, cow dung smoke extract (CDSE) and wood smoke extract (WSE) for use in a range of biological applications. We examined the effect of the biomass smoke extracts on human bronchial epithelial cell expression of a known responder to cigarette smoke exposure (CSE), the platelet-activating factor receptor (PAFR). Similar to the treatment with CSE, we observed a dose-dependent increase in PAFR expression on human airway epithelial cells that were exposed to CDSE and WSE. This method provides biomass smoke in a re-usable form for cell and molecular bioscience studies on the pathogenesis of chronic lung disease.

Keywords: biomass smoke; cigarette smoke extract; chronic obstructive pulmonary disease; platelet-activating factor receptor

Introduction

It is estimated that nearly 3 billion people worldwide are exposed to biomass smoke, generated from burning wood, crop residues, or animal dung for household cooking and heating [1]. Biomass smoke is the leading environmental cause of death and disability, causing over 4 million deaths each year [2]. Several epidemiological studies have associated biomass smoke exposure with lung diseases, including chronic obstructive pulmonary disease (COPD), airway infections, and lung cancer [3–7]. Similarly, *in vitro* studies have found that human lung cells exhibit impaired inflammatory and immune responses following exposure to biomass smoke [8, 9]. Inhalation of animal dung biomass smoke is of particular concern to human health as it has the highest polluting potential per unit energy released compared to wood smoke [10]. Airway epithelial cells are the primary target of inhaled smoke; therefore, the responses of epithelial cells to different types of biomass smoke are of considerable interest. Although, extensive *in vitro* studies have been performed on the effects of tobacco smoke on the expression of host receptors on respiratory epithelial cells and on susceptibility to bacterial infection [11–13], only a small number of

Received: 10 May 2018; **Revised:** 2 August 2018. **Accepted:** 10 August 2018

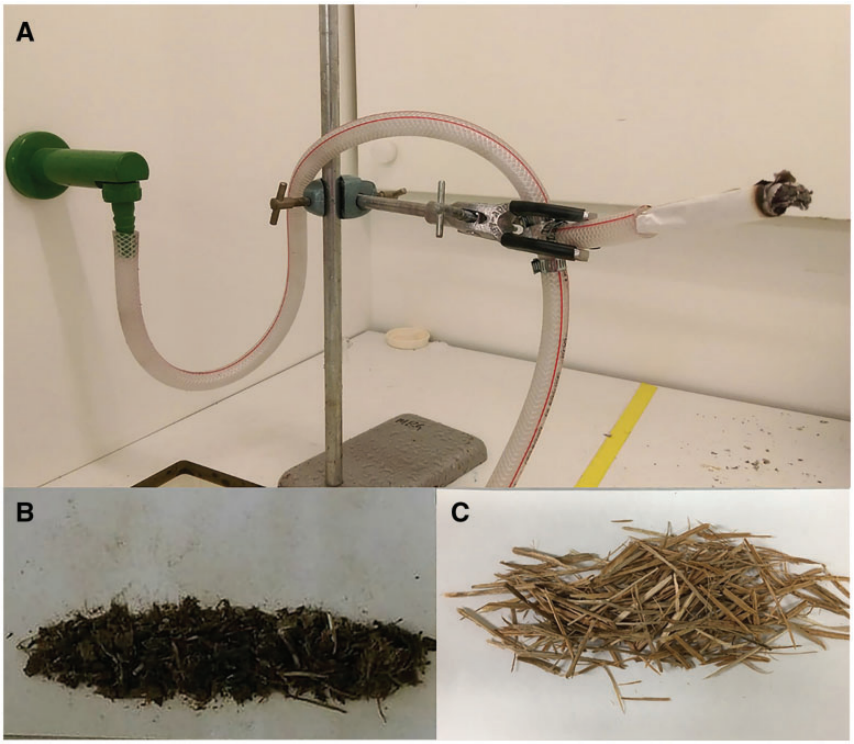


Figure 1: Generation of CDSE and WSE. (**A**) A water aspirator was set up to draw smoke from a burning cow dung or wood shaving roll using the vacuum created by the flow of water. (**B**) Cow dung was sun-dried, crushed into fine particles, and rolled in paper. (**C**) Wood was cut into small chips and rolled in paper.

comparable studies have been performed using biomass smoke. More mechanistic research is therefore needed to understand the cellular and molecular responses to biomass smoke, including animal dung and wood smoke. Currently, we do not have standardized experimental approaches for the preparation of re-usable biomass smoke extract and for the assessment of cellular responses to different types of biomass smoke. Here, we devised a low-cost and reproducible biomass smoke generation system and tested the extracts for their effect on human bronchial epithelial cell expression of platelet-activating factor receptor (PAFR), a G-protein-coupled receptor (GPCR), and an established marker of cigarette smoke exposure [13–15]. GPCRs constitute a large family of membrane-bound receptors that activate intracellular signal transduction pathways in eukaryotic cells in response to extracellular signals [16].

Materials and methods

Preparation of cigarette smoke extract

Cigarette smoke extract (CSE) was prepared at the College of Health and Medicine, University of Tasmania, Australia. Briefly, the filter from a Marlboro cigarette butt was replaced with a sterile cotton wool filter and was smoked using a water aspirator [11]. The water aspirator consisted of a tee with hose barbs on three sides fitted with hoses. The hose-fitted tee was clamped in a stand as shown in Fig. 1A. One of the hoses from the tee fitting was connected to a tap, the second on the opposite side drained water to a sink, and the third hose at a right angle held the cigarette roll. When water was passed through the tube, a vacuum was generated by the Venturi effect, drawing smoke from a burning cigarette [18]. Here, the flow of water was maintained at the constant rate of 110 ml/s. The flow of water ensured the continual smoking of the cigarette, thereby collecting the cigarette smoke material in the cotton wool filter [11, 17]. After complete combustion of a cigarette, the cotton wool filter was removed and then placed into another cigarette from which the filter had been removed. This way, the same cotton filter was used in the smoking of three Marlboro cigarettes.

The cigarette smoke material retained in the cotton filter was quantified by measuring the weight of cotton filter before and after the combustion of the three cigarettes. The cotton filter was then vortexed in 1 ml dimethyl sulfoxide (DMSO). The solubilized smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The CSE was then filter-sterilized through a 0.22 μm membrane filter and the filtrate was re-quantified by weight measurement.

Preparation of cow dung smoke extract

Cow dung, that was collected from a local farm near Hobart, Tasmania, was sun-dried for approximately 5 days and was crushed into fine particles using a mortar and pestle. Cow dung powder was rolled in a paper with a sterile cotton wool filter at one of the ends, similar to a filtered-cigarette (Fig. 1B). Four such rolls were prepared from a total of 7.5 g of cow dung fine particles, such that each roll contained 1.875 g of cow dung powder. The dung roll was then burned and smoked using the water aspirator, as described for cigarette smoking. After complete combustion of a dung roll, the cotton filter was removed and placed in another cow dung roll. In this way, four such cow dung rolls were smoked using the same cotton wool filter. Finally, the cow dung smoke extract (CDSE) was prepared by vortexing the cotton wool filter in 1 ml DMSO. The solubilized

smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The CDSE was then filter-sterilized through a 0.22 μm membrane filter and the filtrate was re-quantified by weight measurement.

Preparation of wood smoke extract

Fire wood was collected from a local supplier near Hobart, Tasmania. For wood smoke generation, 1.875 g of wood shavings were rolled in paper with a sterile cotton wool filter at one of the ends, similar to a filtered-cigarette (Fig. 1C). Four such rolls were prepared from a total of 7.5 g of wood shavings. Each wood shaving roll was then burned and smoked using the water aspirator, as described for cigarette smoking. After complete combustion of a wood shaving roll, the cotton filter was removed and placed into another wood shaving roll. Finally, the wood smoke extract (WSE) was prepared by vortexing the cotton wool filter in 1 ml DMSO. The solubilized smoke material was quantitated by measuring the weight of equal volumes of pure DMSO and smoke material-dissolved DMSO. The WSE was then filter-sterilized through a 0.22 μm membrane filter and the filtrate was re-quantified by weight measurement.

Normalization of smoke extracts

To compare the effects of different smoke extracts, the prepared extracts were normalized to the same concentration and were stored at −20°C in aliquots of 100 μl until use. The normalized smoke extracts were diluted in bronchial epithelial cell growth medium (BEGM) for use in the smoke extract exposure experiments.

In vitro BEAS-2B cell culture

As airway epithelial cells are the primary cells to respond to smoke, an immortalized cell line of human bronchial epithelial cells, BEAS-2B (Catalogue no 95102433, Sigma-Aldrich), was selected for this study. The BEAS-2B cells were maintained at 37°C, 5% CO_2 in BEGM (Lonza, Basel, Switzerland) supplemented with the BulletKit (Lonza). The BEAS-2B cells were sub-cultured in T75 flasks (Corning Inc., Corning, NY, USA), and were used in experiments at passage numbers ≤15 passages. Sterile 8-well chambered glass slides (Millipore, Billerica, MA, USA) were precoated by incubating overnight at 4°C with 200 μl of 5% (v/v) bovine collagen I (ThermoFisher Scientific, USA), prepared in 20 mM acetic acid. The wells were rinsed twice with prewarmed phosphate buffered saline (PBS) followed by seeding of the BEAS-2B cells at a cell density of 30 000 cells per well in 200 μl BEGM and incubated overnight at 37°C, 5% CO_2. On the following day, the culture media was replaced with fresh BEGM and incubated at 37°C, 5% CO_2 for 24 h before the *in vitro* smoke extract exposure experiments.

Exposure of BEAS-2B cells to smoke extracts

Approximately 50 000–60 000 BEAS-2B cells in each well were exposed to 200 μl of BEGM containing five different concentrations of CDSE and WSE, ranging from 8.75 ng/ml to 87.5 μg/ml, for 4 h at 37°C and 5% CO_2. Parallel exposures of BEAS-2B cells to CSE in the concentration range of 8.75 ng/ml to 87.5 μg/ml were also performed for comparison.

Immunofluorescence

After 4 h of exposure to the smoke extracts, the media was discarded and the cells were washed twice with 200 μl PBS pre-warmed at 37°C. The cells were then fixed with 200 μl of 4% (w/v) paraformaldehyde (Sigma-Aldrich) for 20 min at room temperature. The cells were rinsed twice with 200 μl of PBS and permeabilized with 100 μl of chilled (−20°C) acetone for 10 min at room temperature. After washing the cells again twice with 200 μl of PBS, the non-specific binding sites were blocked with 200 μl of 1% (w/v) bovine serum albumin (Sigma-Adrich), prepared in PBS containing 0.1% (v/v) Tween-20 (Sigma-Aldrich), for an hour at room temperature. Cells were then incubated overnight with 100 μl of 2.5 μg/ml monoclonal antibody (mAb) against the human PAFR protein (11A4, Clone 21, Cayman Chemical Company, USA) at 4°C in the dark. The cells were rinsed 3 times with 200 μl of 0.1% (w/v) bovine serum albumin (BSA) in PBS and incubated for an hour with 100 μl of 1:100 dilutions of Alexa Fluor 594 conjugated goat anti-mouse IgG (H + L) secondary antibody (ThermoFisher Scientific, USA) at room temperature. After rinsing 3 times with 200 μl of 0.1% (w/v) BSA in PBS, the cells were stained with 200 μl of 1 μg/mL 4′, 6-diamidino-2-phenylindole dihydrochloride (DAPI) (ThermoFisher Scientific, USA) for 15 min at room temperature. Finally, the cells were washed 3 times with 200 μl of PBS, air-dried and the slides were mounted with Dako fluorescence mounting media (Agilent, USA).

Microscopy and image analysis

Cell preparations were examined under 400× magnification using an Olympus BX50 epifluorescence microscope with NIS elements software (Nikon; Tokyo, Japan) and Cool Snap Hq2 CCD camera (Photometrics, Tucson, AZ, USA). Five images were taken per well from different points using multi-fluorescence channels designed for simultaneous detection of emission from the fluorochromes DAPI (violet excitation and blue emission, 200 ms exposure), and Alexa Fluor 594 (green excitation and red emission, 300-ms exposure). The level of cellular PAFR protein expression was quantified as a measure of total cell fluorescence intensity using the software ImageJ (NIH, USA) [19]. The cellular fluorescence was corrected against the background fluorescence using the following formula:

$$\text{Total cell fluorescence} = \text{integrated density} - (\text{area of selected cell} \times \text{mean fluorescence of the background}).$$

Quantitative real-time polymerase chain reaction analysis

The expression of PAFR was also determined at the transcriptional level using quantitative real-time polymerase chain reaction. The BEAS-2B cells were seeded into sterile clear-flat bottom 12-well plates (Corning Inc.) at a density of 2×10^5 cells per well and incubated overnight at 37°C and 5% CO_2. The next day, cells were exposed to different concentrations of CSE, CDSE, and WSE at 37°C and 5% CO_2. After 3 h, total RNA was extracted with Tri-reagent (Sigma-Aldrich), according to the manufacturer's instructions. It was then treated with DNase (Promega). Using a SensiFAST cDNA synthesis kit (Bioline), 490 ng of RNA was converted into first-stranded cDNA. The cDNA generated was amplified on a LightCycler 480 System (Roche) with the SensiFAST Probe No-ROX kit (Bioline) in a total

volume of 20 µl. The relative fold change of mRNA expression was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

BEAS-2B cell viability assay

The BEAS-2B cells were seeded into a sterile clear-flat bottom 96-well plate (Sigma-Aldrich) at a density of 5000 cells per well and incubated overnight at 37°C and 5% CO_2. The cells were exposed to different concentrations of CSE, CDSE, and WSE. Alamar Blue (Life Technologies) was then added to each well at a final concentration of 10% (v/v). The absorbance readings were taken at 570 and 600 nm at 2 and 4 h post-exposure to smoke extracts at 8.75 and 87.5 µg/ml concentrations using a Spectromax Spectrophotometer Microplate Reader (Molecular Devices, USA). The percent reduction of Alamar Blue was calculated using the following formula:

$$\%\ \text{Reduction of Alamar Blue Reagent} = \frac{(\text{Eoxi600} \times \text{A570}) - (\text{Eoxi570} \times \text{A600})}{(\text{Ered570} \times \text{C600}) - (\text{Ered600} \times \text{C570})} \times 100$$

Molar extinction coefficient of oxidized Alamar Blue at 570 nm (Eoxi570) = 80 586; at 600 nm (Eoxi600) = 117 216

Absorbance of test wells at 570 nm (A570); at 600 nm (A600)

Molar extinction coefficient of reduced Alamar Blue at 570 nm (Ered570) = 155 677; at 600 nm (Ered600) = 14 652

Absorbance of negative control well (no cells) at 570 nm (C570); at 600 nm (C600)

Statistical analysis

Data were expressed as mean ± standard error of the mean (SEM) and median (interquartile range) using the Microsoft Excel Statistics package (Microsoft Corporation, Redmond, WA, USA) and analyzed using GraphPad Prism version 5.0 for Windows (GraphPad Software, San Diego, CA, USA, www.graphpad.com). Comparisons between groups were performed using unpaired two-tailed *t*-tests with Welch's correction and one-way analysis of variance (ANOVA) with Dunnett's multiple comparison analysis.

Results

CSE, CDSE, and WSE preparation

After the combustion of three cigarettes, and 7.5 g each of cow dung powder and wood shavings, 66.52, 124.7, and 131.2 mg of smoke particles were retained in the cotton filter. The retained smoke particles were then solubilized in DMSO. The concentration of DMSO-dissolved cigarette, cow dung, and wood smoke particles were 10.85, 43.7, and 36.0 mg/ml, respectively. After filter sterilization, the final concentration of cigarette, cow dung, and wood smoke material were 8.75, 24.4, and 31.64 mg/ml, respectively. The CSE, CDSE, and WSE concentrations were all normalized to same concentration of 8.75 mg/ml and were used in subsequent exposure experiments over the range from 8.75 ng/ml to 87.5 µg/ml.

CSE exposure increases PAFR expression on bronchial epithelial cells

PAFR expression was measured based on fluorescence intensity following labeling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. Previous studies have reported that PAFR expression is upregulated in BEAS-2B cells exposed to CSE and that maximal induction occurred at 4 h of CSE exposure [11, 12]. Here, CSE exposure for 4 h significantly increased the expression of PAFR on the bronchial epithelial cells (Fig. 2A–C). We observed a dose-dependent increase in PAFR expression upon stimulation with CSE at a concentration range of 8.75 ng/ml to 87.5 µg/ml. In comparison to the mock (1% DMSO) treated control BEAS-2B cells, the mean PAFR expression was approximately 1.18, 1.39, 1.56, 2.06, and 6.10 times higher in the 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 µg/ml, and 87.5 µg/ml CSE treated cells, respectively (Fig. 1C). The level of activity of the CSE was compared over time. There was no significant loss detected in PAFR induction following storage of CSE at −20°C freezer over a 3-month period (Fig. 5). The viability of the BEAS-2B cells exposed to 8.75 and 87.5 µg/ml of CSE for 2 and 4 h was assessed relative to mock (1% DMSO) treated control cells using the Alamar Blue assay. The mean relative viability of BEAS-2B cells treated with 8.75 µg/ml CSE for 2 and 4 h was at 106.2 ± 3.1% (SEM) and 95.8 ± 11.4%, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/ml CSE for 2 and 4 h, the mean relative viability was at 100.7 ± 5.0% and 95.8 ± 11.5%, respectively.

CDSE treatment induces PAFR expression on bronchial epithelial cells

The expression of PAFR on BEAS-2B cells was significantly upregulated by exposure to CDSE (Fig. 3A–C). We observed a dose-dependent increase in PAFR expression upon exposure to CDSE at a concentration range of 8.75 ng/ml to 87.5 µg/ml. The mean cellular PAFR expression was approximately 1.17, 1.24, 1.51, 1.84, and 4.67 times higher than the mock (1% DMSO) treated control BEAS-2B cells in the 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 µg/ml, and 87.5 µg/ml CDSE-stimulated cells, respectively (Fig. 3C). The level of induction of PAFR expression due to CDSE exposure was comparable in experiments conducted 3 months apart (Fig. 5). From the Alamar Blue assay, the mean relative viability of BEAS-2B cells exposed to 8.75 µg/ml CDSE for 2 and 4 h was at 97.4 ± 6.1% and 97.2 ± 13.9%, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/ml CDSE for 2 and 4 h, the mean relative viability was at 102.2 ± 7.9% and 90.8 ± 10.7%, respectively.

WSE exposure upregulates PAFR expression on bronchial epithelial cells

Wood smoke extract exposure was also associated with an induction of PAFR expression on the bronchial epithelial cells (Fig. 4A–C). Treatment with WSE for 4 h resulted in a concentration-dependent increase in the expression of PAFR on BEAS-2B cells. Compared to the 1% DMSO-treated control cells, the mean cellular PAFR expression was approximately 1.28, 1.27, 1.50, 1.99, and 4.34 times higher in 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 µg/ml, and 87.5 µg/ml WSE exposed BEAS-2B cells, respectively (Fig. 4C). The PAFR inducing activity of WSE was similar in experiments conducted 3 months apart (Fig. 5). From the Alamar Blue assay, the mean relative viability of BEAS-2B cells exposed to 8.75 µg/ml WSE for 2 and 4 h was at 101.5 ± 9.2% and 83.8 ± 8.2%, respectively, of the control cell viability. For BEAS-2B cells treated with 87.5 µg/ml WSE for 2 and 4 h, the mean relative viability was at 93.8 ± 7.4% and 79.2 ± 8.1%, respectively.

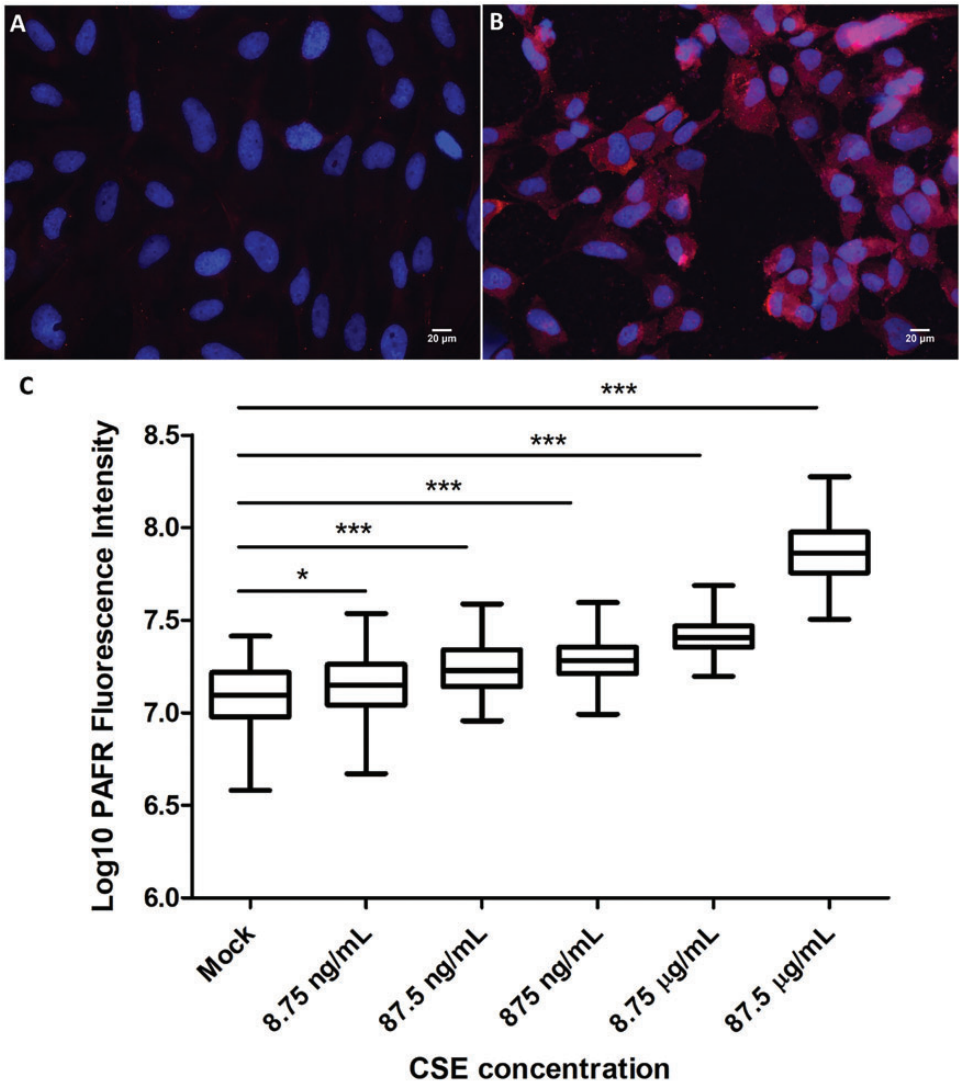


Figure 2: CSE exposure and PAFR expression on bronchial epithelial cells. (**A**) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (**B**) BEAS-2B cells exposed to 87.5 μg/ml CSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 μg/ml, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 μg/ml, blue). Magnification = 400×. (**C**) Response to different concentrations of CSE. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. PAFR expression was significantly increased in 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 μg/ml and 87.5 μg/ml CSE exposed BEAS-2B cells. Data are representative of two independent experiments (*$P < 0.05$, ***$P < 0.0001$, One-way ANOVA with Dunnett's multiple comparison test).

Transcriptional response of BEAS-2B to CSE, CDSE, and WSE

To examine the expression of PAFR at the transcriptional level, the relative PAFR mRNA expression, normalized to glyceraldehyde-3-phosphate (GAPDH), was measured post-exposure to the smoke extracts at 8.75 and 87.5 μg/ml concentrations. Compared to mock (1% DMSO) treated controls, the transcriptional level of PAFR was increased 2.45-, 3.37-, and 2.65-fold after exposure to CSE, CDSE, and WSE, respectively, at 8.75 μg/ml (Table 1). The mRNA levels of PAFR were 3.19-, 4.17-, and 3.38-fold higher in BEAS-2B cells exposed to 87.5 μg/ml concentrations of CSE, CDSE, and WSE, respectively.

Previous studies have reported a respiratory inflammatory response to cigarette, animal dung, and wood smoke exposure [9, 20–23]. To investigate the inflammatory response *in vitro*, the BEAS-2B cells were exposed to 8.75 and 87.5 μg/ml concentrations of CSE, CDSE, and WSE for 3 h and mRNA expression was measured. The mRNA levels for pro-inflammatory cytokines, interleukin-1 beta (IL-1β), IL-6, and IL-8 were increased by 5.56-, 7.15-, and 10.86-fold, respectively, post-exposure to 8.75 μg/ml CSE (Table 1). Exposure of BEAS-2B cells to 87.5 μg/ml CSE resulted in an 8.1-, 13.8-, and 11.4-fold increase in mRNA levels of the inflammatory mediators, IL-1β, IL-6, and IL-8. A similar increase in inflammatory mediators IL-1β, IL-6, and IL-8 was observed upon exposure of BEAS-2B cells to 8.75 and 87.5 μg/ml concentrations of CDSE and WSE (Table 1).

Discussion

Nearly 4.3 million people die every year from illnesses attributable to the inhalation of biomass smoke [1]. Among these deaths, 22% are due to COPD and 12% due to pneumonia [1].

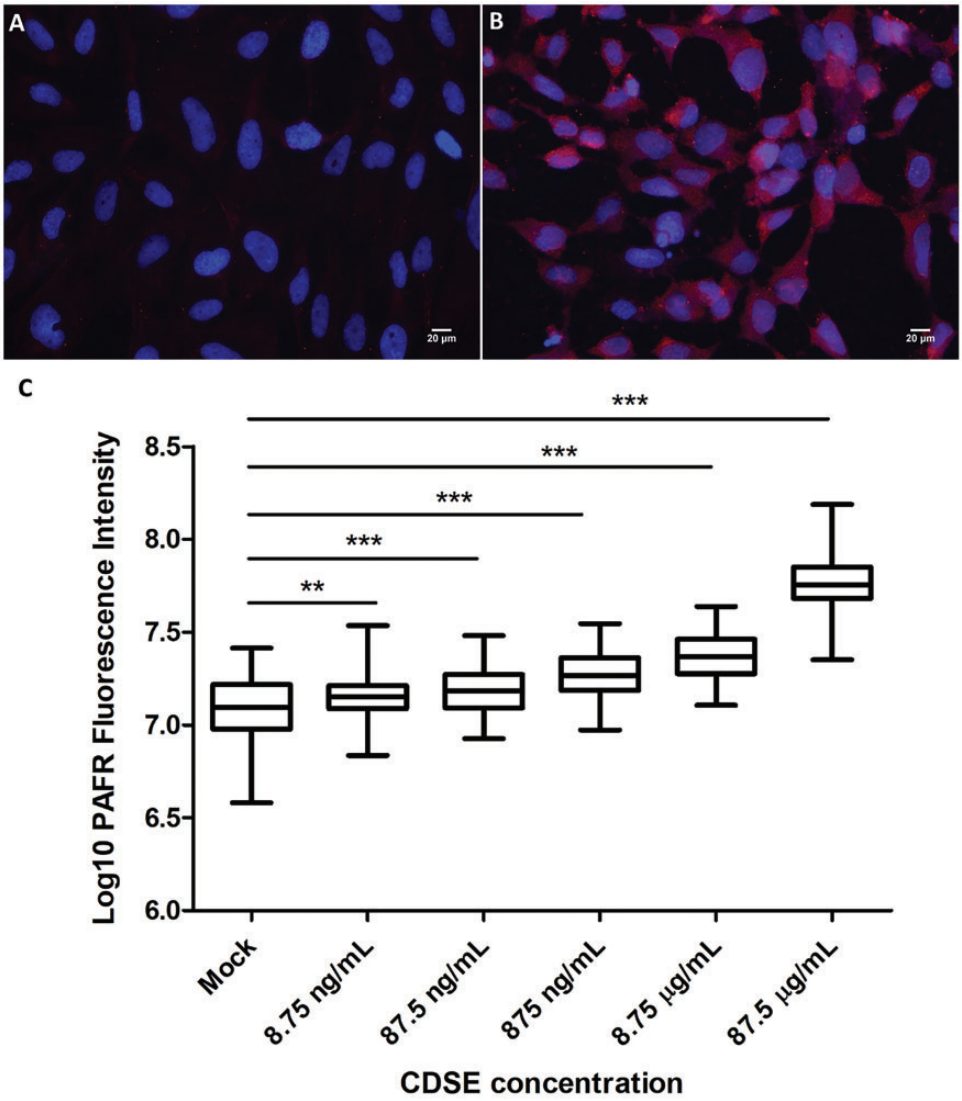


Figure 3: CDSE exposure and PAFR expression on bronchial epithelial cells. (**A**) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (**B**) BEAS-2B cells exposed to 87.5 μg/ml CDSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 μg/ml, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 μg/ml, blue). Magnification = 400×. (**C**) Response to different concentrations of CDSE. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. The PAFR expression was significantly increased in 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 μg/ml and 87.5 μg/ml CDSE treated BEAS-2B cells. Data are representative of two independent experiments (**$P < 0.001$, ***$P < 0.0001$, one-way ANOVA with Dunnett's multiple comparison test).

Biomass fuels (wood, animal dung, and crop residues) are the major source of domestic energy for cooking and household heating, especially in developing countries. Emissions from biomass contain a multitude of pollutants that adversely affect human health, such as suspended particulate matter, methane, free radicals, aldehydes, toxic gases like carbon monoxide and nitrogen oxides, and polycyclic aromatic hydrocarbons like benzo[*a*]pyrene and anthracene [24]. Furthermore, animal dung combustion produces more toxic byproducts, including particulates (23% more $PM_{2.5}$ per kilogram of sample), reactive oxygen species, and microbial products, compared to wood smoke [10, 25]. Several epidemiological studies have correlated biomass smoke exposure with the risk of development of lung diseases, including COPD, lung cancer, and airway infections [3–7]. However, there are only a limited number of studies that have explored mechanisms in biomass smoke induced-pulmonary inflammation and susceptibility to respiratory infections [26]. This knowledge gap is in part due to the lack of a standardized low-cost technique for the generation of biomass smoke in the laboratory.

In work by McCarthy and colleagues, biomass smoke from the combustion of horse dung was pumped into a chamber in which human small airway epithelial cells were exposed [9]. While this method delivered smoke to the epithelial cells, it involved immediate use of the smoke generated and did not allow for storage of batches of biomass smoke for subsequent re-use [9]. It also required the employment of a cigarette smoking machine (Baumgartner-Jaeger CSM2072i) to generate the smoke. In a study by Li and co-workers, biomass smoke from the combustion of rice chaff was bubbled through the cell culture growth medium, Dulbecco's modified Eagle's medium [27]. Again this method generated biomass smoke but the extracts could not be quantified in terms of mass per volume due to the presence of multiple nutrient elements in the growth medium.

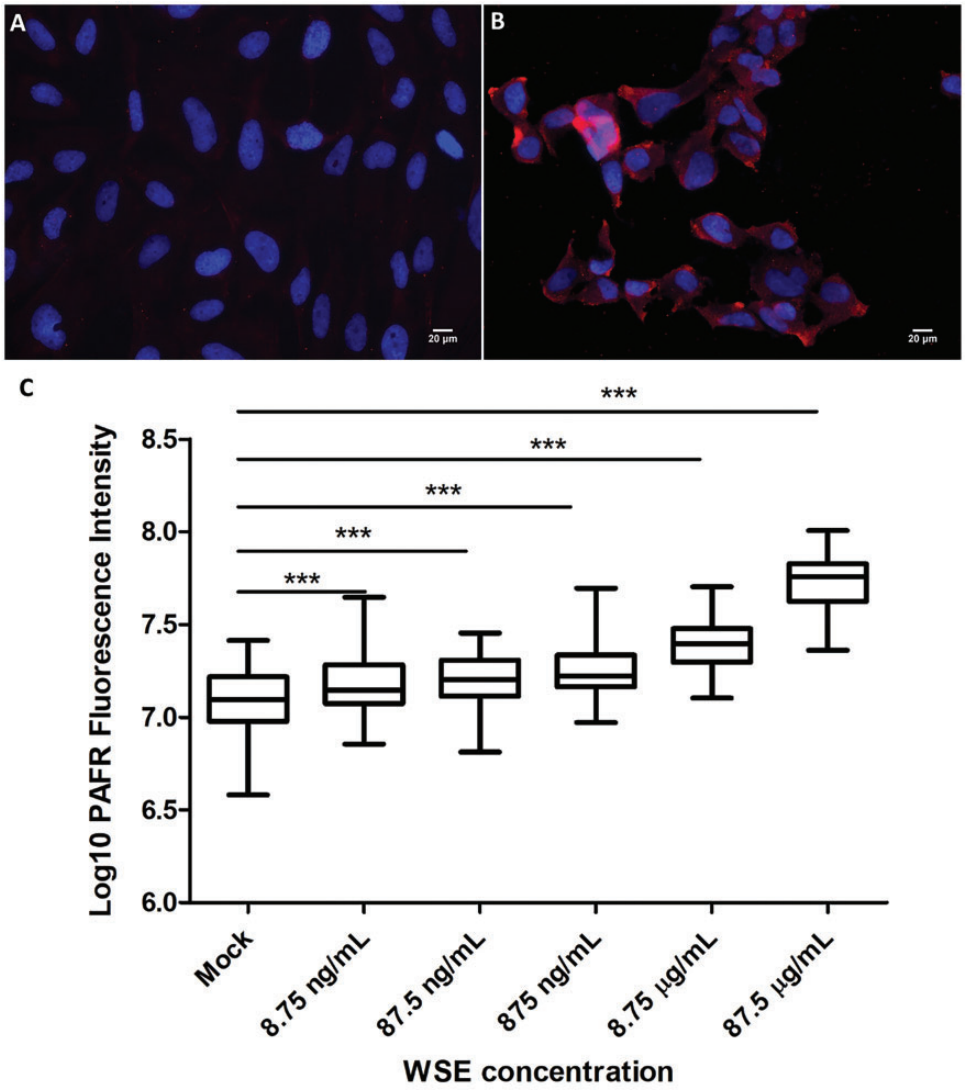


Figure 4: Wood smoke extract (WSE) exposure and PAFR expression on bronchial epithelial cells. (**A**) Mock treatment of BEAS-2B cells with 1% DMSO as a control. (**B**) BEAS-2B cells exposed to 87.5 μg/ml WSE. All immunofluorescence micrographs show BEAS-2B cells with PAFR expression (anti-PAFR monoclonal antibody; 2.5 μg/ml, red) and nuclei stained with 4′, 6-diamidino-2-phenylindole (1 μg/ml, blue). Magnification = 400×. (**C**) Response to different concentrations of WSE. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb. The PAFR expression was significantly increased in 8.75 ng/ml, 87.5 ng/ml, 875 ng/ml, 8.75 μg/ml and 87.5 μg/ml WSE exposed BEAS-2B cells. Data are representative of two independent experiments (***$P < 0.0001$, one-way ANOVA with Dunnett's multiple comparison test).

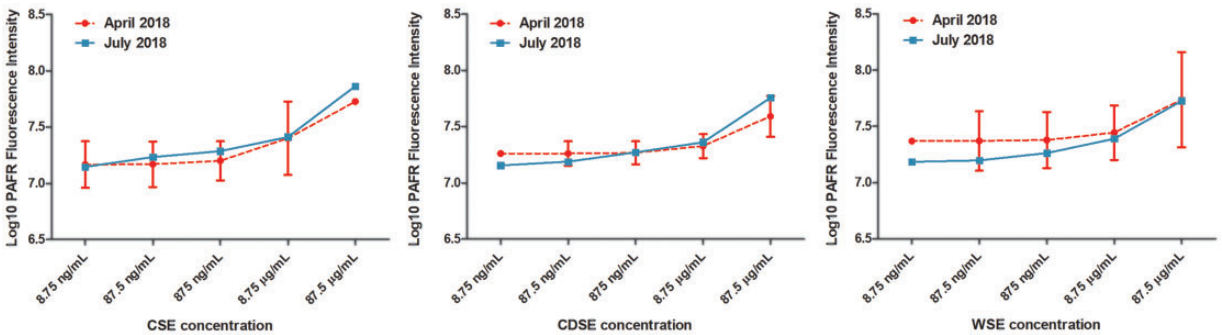


Figure 5: Comparison of effects of smoke extracts on PAFR expression in experiments conducted 3 months apart. All of the smoke extracts, CSE, CDSE, and WSE, were run in the concentration range from 8.75 ng/ml to 87.5 μg/ml in the July 2018 experiment. The CSE extract was run in the same concentration range from 8.75 ng/ml to 87.5 μg/ml in the April 2018 experiment. For the April 2018 experiments, the CDSE and WSE data were interpolated to the 8.75 ng/ml to 87.5 μg/ml concentration range using non-linear least squares regression. The levels of PAFR expression in the July 2018 experiments with the CSE, CDSE, and WSE samples were within 95% CI of the levels obtained for the April 2018 experiments. Therefore, no significant decay in the PAFR inducing activity of the smoke extracts was detected following storage at −20°C over a 3-month period.

Table 1: The smoke extracts CSE, CDSE, and WSE induce PAFR, IL-1β, IL-6, and IL-8 mRNA expression in BEAS-2B cells

Marker	CSE		CDSE		WSE	
	A	B	A	B	A	B
PAFR	2.45 ± 0.08***	3.19 ± 0.69	3.37 ± 0.18***	4.17 ± 1.13	2.65 ± 0.25**	3.38 ± 0.5*
IL-1β	5.56 ± 0.65**	8.06 ± 2.95	8.26 ± 0.27***	14.32 ± 4.73	9.6 ± 1.24**	6.69 ± 0.45**
IL-6	7.15 ± 0.58**	13.83 ± 4.95	12.39 ± 0.4***	12.02 ± 1.57**	10.93 ± 1.05**	13.42 ± 5.74
IL-8	10.86 ± 1.61**	11.44 ± 4.42	11.64 ± 1.0**	20.25 ± 2.43**	10.62 ± 0.6***	8.67 ± 1.81*

Relative fold change in mRNA of PAFR and inflammatory cytokines normalized with GAPDH among BEAS-2B cells exposed to CSE, CDSE, and WSE for 3 h at concentrations (A) 8.75 μg/ml and (B) 87.5 μg/ml (*$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$, unpaired two-tailed *t*-test with Welch's correction. The data are presented as the mean of the observed fold change ± SEM, $n = 4$ per group.

Furthermore, many components of biomass smoke are not directly soluble in aqueous solutions such as growth medium, and as such will not be retained when the smoke is bubbled through the medium [27]. In work by Huang and co-investigators, wood smoke from the burning of Chinese fir was collected directly onto a glass filter with a 1.6-μm pore size [28]. This method collected wood smoke particles as intended but it is likely that many of the smaller components of the smoke, such as volatile organic compounds, from the combustion of the wood would have passed through rather than have been captured on the glass filter.

In our method, the smoke material that was collected in the cotton wool was first incubated in the solvent DMSO, which dissolves both polar and nonpolar compounds, overnight before the filtration step to maximize solubilization of the components. Furthermore, most of the smoke material in the DMSO was retained after the filter sterilization step based on both weight and absorbance measurements. By including quantification measurement at several of the preparation steps, we were able to determine the concentration of smoke-derived material in milligram per milliliter in each of the smoke extracts. In addition, we were able to generate batches of smoke extracts that could be preserved indefinitely and used in multiple exposure experiments, minimizing inter-assay variation. And importantly, our protocol does not require the purchase of expensive equipment and therefore, is suitable for use in resource-limited situations.

To test our biomass smoke extracts, we compared their effect on the expression of PAFR on the human bronchial epithelial cells. PAFR is a G-protein-coupled seven transmembrane domain receptor, involved in various leukocyte functions, platelet aggregation, and inflammation [29]. Previous studies have shown that PAFR expression is upregulated in response to a variety of insults including cigarette smoke, e-cigarette vapor, urban particulate matter, and welding fumes [30–33]. In terms of infection, PAFR is utilized by major respiratory bacterial pathogens including non-typeable *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Pseudomonas aeruginosa* as a surface receptor for adhesion of airway epithelial cells. These species express a common adhesin, known as phosphorylcholine (ChoP), in their cell wall that recognizes and binds host cell PAFR enabling establishment of infection of the respiratory tract [11, 34, 35].

We determined that PAFR expression is increased in bronchial epithelial cells following exposure to CDSE and WSE in a dose-dependent manner at both the protein and mRNA levels (Figs. 2–4, Table 1). In addition, the PAFR-inducing activity of the smoke extracts was comparable in experiments conducted 3 months apart (Fig. 5). Therefore, the activity of the smoke extracts was preserved during storage at −20°C in DMSO for at least the 3-month period tested. The upregulation of PAFR may represent a molecular mechanism through which these biomass smoke types could increase susceptibility to lung diseases including airway infections. Furthermore, we detected increased expression of pro-inflammatory mediators IL-1β, IL-6, and IL-8 following exposure of BEAS-2B cells to our CSE, CDSE, and WSE preparations in accordance with earlier studies on the effect of smoke on respiratory cells (Table 1). Therefore, the ability to produce CDSE and WSE in a usable form, by applying a simple and cost-effective water aspirator-based method, will enable further research on their mechanistic role in the inflammatory response and pathogenesis of respiratory disease including COPD. In addition, it will facilitate the discovery of novel therapeutic compounds that reduce the effects of biomass smoke on host cells and tissues of the respiratory system.

Acknowledgements

The authors acknowledge the support of the College of Health and Medicine, University of Tasmania. They are grateful to Conall O'Toole for collection of the wood and cow dung samples.

Funding

R.K.C. is a recipient of a Health Tasmania Graduate Research Scholarship.

References

1. World Health Organisation. *Household Air Pollution and Health. Fact Sheet Number 292 Updated in February 2016: World Health Organisation*. 2016 http://www.who.int/mediacentre/factsheets/fs292/en/.
2. Lim SS, Vos T, Flaxman AD *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**:2224–60.
3. Assad NA, Balmes J, Mehta S *et al.* Chronic obstructive pulmonary disease secondary to household air pollution. *Semin Respir Crit Care Med* 2015;**36**:408–21.
4. Assad NA, Kapoor V, Sood A. Biomass smoke exposure and chronic lung disease. *Curr Opin Pulm Med* 2016;**22**:150–7.
5. Gordon SB, Bruce NG, Grigg J *et al.* Respiratory risks from household air pollution in low and middle income countries. *Lancet Respir Med* 2014;**2**:823–60.

6. Bruce N, Dherani M, Liu R *et al*. Does household use of biomass fuel cause lung cancer? A systematic review and evaluation of the evidence for the GBD 2010 study. *Thorax* 2015;**70**: 433–41.
7. Capistrano SJ, van Reyk D, Chen H *et al*. Evidence of biomass smoke exposure as a causative factor for the development of COPD. *Toxics* 2017;**5**:36.
8. Mehra D, Mazumder S, Dube B *et al*. Respiratory health effects of dung biomass smoke exposure. *Am J Resp Crit Care Med* 2012;**185** (Meeting Abstracts).
9. McCarthy CE, Duffney PF, Gelein R *et al*. Dung biomass smoke activates inflammatory signaling pathways in human small airway epithelial cells. *Am J Physiol – Lung Cell Mol Physiol* 2016; **311**:L1222–L33.
10. Mudway IS, Duggan ST, Venkataraman C *et al*. Combustion of dried animal dung as biofuel results in the generation of highly redox active fine particulates. *Particle Fibre Toxicol* 2005; **2**:6.
11. Grigg J, Walters H, Sohal SS *et al*. Cigarette smoke and platelet-activating factor receptor dependent adhesion of Streptococcus pneumoniae to lower airway cells. *Thorax* 2012;**67**:908–13.
12. Shukla SD, Fairbairn RL, Gell DA *et al*. An antagonist of the platelet-activating factor receptor inhibits adherence of both nontypeable Haemophilus influenzae and Streptococcus pneumoniae to cultured human bronchial epithelial cells exposed to cigarette smoke. *Int J COPD* 2016;**11**:1647–55.
13. O'Toole RF, Shukla SD, Walters EH. Does upregulated host cell receptor expression provide a link between bacterial adhesion and chronic respiratory disease? *J Transl Med* 2016;**14**: 304.
14. Rajendra KC, Shukla SD, Walters EH *et al*. Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease. *Microbiology* 2017;**163**: 421–30.
15. Shukla SD, Sohal SS, O'Toole RF *et al*. Platelet activating factor receptor: gateway for bacterial chronic airway infection in chronic obstructive pulmonary disease and potential therapeutic target. *Expert Rev Respir Med* 2015;**9**:473–85.
16. Cashin P, Goldsack L, Hall D *et al*. Contrasting signal transduction mechanisms in bacterial and eukaryotic gene transcription. *FEMS Microbiol Lett* 2006;**261**:155–64.
17. Granstrom BW, Xu CB, Nilsson E *et al*. Smoking particles enhance endothelin A and endothelin B receptor-mediated contractions by enhancing translation in rat bronchi. *BMC Pulm Med* 2006;**6**:6.
18. Baylar A, Ozkan F. Applications of venturi principle to water aeration systems. *Environ Fluid Mech* 2006;**6**:341–57.
19. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Method* 2012;**9**:671–5.
20. Mehra D, Geraghty PM, Hardigan AA *et al*. A comparison of the inflammatory and proteolytic effects of dung biomass and cigarette smoke exposure in the lung. *PLoS One* 2012;**7**: e52889.
21. Kocbach A, Herseth JI, Lag M *et al*. Particles from wood smoke and traffic induce differential pro-inflammatory response patterns in co-cultures. *Toxicol Appl Pharmacol* 2008;**232**: 317–26.
22. Swanson KJ, Kado NY, Funk WE *et al*. Release of the pro-inflammatory markers by BEAS-2B cells following in vitro exposure to biodiesel extracts. *Open Toxicol J* 2009;**3**:8–15.
23. Shao J, Wheeler AJ, Chen L *et al*. The pro-inflammatory effects of particulate matter on epithelial cells are associated with elemental composition. *Chemosphere* 2018;**202**:530–7.
24. Larson TV, Koenig JQ. Wood smoke: emissions and noncancer respiratory effects. *Annu Rev Public Health* 1994;**15**: 133–56.
25. Sussan TE, Ingole V, Kim JH *et al*. Source of biomass cooking fuel determines pulmonary response to household air pollution. *Am J Respir Cell Mol Biol* 2014;**50**:538–48.
26. Lee A, Kinney P, Chillrud S *et al*. A systematic review of innate immunomodulatory effects of household air pollution secondary to the burning of biomass fuels. *Annals Global Health* 2015;**81**:368–74.
27. Li C, Zhihong H, Wenlong L *et al*. The nucleotide-binding oligomerization domain-like receptor family pyrindomain-containing 3 inflammasome regulates bronchial epithelial cell injury and proapoptosis after exposure to biomass fuel smoke. *Am J Respir Cell Mol Biol* 2016;**55**:815–24.
28. Huang L, Pu J, He F *et al*. Positive feedback of the amphiregulin-EGFR-ERK pathway mediates PM2.5 from wood smoke-induced MUC5AC expression in epithelial cells. *Sci Rep* 2017;**7**:11084.
29. Ishii S, Nagase T, Shimizu T. Platelet-activating factor receptor. *Prostaglandins Other Lipid Mediat* 2002;**68-69**:599–609.
30. Suri R, Periselneris J, Lanone S *et al*. Exposure to welding fumes and lower airway infection with Streptococcus pneumoniae. *J Allergy Clin Immunol* 2016;**137**:527–34.e7.
31. Grigg J, Miyashita L, Suri R. Pneumococcal infection of respiratory cells exposed to welding fumes: role of oxidative stress and HIF-1 alpha. *PLoS One* 2017;**12**:e0173569.
32. Miyashita L, Suri R, Dearing E *et al*. E-cigarette vapour enhances pneumococcal adherence to airway epithelial cells. *Eur Respir J* 2018;**51**:1701592.
33. Mushtaq N, Ezzati M, Hall L *et al*. Adhesion of Streptococcus pneumoniae to human airway epithelial cells exposed to urban particulate matter. *J Allergy Clin Immunol* 2011;**127**: 1236–42.e2.
34. Swords WE, Buscher BA, Ver Steeg Ii K *et al*. Non-typeable Haemophilus influenzae adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol* 2000;**37**:13–27.
35. Barbier M, Oliver A, Rao J *et al*. Novel phosphorylcholine-containing protein of Pseudomonas aeruginosa chronic infection isolates interacts with airway epithelial cells. *J Infect Dis* 2008;**197**:465–73.

respiratory symptoms in Dutch schoolchildren. Environ Health Perspect 111(12):1512–1518
67. Kulkarni N, Pierse N, Rushton L, Grigg J (2006) Carbon in airway macrophages and lung function in children. N Engl J Med 355(1):21–30
68. Gauderman WJ, Vora H, McConnell R, Berhane K, Gilliland F, Thomas D et al (2007) Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. Lancet 369(9561):571–577
69. Kan H, Heiss G, Rose KM, Whitsel E, Lurmann F, London SJ (2007) Traffic exposure and lung function in adults: the Atherosclerosis Risk in Communities study. Thorax 62(10):873–879
70. Schikowski T, Sugiri D, Ranft U, Gehring U, Heinrich J, Wichmann HE et al (2005) Long-term air pollution exposure and living close to busy roads are associated with COPD in women. Respir Res 6:152
71. Wordley J, Walters S, Ayres JG (1997) Short term variations in hospital admissions and mortality and particulate air pollution. Occup Environ Med 54(2):108–116
72. Zanobetti A, Bind MA, Schwartz J (2008) Particulate air pollution and survival in a COPD cohort. Environ Health 7:48
73. Sunyer J (2001) Urban air pollution and chronic obstructive pulmonary disease: a review. Eur Respir J 17(5):1024–1033
74. Attfield MD (1985) Longitudinal decline in FEV_1 in United States coalminers. Thorax 40(2):132–137
75. Gantt RC, Lincoln JE (1988) Determinants of chronic bronchitis and lung dysfunction in Western Australian gold miners. Br J Ind Med 45(7):503
76. Ulvestad B, Bakke B, Eduard W, Kongerud J, Lund MB (2001) Cumulative exposure to dust causes accelerated decline in lung function in tunnel workers. Occup Environ Med 58(10):663–669
77. Meijer E, Kromhout H, Heederik D (2001) Respiratory effects of exposure to low levels of concrete dust containing crystalline silica. Am J Ind Med 40(2):133–140
78. Niven RM, Fletcher AM, Pickering CA, Fishwick D, Warburton CJ, Simpson JC et al (1997) Chronic bronchitis in textile workers. Thorax 52(1):22–27
79. Bradshaw LM, Fishwick D, Slater T, Pearce N (1998) Chronic bronchitis, work related respiratory symptoms, and pulmonary function in welders in New Zealand. Occup Environ Med 55(3):150–154
80. Schwartz DA, Thorne PS, Yagla SJ, Burmeister LF, Olenchock SA, Watt JL et al (1995) The role of endotoxin in grain dust-induced lung disease. Am J Respir Crit Care Med 152(2):603–608
81. Post W, Heederik D, Houba R (1998) Decline in lung function related to exposure and selection processes among workers in the grain processing and animal feed industry. Occup Environ Med 55(5):349–355
82. Hamzah NA, Mohd Tamrin SB, Ismail NH (2016) Metal dust exposure and lung function deterioration among steel workers: an exposure-response relationship. Int J Occup Environ Health 22(3):224–232
83. Boschetto P, Quintavalle S, Miotto D, Lo Cascio N, Zeni E, Mapp CE (2006) Chronic obstructive pulmonary disease (COPD) and occupational exposures. J Occup Med Toxicol 1:11
84. Hart JE, Laden F, Schenker MB, Garshick E (2006) Chronic obstructive pulmonary disease mortality in diesel-exposed railroad workers. Environ Health Perspect 114(7):1013–1017
85. Weinmann S, Vollmer WM, Breen V, Heumann M, Hnizdo E, Villnave J et al (2008) COPD and occupational exposures: a case-control study. J Occup Environ Med 50(5):561–569
86. Wright JL, Harrison N, Wiggs B, Churg A (1988) Quartz but not iron oxide causes air-flow obstruction, emphysema, and small airways lesions in the rat. Am Rev Respir Dis 138(1):129–135
87. Churg A, Hobson J, Wright J (1989) Functional and morphologic comparison of silica- and elastase-induced airflow obstruction. Exp Lung Res 15(6):813–822
88. McCarthy CE, Duffney PF, Gelein R, Thatcher TH, Elder A, Phipps RP et al (2016) Dung biomass smoke activates inflammatory signaling pathways in human small airway epithelial cells. Am J Physiol Lung Cell Mol Physiol 311(6):L1222–L1233
89. Banerjee A, Mondal NK, Das D, Ray MR (2012) Neutrophilic inflammatory response and oxidative stress in premenopausal women chronically exposed to indoor air pollution from biomass burning. Inflammation. 35(2):671–683
90. Hu G, Zhou Y, Hong W, Tian J, Hu J, Peng G et al (2013) Development and systematic oxidative stress of a rat model of chronic bronchitis and emphysema induced by biomass smoke. Exp Lung Res 39(6):229–240
91. Kurmi OP, Dunster C, Ayres JG, Kelly FJ (2013) Oxidative potential of smoke from burning wood and mixed biomass fuels. Free Radic Res 47(10):829–835
92. Warheit DB, Hansen JF, Yuen IS, Kelly DP, Snajdr SI, Hartsky MA (1997) Inhalation of high concentrations of low toxicity dusts in rats results in impaired pulmonary clearance mechanisms and persistent inflammation. Toxicol Appl Pharmacol 145(1):10–22
93. Krimmer D, Ichimaru Y, Burgess J, Black J, Oliver B (2013) Exposure to biomass smoke extract enhances fibronectin release from fibroblasts. PLoS ONE 8(12):e83938
94. Moorthy B, Chu C, Carlin DJ (2015) Polycyclic aromatic hydrocarbons: from metabolism to lung cancer. Toxicol Sci 145(1):5–15
95. Mastrangelo G, Fadda E, Marzia V (1996) Polycyclic aromatic hydrocarbons and cancer in man. Environ Health Perspect 104(11):1166–1170
96. Siddens LK, Larkin A, Krueger SK, Bradfield CA, Waters KM, Tilton SC et al (2012) Polycyclic aromatic hydrocarbons as skin carcinogens: comparison of benzo[a]pyrene, dibenzo[def, p]chrysene and three environmental mixtures in the FVB/N mouse. Toxicol Appl Pharmacol 264(3):377–386
97. Rengarajan T, Rajendran P, Nandakumar N, Lokeshkumar B, Rajendran P, Nishigaki I (2015) Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. Asian Pac J Trop Biomed 5(3):182–189
98. Lim WY, Seow A (2012) Biomass fuels and lung cancer. Respirology (Carlton, Vic). 17(1):20–31
99. Denissenko MF, Pao A, Tang M, Pfeifer GP (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science 274(5286):430–432
100. Ross JA, Nesnow S (1999) Polycyclic aromatic hydrocarbons: correlations between DNA adducts and ras oncogene mutations. Mutat Res 424(1–2):155–166
101. Peluso M, Munnia A, Hoek G, Krzyzanowski M, Veglia F, Airoldi L et al (2005) DNA adducts and lung cancer risk: a prospective study. Can Res 65(17):8042–8048
102. Rylance J, Fullerton DG, Scriven J, Aljurayyan AN, Mzinza D, Barrett S et al (2015) Household air pollution causes dose-dependent inflammation and altered phagocytosis in human macrophages. Am J Respir Cell Mol Biol 52(5):584–593
103. Kc R, Zosky GR, Shukla SD, O'Toole RF (2018) A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells. Biol Methods Protocol 3:bpy010
104. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Lam KB et al (2014) Respiratory risks from household air pollution in low and middle income countries. Lancet Respir Med 2(10):823–860
105. Roebuck KA, Rahman A, Lakshminarayanan V, Janakidevi K, Malik AB (1995) H_2O_2 and tumor necrosis factor-alpha activate intercellular adhesion molecule 1 (ICAM-1) gene transcription through distinct cis-regulatory elements within the ICAM-1 promoter. J Biol Chem 270(32):18966–18974
106. Ishizuka S, Yamaya M, Suzuki T, Nakayama K, Kamanaka M, Ida S et al (2001) Acid exposure stimulates the adherence of Streptococcus pneumoniae to cultured human airway epithelial cells: effects on platelet-activating factor receptor expression. Am J Respir Cell Mol Biol 24(4):459–468
107. Kc R, Shukla SD, Walters EH, O'Toole RF (2017) Temporal upregulation of host surface receptors provides a window of opportunity for bacterial adhesion and disease. Microbiology 163(4):421–430
108. Shukla SD, Mahmood MQ, Weston S, Latham R, Muller HK, Sohal SS et al (2017) The main rhinovirus respiratory tract adhesion site (ICAM-1) is upregulated in smokers and patients with chronic airflow limitation (CAL). Respir Res 18(1):6
109. Lopez-Campos JL, Calero C, Arellano-Orden E, Marquez-Martin E, Cejudo-Ramos P, Ortega Ruiz F et al (2012) Increased levels of soluble ICAM-1 in chronic obstructive pulmonary disease and resistant smokers are related to active smoking. Biomark Med 6(6):805–811
110. Shukla SD, Muller HK, Latham R, Sohal SS, Walters EH (2016) Platelet-activating factor receptor (PAFr) is upregulated in small airways and alveoli of smokers and COPD patients. Respirology 21(3):504–510

Exposure and Health
https://doi.org/10.1007/s12403-020-00351-y

ORIGINAL PAPER

Cow Dung Biomass Smoke Exposure Increases Adherence of Respiratory Pathogen Nontypeable *Haemophilus influenzae* to Human Bronchial Epithelial Cells

Rajendra KC[1] · Isabel K. Hyland[2] · Jason A. Smith[2] · Shakti D. Shukla[3] · Philip M. Hansbro[4,5] · Graeme R. Zosky[1] · Gunasegaran Karupiah[1] · Ronan F. O'Toole[1,6,7]

Received: 1 August 2019 / Revised: 5 December 2019 / Accepted: 22 February 2020

Abstract

Biomass smoke exposure is associated with a heightened risk of development of respiratory diseases that include chronic obstructive pulmonary disease (COPD). The aim of this study was to increase our understanding of how biomass smoke could contribute to an increased susceptibility to respiratory infection. We investigated the effects of cow dung and wood smoke exposure on human bronchial epithelial cells with respect to adherence of a major respiratory bacterial pathogen in COPD, nontypeable *Haemophilus influenzae* (NTHi), using immunofluorescence microscopy. In addition, expression of a known receptor of NTHi, platelet-activating factor receptor (PAFR), and two pro-inflammatory cytokines, interleukin 6 (IL-6) and interleukin-8 (IL-8), were determined using quantitative polymerase chain reaction. We observed a dose-dependent increase in NTHi adhesion to human bronchial epithelial cells following exposure to cow dung but not wood smoke extracts. Pre-treatment with PAFR antagonists, WEB-2086 and its analogue, C17, decreased adherence by NTHi to airway epithelial cells exposed to cow dung smoke. Both cow dung and wood smoke-induced expression of PAFR, as well as of IL-6 and IL-8, which was inhibited by WEB-2086 and C17. In conclusion, biomass smoke from combustion of cow dung and wood-induced expression of PAFR and airway inflammatory markers in human bronchial epithelial cells. Cow dung exposure, but not wood smoke exposure, mediated a measurable increase in NTHi adhesion to airway epithelial cells that was inhibited by PAFR antagonists. This work highlights the potential of PAFR as a therapeutic target for reducing the impact of hazardous biomass smoke exposure on respiratory health.

Keywords Chronic obstructive pulmonary disease · Cow dung smoke · Biomass smoke · Cigarette smoke · Nontypeable *Haemophilus influenzae*

✉ Ronan F. O'Toole
r.otoole@latrobe.edu.au

[1] Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia

[2] School of Natural Sciences, College of Sciences and Engineering, University of Tasmania, Hobart, Tasmania, Australia

[3] Priority Research Centre for Healthy Lungs and School of Medicine and Public Health, Faculty of Health and Medicine, and Hunter Medical Research Institute, University of Newcastle, Callaghan, New South Wales, Australia

[4] Centre for Inflammation, Faculty of Science, University of Technology Sydney, Ultimo, New South Wales, Australia

[5] Centenary Institute, Sydney, New South Wales, Australia

[6] Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin, Ireland

[7] Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria, Australia

Published online: 09 March 2020

Introduction

Almost half of the world's population is reliant on biomass, such as wood, animal dung, and crop residues as their primary source of fuel for cooking and heating purposes (WHO 2018a, b). Rural human populations of low- and middle-income countries, in particular women and young children, are regularly exposed to extremely high levels of smoke from wood or cow dung that is used for cooking. For instance, in a rural Costa Rican setting, Park and Lee (Park and Lee 2003) monitored the quality of indoor air of households with wood burning stoves for 24-h and reported elevated concentrations of $PM_{2.5}$ (44 µg/m^3), PM_{10} (132 µg/m^3), and carbon monoxide (0.5 and 3.3 ppm). The concentration of these parameters peaked considerably during cooking, with $PM_{2.5}$ and PM_{10} levels reaching a concentration of approximately 310–8170 µg/m^3 and 500–18,900 µg/m^3, respectively (Park and Lee 2003).

Household biomass smoke exposure has been linked to an increased risk of airway disease including chronic obstructive pulmonary disease (COPD) (Capistrano et al. 2017; KC et al. 2018a, b). An Indian study involving non-smoking women reported an increased prevalence of respiratory symptoms (chronic bronchitis and dyspnea) in those who used biomass fuel for cooking (Behera and Jindal 1991). Another recent Indian study reported that a substantial proportion (> 60%) of biomass fuel (cow dung cakes and firewood) users are at higher risk of developing COPD, bronchial asthma and tuberculosis (Faizan and Thakur 2019). Worldwide, COPD is responsible for 3 Mio. deaths each year making it the third largest cause of human mortalities after heart disease and stroke (WHO 2018a, b). While tobacco smoking is still the primary causative factor of COPD in high-income countries, in low and middle countries over 45% of COPD patients have been found to be never smokers (KC et al. 2018a, b). There is emerging evidence that smoke from biomass fuels is an independent risk factor for the development of COPD where reliance on biomass fuel is high (Ehrlich et al. 2004; Magitta et al. 2018). However, the exact mechanism regarding how biomass smoke exposure contributes to the development and progression of COPD is still unknown. Since, biomass fuel smoke has been linked to acute respiratory infections (Po et al. 2011), we hypothesized that the exposure of bronchial epithelial cells to biomass smoke extracts increases adherence of nontypeable *Haemophilus influenzae* (NTHi), the major respiratory pathogen involved in the pathogenesis of COPD.

This study was conducted to advance our understanding of how biomass smoke could contribute to an increased susceptibility to respiratory infection, In our work, we investigated the response of human bronchial epithelial cells to cow dung and wood smoke extract exposures, in particular, the expression of platelet-activating factor receptor (PAFR) and of pro-inflammatory cytokines, interleukin 6 (IL-6) and IL-8, and the level of adhesion of NTHi to smoke extract-treated cells. In particular, we demonstrate here that exposure to cow dung smoke extract results in increased adhesion of NTHi to human bronchial epithelial cells.

Methods

Generation of Cigarette, Cow Dung and Wood Smoke Extracts

In our previous paper, we described in detail a cost-effective and reproducible method for the generation of different smoke extracts, including cigarette, cow dung and wood smoke extracts (CSE, CDSE and WSE, respectively) (KC et al. 2018a, b). Briefly, three cigarette rolls, four rolls of cow dung powder, and four rolls of wood shavings were smoked using a water aspirator. The cigarette, cow dung and wood smoke extracts were prepared by vortexing cotton filter containing the trapped smoke particles in 1 mL of dimethyl sulfoxide (DMSO). The smoke extracts were then filter-sterilized through a 0.22 µm membrane filter and quantified by weight measurement. Finally, the prepared extracts were normalized to the same concentration (8.75 mg/mL) and stored at − 20 °C. The extracts were diluted in bronchial epithelial cell growth medium (BEGM) for use in exposure experiments.

In Vitro BEAS-2B Cell Culture and Fluorescence Labelling of Bacteria

Airway epithelial cells are the primary cells to respond to environmental insults to the respiratory system, such as smoke and dust particles (Crystal et al. 2008). These cells are of utility for investigating the biological and/or pathological effects of environmental air pollutants and pathogens (Gruenert et al. 1995). The in vitro system comprising airway epithelial cells, including bronchial and alveolar epithelial cells are routinely utilised in both basic and clinical research settings, with the advantage of being able to expose these cells directly to environmental stimuli. Bronchial epithelial cells are one of the primary sites in the respiratory tract to interact with and respond to pathogenic microbes and environmental substances. Primary human bronchial epithelial cells are mortal, show a high degree of variability and are difficult to isolate, whereas, commercially available bronchial epithelial cells, such as BEAS-2B cells, are immortal, much more

homogenous under normal conditions, and are widely accessible. Thus, we utilized BEAS-2B cells to evaluate the effect of smoke extract exposures. BEAS-2B cells (Catalogue no 95102433, Sigma-Aldrich) were cultured in serum free bronchial epithelial basal medium (BEBM) supplemented with the BulletKit (BEGM) (Lonza, Basel, Switzerland) at 37 °C and 5% CO_2. The BEAS-2B cells were sub-cultured in a T-75 flask (Corning, NY, USA), and were used for experiments at ≤ 15 passages. The NTHi strain RHH3 was collected as part of routine hospital laboratory diagnosis at the Royal Hobart Hospital, Tasmania, Australia and no research participants or patients were recruited for this purpose. The study was approved by the Tasmanian Health and Medical Human Research Ethics Committee (EC00337) and conducted in accordance with the Declaration of Helsinki. The bacteria were labelled with fluorescein isothiocyanate (FITC) (Sigma-Aldrich, USA) as previously described (Geelen et al. 1993). Briefly, bacteria were grown overnight in brain heart infusion medium supplemented with hemin (1 mg/mL) and nicotinamide adenine dinucleotide (10 mg/mL). Subsequently, bacteria (10^8 CFU/mL) were incubated for 1 h at 4 °C with 1 mg/mL (w/v) FITC prepared in a carbonate buffer (0.09 M sodium carbonate, 0.015 M sodium bicarbonate and 0.1 M sodium chloride). The FITC-labelled bacteria were washed three times with phosphate-buffered saline (PBS) and resuspended in BEGM without antibiotics at a density of 5×10^7 CFU/mL.

Smoke Extract Exposure

For adherence assays, the BEAS-2B cells were transferred to sterile 8-well-chambered glass slides (Millipore, MA, USA) pre-coated with 5% (v/v) bovine collagen I (Thermo Fisher Scientific, USA) at a density of 30,000 cells per well and cultured for an additional 24 h to form a confluent monolayer. The cell monolayer was then exposed to CSE, CDSE and WSE at concentrations ranging from 87.5 ng/mL to 87.5 µg/mL for 4 h at 37 °C and 5% CO_2 as described in our previous paper (KC et al. 2018b).

Pre-treatment with PAFR Antagonists

After exposure to smoke extracts, the BEAS-2B cells were washed with pre-warmed PBS and treated with PAFR antagonists, WEB-2086 (Sigma-Aldrich, USA) and a new analogue C17 (synthesized at the University of Tasmania whereby the fused triazolo-diazepine ring and oxopropylmorpholine side chain group were replaced by a pyrrole ring and a dodecyl group, respectively) (Fig. 1). A 10 mM stock solution of the compounds was prepared in DMSO (Sigma, USA) and diluted in antibiotic-free BEGM to final concentrations of 100 µM, 10 µM, 1 µM and 100 nM. The BEAS-2B cells were pre-incubated with 200 µL of 100 µM, 10 µM, 1 µM and 100 nM concentrations of WEB-2086 or C17 for 1 h at 37 °C and 5% CO_2.

Bacterial Adhesion Assays

Following pre-treatment of the BEAS-2B cell monolayer with PAFR antagonists, 100 µL of FITC-tagged NTHi was added to the monolayer at a multiplicity of infection of 100:1. We conducted the smoke extract exposure and bacterial adhesion experiments at 37 °C mimicking the in vivo body temperature and the optimal temperature for obligate human pathogen NTHi. The cells were incubated for 1 h at 37 °C and 5% CO_2 and subsequently washed three times with pre-warmed PBS. BEAS-2B cells were fixed with 200 µL of 4% (w/v) paraformaldehyde (Sigma-Aldrich, USA) for 20 min at room temperature. After 2 washes with PBS, cells were permeabilized with 100 µL of chilled (− 20 °C) acetone for 10 min at room temperature. Subsequently, the cells were washed three times with PBS and blocked with 200 µL of 1% (w/v) bovine serum albumin (BSA) (Sigma-Aldrich, USA), prepared in PBS containing 1% (v/v) Tween-20 (Sigma-Aldrich, USA) for 1 h at room temperature. The cells were then incubated with mouse anti-human platelet-activating factor receptor monoclonal antibody (anti-PAFR

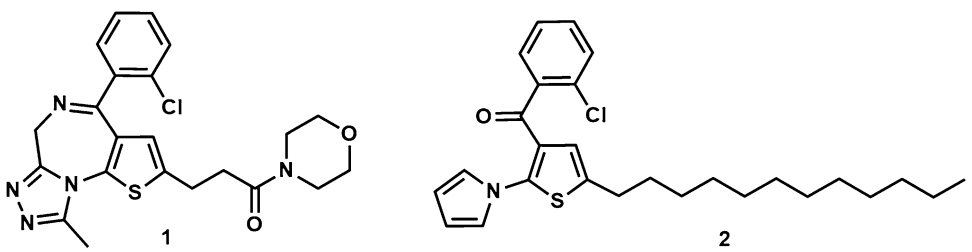


Fig. 1 Chemical structure of WEB-2086 (1) and its analogue C17 (2). WEB-2086 is a PAFR antagonist and belongs to the group of thienotriazolo-1,4-diazepines (hetrazepines). C17 is an analogue of WEB-2086 in which the fused triazolo-diazepine ring and the oxopropylmorpholine side chain group are replaced by a pyrrole ring and a dodecyl group, respectively

mAb, 2.5 µg/mL) (11A4, Clone 21, Cayman Chemical Company, USA) overnight in the dark at 4 °C. After 3 washes with 0.1% (w/v) BSA, cells were incubated with a 1:100 dilution of Alexa Fluor 594-conjugated goat anti-mouse IgG (H+L) secondary antibody (Thermo Fisher Scientific, USA) for 1 h at room temperature. The cells were then rinsed three times with 0.1% (w/v) BSA, followed by staining with 4′,6-diamidino-2-phenylindole (DAPI, 1 µg/mL) (Thermo Fisher Scientific, USA) in the dark at room temperature for 15 min. Finally, the cells were washed three times with PBS and air-dried before the slides were mounted with Dako fluorescence mounting media (Agilent, USA).

Microscopy and Image Analysis

The cells were examined with an Olympus BX50 epifluorescence microscope equipped with NIS elements software (Nikon, Japan) and a Cool Snap Hq2 CCD camera (Photometrics, USA) under ×400 magnification. Ten images were taken per well from different points using three fluorescence channels for DAPI (violet excitation and blue emission, 200 ms exposure), Alexa Fluor 594 (green excitation and red emission, 300 ms exposure) and FITC (blue excitation and green emission, 500 ms exposure). Image merging and bacterial and BEAS-2B cell counting were performed using ImageJ (NIH, USA).

Quantitative Real-time Polymerase Chain Reaction (qRT-PCR) Analysis

To determine the transcriptional level of expression of PAFR, IL-6 and IL-8, the BEAS-2B cells were seeded into sterile clear flat-bottom 12-well plates (Corning, USA), not coated with collagen, at a density of 2×10^5 cells per well and incubated overnight at 37 °C and 5% CO_2. The cells were then exposed to smoke extracts CSE, CDSE or WSE for 4 h at 37 °C and 5% CO_2. After 4 h, total RNA was extracted using Tri-reagent (Sigma-Aldrich, USA), according to the manufacturer's instructions. Any traces of genomic DNA were degraded by treating the total RNA extract with DNase (Promega, USA). This was followed by first-stranded cDNA synthesis using a SensiFAST cDNA synthesis kit (Bioline, UK). Subsequently, the cDNA generated was amplified on a LightCycler 480 System (Roche, Switzerland) using SensiFAST Probe No-ROX kit (Bioline, UK). Finally, mRNA expression was normalized to the expression of a housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and the fold-changes in PAFR, IL-6 and IL-8 mRNA levels in smoke extract-exposed cells were expressed relative to mRNA levels in mock (1% DMSO) treated cells using the primer pairs: GAPDH_F, GCCAAGGTCATCCATGACAACT; GAPDH_R, GGGCCATCCA-CAGTCTTCTG; PAFR_F, GACAGCATAGAGGCTGAGGC; PAFR_R, TAGCCATT-AGCAATGACCCC; IL-6_F, ACAGCCACTCACCTCTTCAG; IL-6_R, CCATCTTT- TTCAGCCATCTTT; IL-8_F, CTTGGCAGCCTTCCTGATTT; IL-8_R, TTCTTTAG- CACTCCTTGGCAAAA, respectively.

Cytotoxic Effect of WEB-2086 and C17 on BEAS-2B Cells

To ascertain whether the compounds WEB-2086 and C17 exhibited toxicity towards bronchial epithelial cells, BEAS-2B cells were seeded into sterile clear flat-bottom 96-well plates (Sigma-Aldrich, USA), not coated with collagen, at a density of 5000 cells per well and incubated overnight at 37 °C and 5% CO_2. The cells were then treated in triplicate with different concentrations of WEB-2086 or C17 ranging from 100 nM to 100 µM, followed by the addition of Alamar Blue (Life Technologies, USA) to each well at a final concentration of 10% (v/v). The cells were then incubated for 4 h at 37 °C and 5% CO_2 to allow reduction of resazurin (blue) to resorufin (red). The absorbance signal was then measured at 570 nm with 600 nm as a reference wavelength after 4 and 24 h of treatment with WEB-2086 or C17, using a Spectromax Spectrophotometer Microplate Reader (Molecular Devices, USA). The percent reduction of Alamar Blue was calculated as previously described (KC et al. 2018b).

Statistical Analysis

Two independent experiments, each including two or three replicates, were analyzed. Data from four or six biological replicates from two experiments were included in the analysis and were expressed as the mean ± standard error. Differences between groups were tested by one-way or two-way analysis of variance (ANOVA) with Dunnett's or Bonferroni post-hoc analysis using GraphPad Prism version 5.0 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

Results

CDSE but not WSE Significantly Upregulated NTHi Adherence to Human Bronchial Epithelial Cells

We investigated whether exposure to cow dung and wood smoke extracts increases the susceptibility of bronchial epithelial cells to NTHi adhesion. For this BEAS-2B cells were exposed to CDSE and WSE at concentrations ranging from 87.5 ng/mL to 87.5 µg/mL for 4 h. Parallel exposures of BEAS-2B cells to CSE were performed for comparison. Treatment of BEAS-2B cells with CSE, CDSE and WSE correlated with increased PAFR expression as measured

by immunofluorescence as previously described (KC et al. 2018b). Challenge of BEAS-2B with FITC-labelled NTHi for an hour revealed that bacterial adhesion to unstimulated BEAS-2B cells was low *i.e.* 1.859 ± 0.023 bacteria per cell. The number of bacteria adhering to BEAS-2B cells exposed to CSE and CDSE increased in a dose-dependent manner.

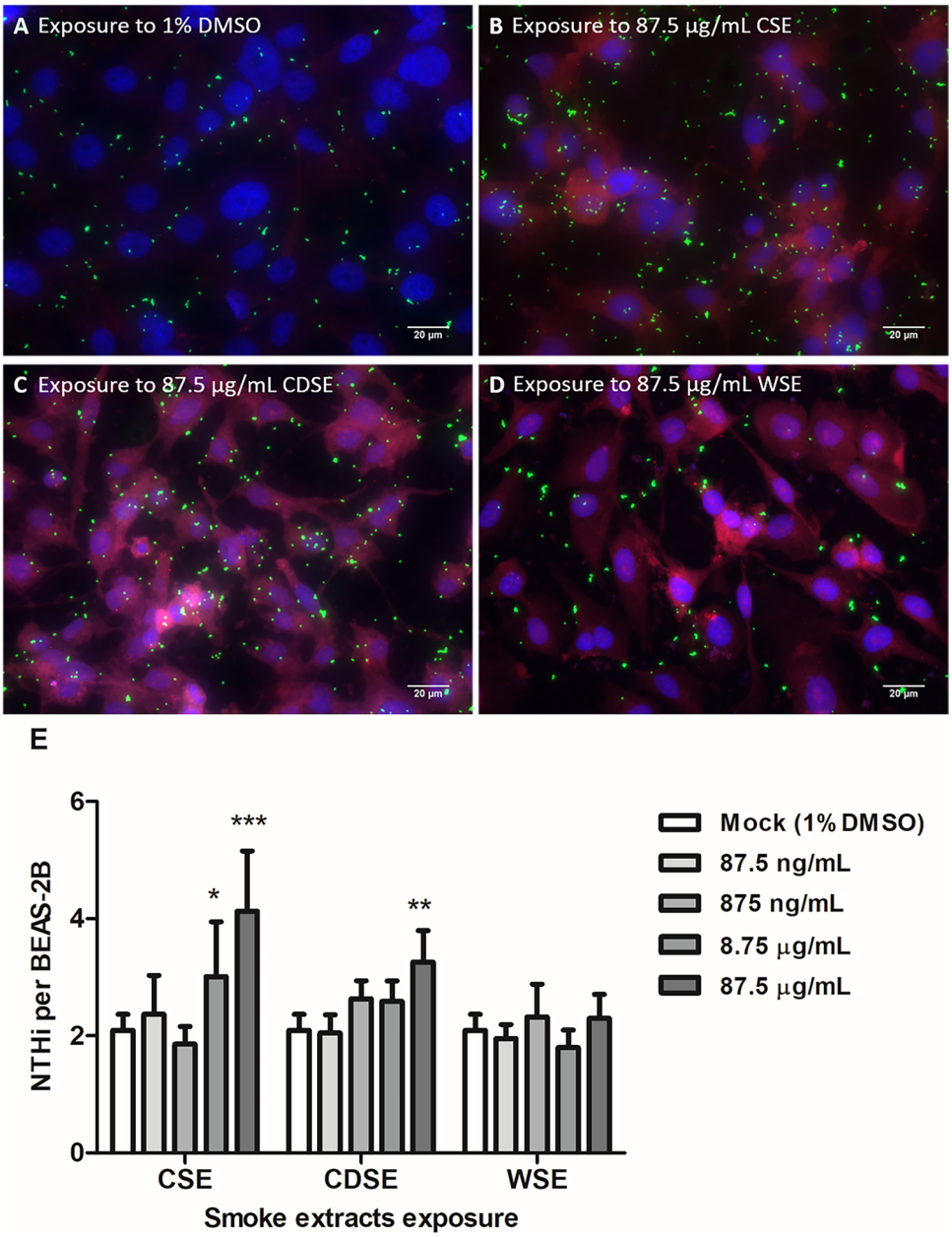


Fig. 2 Effect of biomass smoke exposure on the adhesion of nontypeable *H. influenzae* strain RHH3 to human bronchial epithelial cells. **a** Mock treatment of BEAS-2B cells with 1% DMSO as a control. **b** BEAS-2B cells exposed to 87.5 µg/mL CSE for comparison. **c** BEAS-2B cells exposed to 87.5 µg/mL CDSE. **d** BEAS-2B cells exposed to 87.5 µg/mL WSE. All immunofluorescence micrographs show NTHi labelled with FITC (1 mg/mL, green) and BEAS-2B cells stained for PAFR expression (anti-PAFR monoclonal antibody; 2.5 µg/mL, red), and nuclei counterstained with DAPI (1 µg/mL, blue). Magnification = × 400. **e** Exposure of BEAS-2B cells to different concentrations of CSE, CDSE and WSE ranging from 87.5 ng/mL to 87.5 µg/mL. NTHi adherence was significantly increased in CSE (8.75 and 87.5 µg/mL) and CDSE (87.5 µg/mL) exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, relative to mock-treated cells, one-way ANOVA with Dunnett's multiple comparison test. *CSE* cigarette smoke extract, *CDSE* cow dung smoke extract, *WSE* wood smoke extract, *NTHi* nontypeable *Haemophilus influenzae*, *PAFR* platelet-activating factor receptor, *DMSO* dimethyl sulfoxide, *FITC* fluorescein isothiocyanate, *DAPI* 4′, 6-diamidino-2-phenylindole

NTHi adherence to CSE- and CDSE-exposed BEAS-2B cells was 2.0 ($p < 0.001$), and 1.6 ($p < 0.01$) fold higher, respectively, with respect to 1% DMSO (mock) treated BEAS-2B cells, at the 87.5 µg/mL concentration (Fig. 2). At lower concentrations of CSE and CDSE *i.e.* 875 ng/mL and 87.5 ng/mL, NTHi adhesion did not differ significantly between mock-treated and smoke extract-exposed BEAS-2B cells. Regression analyses were performed on levels of PAFR expression based on PAFR fluorescence intensity with respect to NTHi adhesion. The adherence of NTHi to CSE and CDSE-exposed BEAS-2B cells was positively correlated with levels of PAFR expression ($r = 0.96$ $p = 0.04$ and $r = 0.979$ $p = 0.021$, respectively) (Fig. 3). WSE exposure did not result in increased attachment of NTHi to BEAS-2B cells across the concentration range tested *i.e.* 87.5 ng/mL to 87.5 µg/mL. Compared to 87.5 µg/mL WSE exposure, NTHi adhesion to bronchial epithelial cells was 80% ($p < 0.001$) and 42% ($p < 0.05$) higher in CSE and CDSE exposures, respectively.

WEB-2086 and C17 Significantly Reduced the Adhesion of NTHi to Cow Dung Smoke Exposed BEAS-2B Cells

To determine whether PAFR played a role in adhesion by NTHi to cow dung smoke exposed bronchial epithelial cells, BEAS-2B cells were treated with a well-known PAFR antagonist, WEB-2086 at concentrations ranging from 100 nM to 100 µM for an hour. We observed a significant decrease in CDSE-induced PAFR expression upon treatment with WEB-2086 as shown in Fig. 4. The treated BEAS-2B cells were then challenged with FITC-tagged NTHi. The adherence of NTHi to both CSE- and CDSE- exposed BEAS-2B cells was reduced in the presence of WEB-2086 in a dose-dependent manner (Fig. 5). Compared to mock-stimulated cells, NTHi adhesion to WSE-exposed BEAS-2B cells was not affected by treatment with WEB-2086 ($p > 0.05$) (Fig. 5). C17, an analogue of WEB-2086 was also tested for its activity towards the expression of PAFR and adhesion of NTHi to CSE, CDSE and WSE-exposed bronchial epithelial cells. Consistent with WEB-2086 results, C17 exhibited significant inhibition of CSE-, CDSE- and WSE-induced expression of PAFR on the airway epithelial cells (Fig. 4). Notably, C17 exhibited significant abrogation of attachment of NTHi to both CSE- and CDSE-treated bronchial epithelial cells in a dose-dependent manner (Fig. 6).

Transcriptional Response of BEAS-2B Cells Following CSE, CDSE, and WSE Exposure

The effect of CSE-, CDSE-, and WSE-exposure on BEAS-2B cell expression of a known receptor of NTHi, PAFR, was investigated at the transcriptional level. PAFR mRNA levels in 87.5 µg/mL CSE-, CDSE-, and WSE-exposed BEAS-2B cells increased by 9.0 ($p < 0.001$), 7.2 ($p < 0.01$) and 9.5-fold ($p < 0.001$), respectively, compared to mock (1% DMSO) treated BEAS-2B cells (Fig. 7). mRNA transcripts for the pro-inflammatory cytokines, interleukin 6 (IL-6) and interleukin 8 (IL-8), were significantly induced in BEAS-2B cells by 87.5 µg/mL CSE, CDSE, and WSE as illustrated in Fig. 7. In comparison to the mock (1% DMSO) treated control BEAS-2B cells, the mean IL-6 and IL-8 levels of expression were approximately 8.4 ($p < 0.001$) and 11.2 ($p < 0.001$) times higher in CSE-treated cells, respectively; 7.2 ($p < 0.001$) and 11.9 ($p < 0.001$) times higher in CDSE-stimulated cells, respectively; and 6.6 ($p < 0.001$) and 12.2 ($p < 0.001$) times higher in WSE-exposed cells, respectively.

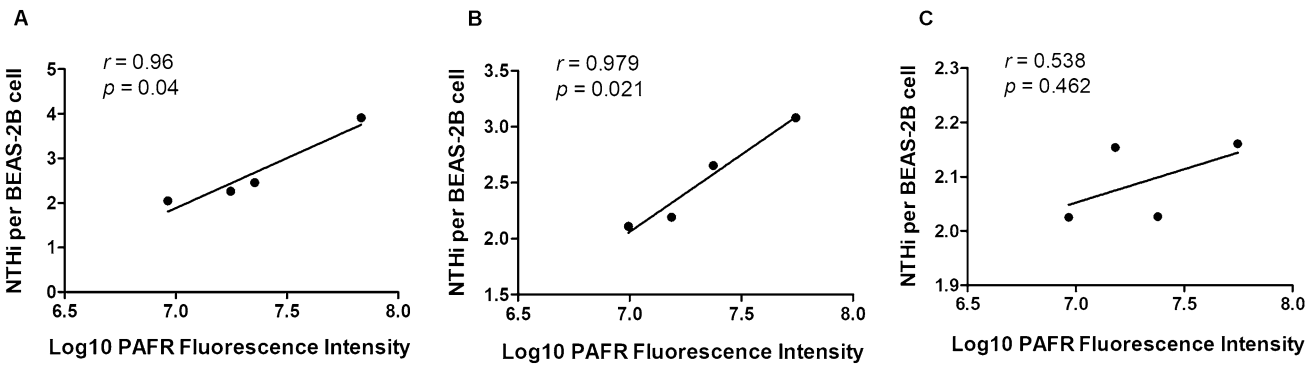


Fig. 3 Correlation between PAFR expression and NTHi adhesion on BEAS-2B cells exposed to **a** CSE **b** CDSE **c** WSE. PAFR expression corresponds to log10 of fluorescence intensity, which is significantly positively correlated with the number of NTHi attached per BEAS-2B cell for CSE and CDSE exposures. For WSE exposure, NTHi adhesion is not associated with expression of PAFR. Data are presented from four independent replicates from two experiments. *CSE* cigarette smoke extract, *CDSE* cow dung smoke extract, *WSE* wood smoke extract, *NTHi* nontypeable *Haemophilus influenzae*, *PAFR* platelet-activating factor receptor

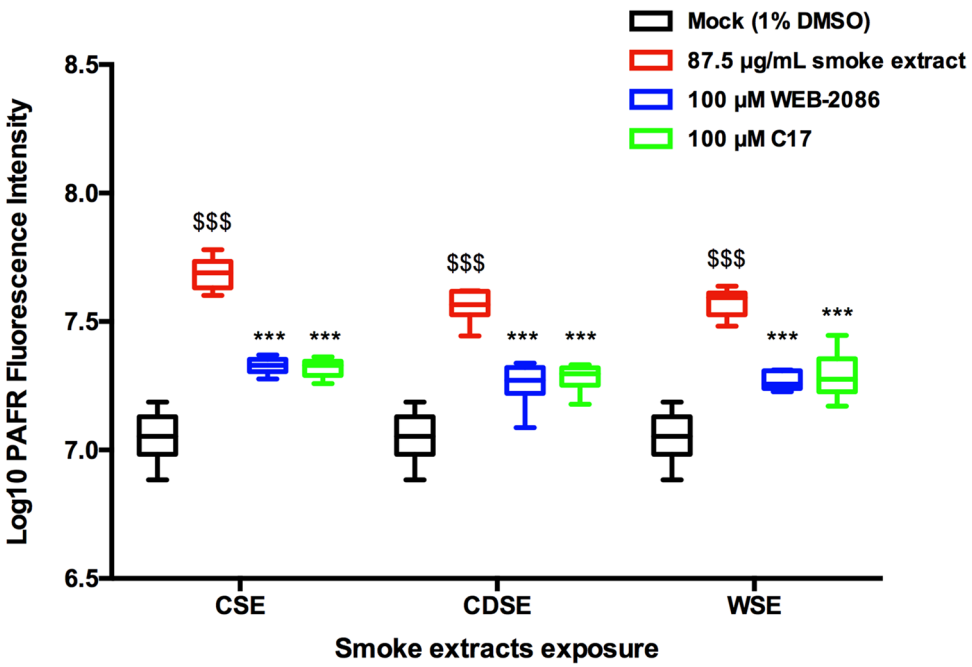


Fig. 4 Effect of smoke extracts (CSE, CDSE and WSE) on levels of PAFR protein expression on bronchial epithelial cells and inhibition by PAFR antagonists WEB-2086 and C17. PAFR expression corresponds to $\log_{10}$ of fluorescence intensity following labelling with Alexa Fluor 594 conjugate antibody targeting anti-PAFR mAb in immunofluorescence assays. The data represent the mean ± standard error for four independent replicates from two independent experiments. $^{\$\$\$}p < 0.001$ relative to 1% DMSO treated control cells, $^{***}p < 0.001$ relative to vehicle-treated smoke extract-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. *CSE* cigarette smoke extract, *CDSE* cow dung smoke extract, *WSE* wood smoke extract, *PAFR* platelet-activating factor receptor, *DMSO* dimethyl sulfoxide

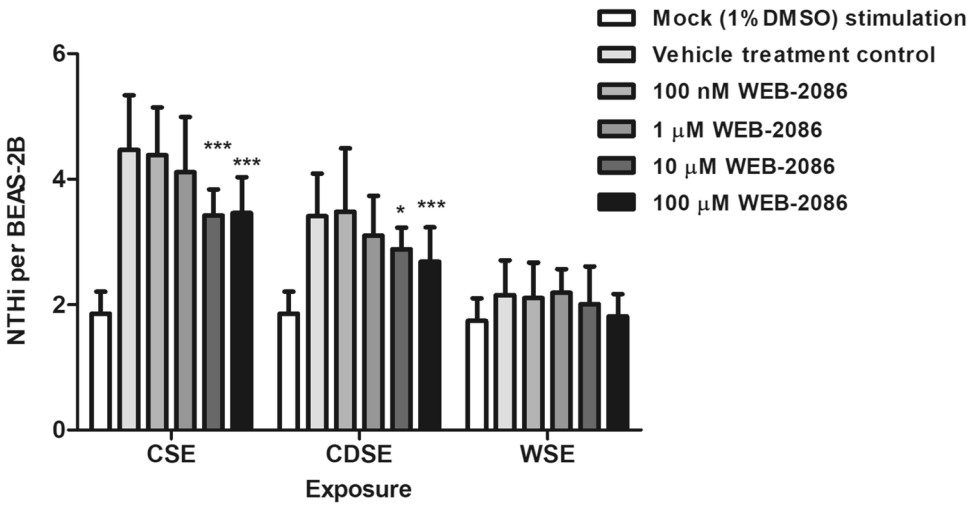


Fig. 5 Effect of PAFR antagonist, WEB-2086, on attachment of NTHi to CSE and CDSE-exposed BEAS-2B cells. The BEAS-2B cells were exposed to 1% DMSO or 87.5 µg/mL of CSE, CDSE, or WSE for 4 h. This was followed by treatment with vehicle (1% DMSO) or PAFR antagonist, WEB-2086, at a concentration range of 100 nM to 100 µM for an hour and subsequent challenge with FITC-labelled NTHi for an additional hour. WEB-2086 exhibited dose-dependent inhibition of NTHi adhesion to the CSE and CDSE-exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; $^{*}p < 0.05$, $^{***}p < 0.001$, relative to vehicle-treated CSE-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. *CSE* cigarette smoke extract, *CDSE* cow dung smoke extract, *WSE* wood smoke extract, *NTHi* nontypeable *Haemophilus influenzae*, *PAFR* platelet-activating factor receptor, *DMSO* dimethyl sulfoxide, *FITC* fluorescein isothiocyanate

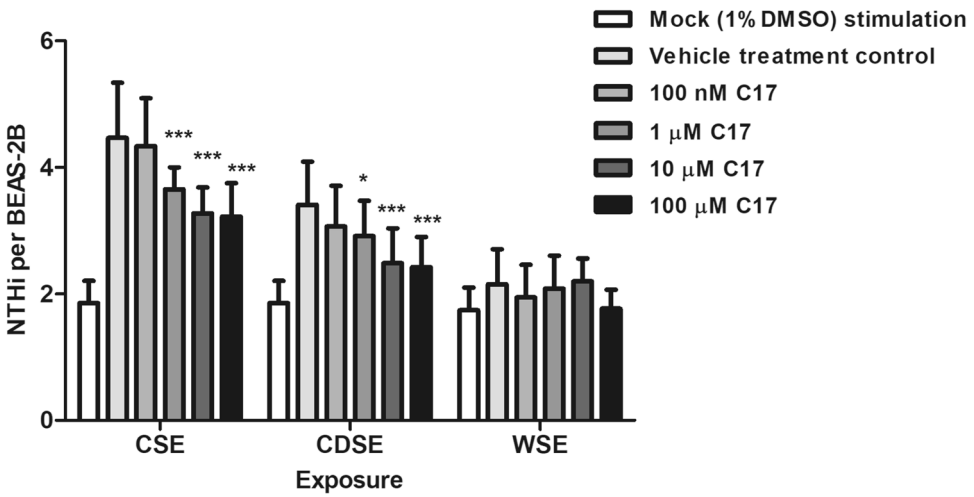


Fig. 6 Effect of new compound, C17, on attachment of NTHi to CSE and CDSE-exposed BEAS-2B cells. The BEAS-2B cells were exposed to 1% DMSO or 87.5 μg/mL of CSE, CDSE, or WSE for 4 h. This was followed by treatment with vehicle (1% DMSO) or compound C17 at a concentration range of 100 nM to 100 μM for an hour and subsequent challenge with FITC-labelled NTHi for an additional hour. C17 exhibited dose-dependent inhibition of NTHi adhesion to the CSE and CDSE-exposed BEAS-2B cells. The data represent the mean ± standard error for four independent replicates (comprising 5 images from each replicate) from two experiments; $*p < 0.05$, $***p < 0.001$, relative to vehicle-treated CSE-exposed cells, one-way ANOVA with Dunnett's multiple comparison test. *CSE* cigarette smoke extract, *NTHi* nontypeable *Haemophilus influenzae*, *PAFR* platelet-activating factor receptor, *DMSO* dimethyl sulfoxide, *FITC* fluorescein isothiocyanate

Effects of the WEB-2086 and the C17 Treatment on Induced Transcriptional Expression of PAFR and Inflammatory Cytokines IL-6 and IL-8

The effects of WEB-2086 and C17 on the transcriptional expression of PAFR, IL-6 and IL-8 in CSE-, CDSE-, and WSE-treated BEAS-2B cells were also examined. Treatment of smoke extract-exposed BEAS-2B cells with WEB-2086 or C17 for an hour significantly downregulated the expression of PAFR mRNA (Fig. 7). WEB-2086 and C17 at a concentration of 100 μM reduced the levels of PAFR mRNA by 54% ($p < 0.05$) and 56% ($p < 0.05$), respectively. Compared to CSE-exposed BEAS-2B cells (87.5 μg/mL) that were treated with vehicle (1% DMSO), transcriptional levels of IL-6 and IL-8 were significantly inhibited by 68% ($p < 0.001$) and 73% ($p < 0.001$), respectively, with 100 μM WEB-2086 treatment, and by 85% ($p < 0.001$) and 79% ($p < 0.001$), respectively, with 100 μM C17 treatment. In addition, both WEB-2086 and C17 treatment downregulated the CDSE-induced expression of IL-6 in a dose-dependent manner *i.e.* for 10 μM versus 100 μM of WEB-2086 and C17, 49% ($p < 0.001$) versus 60% ($p < 0.001$) and 43% ($p < 0.01$) versus 69% ($p < 0.001$), respectively. Similarly, the levels of IL-8 mRNA were also reduced in a dose-dependent manner by both WEB-2086 and C17 *i.e.* 68% ($p < 0.001$) versus 81% ($p < 0.001$) and 79% ($p < 0.001$) versus 83% ($p < 0.001$), for 10 μM versus 100 μM of WEB-2086 and C17 treatments, respectively (Fig. 7). Furthermore, the compounds were also found to suppress the expression of IL-6 and IL-8 in WSE-treated cells.

Tolerance of Compounds WEB-2086 and C17 by Human Bronchial Epithelial Cells

WEB-2086 and C17 were tested for toxicity towards BEAS-2B cells across the concentration range of 100 nM to 100 μM using the Alamar Blue assay. No significant signs of cytotoxicity were seen after 24-h treatment. The reduction of Alamar Blue by BEAS-2B cells was similar for treatment with both vehicle (1% DMSO) and 100 μM WEB-2086 *i.e.* 49.2% versus 44.8% at 4 h ($p > 0.05$), and 102.3% versus 104.1% at 24 h ($p > 0.05$). Furthermore, the viability of the bronchial epithelial cells was similar across treatment with vehicle and 100 μM C17 *i.e.* 49.2% versus 45.9% at 4 h ($p > 0.05$), and 102.3% versus 102.5% at 24 h ($p > 0.05$). Thus, both compounds, WEB-2086 and C17, were found to be tolerated by BEAS-2B cells up to the highest concentration tested of 100 μM.

Discussion

Nearly 4 Mio. deaths annually have been attributed to household air pollution (WHO 2018a, b). A number of meta-analyses of epidemiological studies have reported an overall 1.4 to 2.8-fold increased risk of COPD development in solid biomass fuel users compared to clean fuel users (Hu et al.

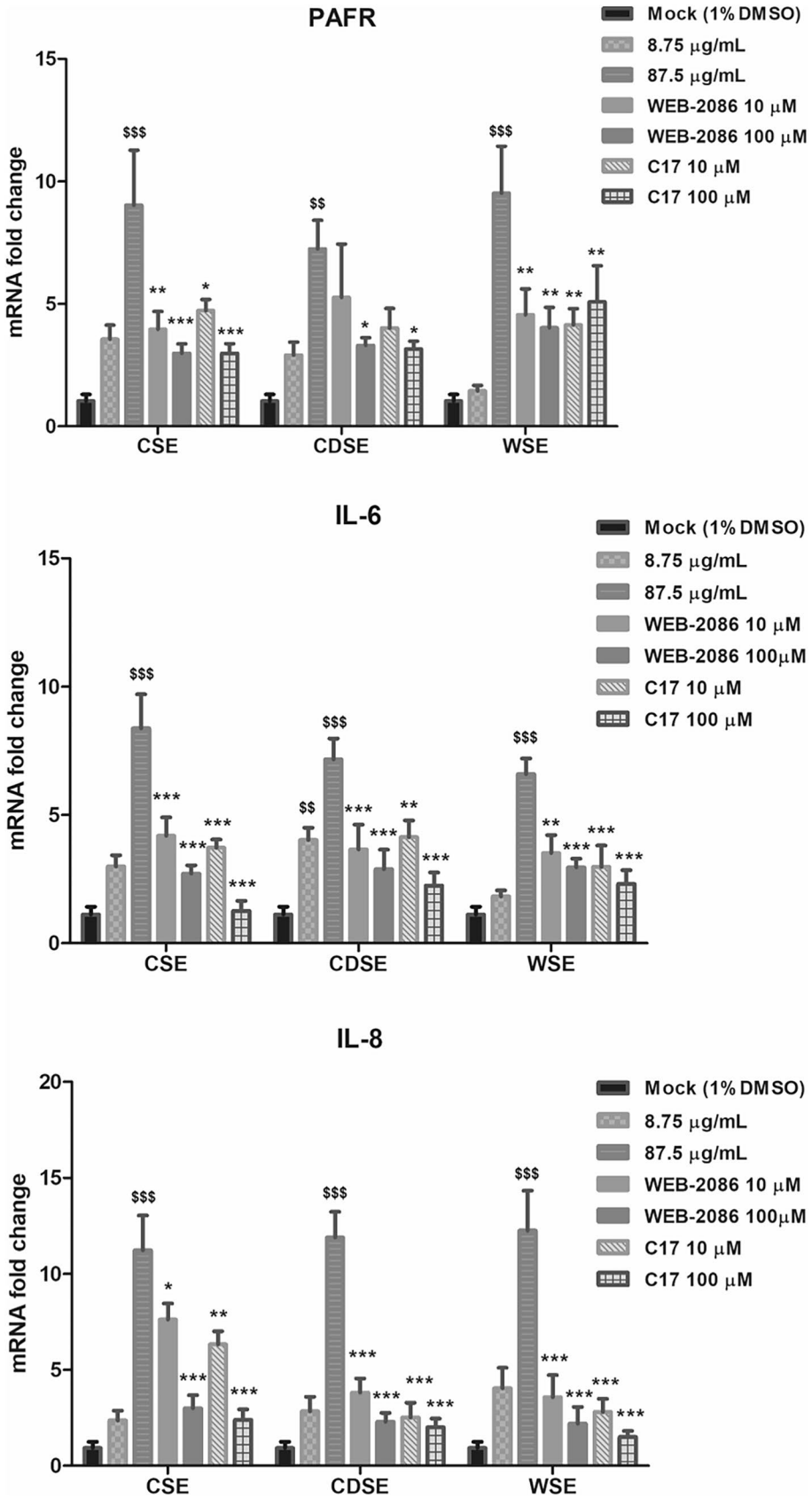


Fig. 7 Relative fold change in PAFR, IL-8, and IL-6 mRNA normalized with GAPDH in BEAS-2B cells exposed to CSE, CDSE, and WSE for 4 h at concentrations 8.75 and 87.5 µg/mL following treatment with PAFR antagonists, WEB-2086 and C17 at concentrations 10 µM and 100 µM for an hour. The data are presented as the mean of the observed fold-change ± standard error for four independent replicates from two experiments; $^{\$\$}p < 0.01$, $^{\$\$\$}p < 0.001$, relative to mock (1% DMSO) treated control; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, relative to the 87.5 µg/mL smoke extract-exposed BEAS-2B cells; two-way ANOVA with Bonferroni post-hoc test. *CSE* cigarette smoke extract, *CDSE* cow dung smoke extract, *WSE* wood smoke extract, *NTHi* nontypeable *Haemophilus influenzae*, *PAFR* platelet-activating factor receptor, *IL-6* interleukin 6, *IL-8* interleukin 8, *GAPDH* glyceraldehyde-3-phosphate dehydrogenase

2010; Kurmi et al. 2010; Po et al. 2011; Sana et al. 2018; Yang et al. 2017). A recent PUMA (Prevalence Study and Regular Practice, Diagnosis, and Treatment Among General Practitioners in Populations at Risk of COPD in Latin America) study conducted in primary care settings in four Latin American countries reported a higher proportion of subjects with COPD were exposed to biomass smoke compared to those without COPD (Montes de Oca et al. 2017). Da Silva and colleagues examined the dose–response relationship and reported a significant correlation between reduced pulmonary function and the level and duration of biomass exposure (da Silva et al. 2012). Several short-term trials using improved cooking stoves and kitchen ventilation have provided support for the efficacy of interventions, suggesting an association between biomass smoke exposure and lung function impairment (Pope et al. 2015; Romieu et al. 2009; Zhou et al. 2014).

Smith et al., in a population-based study of approximately 317,000 non-smoking Chinese adults, reported airflow obstruction to be positively associated with cooking with coal but not with other sources of household air pollution (Smith et al. 2014). Recently, two large-scale multinational studies were conducted to address the role of household air pollution in airflow obstruction and were published in the same issue of the *American Journal of Respiratory and Critical Care Medicine* in 2018 (Amaral et al. 2018; Siddharthan et al. 2018). Amaral and colleagues, using data from the BOLD (Burden of Obstructive Lung Disease) study, reported no evidence of association between airflow obstruction and the use of solid fuels (Amaral et al. 2018). However, they reported a 28% higher risk of chronic phlegm in never-smoker women who had used solid fuels for cooking or heating over twenty years or more (Amaral et al. 2018). Siddharthan and co-investigators analyzed population data from the PRISA (Pulmonary Risk in South America) and CRONICAS cohort studies in South America, a longitudinal study conducted by the Centre for Control of Chronic Diseases in Bangladesh, and the LINK (Lung Function Study in Nakaseke and Uganda) and FRESH AIR studies in Africa. They reported that study participants with household air pollution exposure from biomass fuel use were 41% more likely to have COPD than non-exposure individuals and that the association between household air pollution exposure and COPD was more pronounced in women (adjusted odds ratio 1.70; 95% confidence interval, 1.24–2.32) than in men (1.21; 0.92–1.58) (Siddharthan et al. 2018). The findings from the above two studies are, therefore, not in agreement and it has been acknowledged that the lack of direct measurement of exposure and level of ventilation may have impacted the studies (Balmes and Eisen 2018). The role of biomass smoke in airway obstructive disease requires further extensive research including examining the effect of biomass smoke at a cellular level, which was the focus of our study.

In this work, we report for the first time that exposure of bronchial epithelial cells to biomass smoke in the form of cow dung smoke exposure results in increased adhesion by NTHi, a major respiratory bacterial pathogen in the chronic respiratory disease COPD. NTHi are opportunistic bacterial pathogens that are normal commensals of the upper respiratory tract (Bakaletz and Novotny 2018). However, in COPD patients, due to impaired immunity they colonize the lower airways of COPD patients, thereby inducing inflammation, resulting in a worsening of symptoms and an increased frequency of exacerbations (Leung et al. 2017). More than half of acute COPD exacerbations are caused by bacterial infections, particularly NTHi (Sethi et al. 2002). This study suggests that cow dung smoke exposure could translate into an increased susceptibility to NTHi infection in the respiratory tract of patients with biomass smoke associated COPD.

We further investigated the effect of CDSE and WSE on airway cells. As shown in Fig. 7, expression of both IL-6 and IL-8 was upregulated several-fold in bronchial epithelial cells following exposure to CDSE and WSE with respect to controls. IL-6 is a major stimulator of the production of acute phase proteins (Castell et al. 1989) and affects the Th17/Treg balance which can contribute to the development of autoimmune and chronic inflammatory diseases (Kimura and Kishimoto 2010). IL-8, on the other hand, is a key regulator in neutrophil-mediated acute inflammation via neutrophil transmigration and activation, including degranulation and the release of leukotrienes and platelet-activating factor (Mukaida 2000, 2003; Schröder 1989). A significant upregulation of expression of pro-inflammatory cytokines, IL-6 and IL-8 upon exposure to biomass smoke plays an important role in the pathogenesis of chronic inflammatory diseases, particularly in COPD (Gabay 2006; Oishi et al. 1994; Tamimi et al. 2012). The results obtained here are, therefore, in agreement with human and animal model studies. Increased sputum and serum levels of IL-6 and IL-8 have been observed in women cooking with biomass fuels compared to the clean fuel users (Dutta et al. 2012, 2013). Similar upregulation of cytokines was also observed in a mouse model of biomass exposure and in the in vitro exposure of biomass smoke to primary human small airway epithelial cells (McCarthy et al. 2016; Sussan et al. 2014). This may in part be linked to the potential role of biomass smoke exposure in airway obstructive disease in never smokers, particularly, in low- and middle-income countries where biomass fuel is the primary source of domestic energy (KC et al. 2018a, b).

Similar to CDSE, WSE exposure also increases the expression of PAFR (Figs. 4 and 7). However, WSE exposure did not increase the attachment of NTHi to respiratory epithelial cells. There are two possible reasons for a lack of a measured effect of WSE on NTHi adhesion to bronchial epithelial cells. Firstly, the possible presence of compounds

Journal of Biological Methods | **2019** | **Vol.** 6(1) | e110
DOI: 10.14440/jbm.2019.276

PROTOCOL

A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens

Sanjay S. Gautam[1], Rajendra KC[1], Kelvin WC Leong[3], Micheál Mac Aogáin[2], Ronan F. O'Toole[1,2,3]*

[1]School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, TAS 7001, Australia
[2]Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin D08 W9RT, Ireland
[3]School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria 3690, Australia

*Corresponding Author: Ronan F. O'Toole, Email: r.otoole@latrobe.edu.au

Competing interests: The authors have declared that no competing interests exist.

Abbreviations used: DAL, diluted amplified library; NGS, next-generation sequencing; NTA, Nextera XT tagment amplicon; PAL, pooled amplified library; PCR, polymerase chain reaction; VNTR, variable number tandem repeat; WES, whole exome sequencing; WGS, whole genome sequencing

Received September 28, 2018; Revision received January 13, 2019; Accepted January 28, 2019; Published March 15, 2019

ABSTRACT

Bacterial whole genome sequencing (WGS) is becoming a widely-used technique in research, clinical diagnostic, and public health laboratories. It enables high resolution characterization of bacterial pathogens in terms of properties that include antibiotic resistance, molecular epidemiology, and virulence. The introduction of next-generation sequencing instrumentation has made WGS attainable in terms of costs. However, the lack of a beginner's protocol for WGS still represents a barrier to its adoption in some settings. Here, we present detailed step-by-step methods for obtaining WGS data from a range of different bacteria (Gram-positive, Gram-negative, and acid-fast) using the Illumina platform. Modifications have been performed with respect to DNA extraction and library normalization to maximize the output from the laboratory consumables invested. The protocol represents a simplified and reproducible method for producing high quality sequencing data. The key advantages of this protocol include: simplicity of the protocol for users with no prior genome sequencing experience and reproducibility of the protocol across a wide range of bacteria.

Keywords: whole genome sequencing, *Enterococcus faecium*, *Haemophilus influenzae*, *Mycobacterium tuberculosis*

BACKGROUND

Using Sanger sequencing, the Human Genome Project expended approximately USD $2.7 billion and took more than 10 years to produce the first human genome sequence. Today, a human genome can be sequenced in a matter of days for less than USD $1000 on a single next-generation sequencing (NGS) machine. This step change in throughput and per-base cost has transformed the use of DNA sequencing in biomedical research and is being translated in an expanding number of ways into medicine. NGS is increasingly being applied to understanding and managing infectious diseases. This includes the sequencing of microbial genomes for the purposes of laboratory identification of infectious agents [1], detection of antibiotic resistance markers [2], and the public health surveillance of epidemiological clusters and outbreaks [3]. Examples include its deployment in public health surveillance and control of community cases of *Escherichia coli* [4], *Campylobacter jejuni* [5], *Legionella pneumophila* [6] and *Mycobacterium tuberculosis* [7] disease, or global and regional epidemics caused by influenza [8], Ebola [9], and Zika [10] viruses. It has also been utilised to track the source and spread of healthcare-associated infections caused by *Staphylococcus aureus* [11], *Pseudomonas aeruginosa* [12], *Acinetobacter baumannii* [13], and *Enterococcus faecium* [14] in order to guide infection prevention and control in hospitals.

In addition to its whole genome (WGS), whole exome (WES), transcriptome (RNA-Seq), bisulphite methylome, and metagenomic sequencing capabilities, NGS can be directed to the detection of specific genes or mutations associated with human disease through targeted-panel amplicon screening. However, barriers remain with regard to establishing NGS in a laboratory for the first time and this hinders its uptake in clinical microbiology and other settings. One of these challenges is the lack of a simplified step-by-step protocol that can be picked up by laboratory personnel with no prior training or experience in NGS and used to generate reliable, high quality sequence data. Illumina dye-sequencing is currently considered the gold standard internationally in terms of read depth and base-calling accuracy, genome coverage, scalability, and the range of sequencing applications it delivers.

In this work, we produced an easy-to-follow, step-by-step NGS protocol with consistent genome coverage and average read depth that

How to cite this article: Gautam SS, Rajendra KC, Leong KWC, Mac Aogáin M, O'Toole RF. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *J Biol Methods* 2019;6(1):e110. DOI: 10.14440/jbm.2019.276

was applicable to a range of bacterial pathogens *i.e.*, Gram-positive vancomycin-resistant *Enterococcus faecium*, Gram-negative non-typeable *Haemophilus influenzae*, and acid-fast high-GC content *Mycobacterium tuberculosis*. This protocol can be used to generate Illumina-based WGS data for clinical isolates of bacterial pathogens of importance to human health.

Figure 1 is the graphical summary of the process of obtaining whole genome sequence data from bacterial culture. This wet laboratory procedure generated FastQ reads from the sequencer within three days of start. We modified a number of the DNA extraction steps to obtain a sufficient quantity of contamination free template. Similarly, we replaced library normalization plates and Nextera XT tagment amplicon (NTA) plates with conventional polymerase chain reaction (PCR) tubes which may represent a cost-effective alternative. In addition, we have recommended the use of equal DNA concentrations of each library during library normalization to ensure better coverage and minimize bias. Simplification of bacterial NGS may assist in its uptake by beginner users.

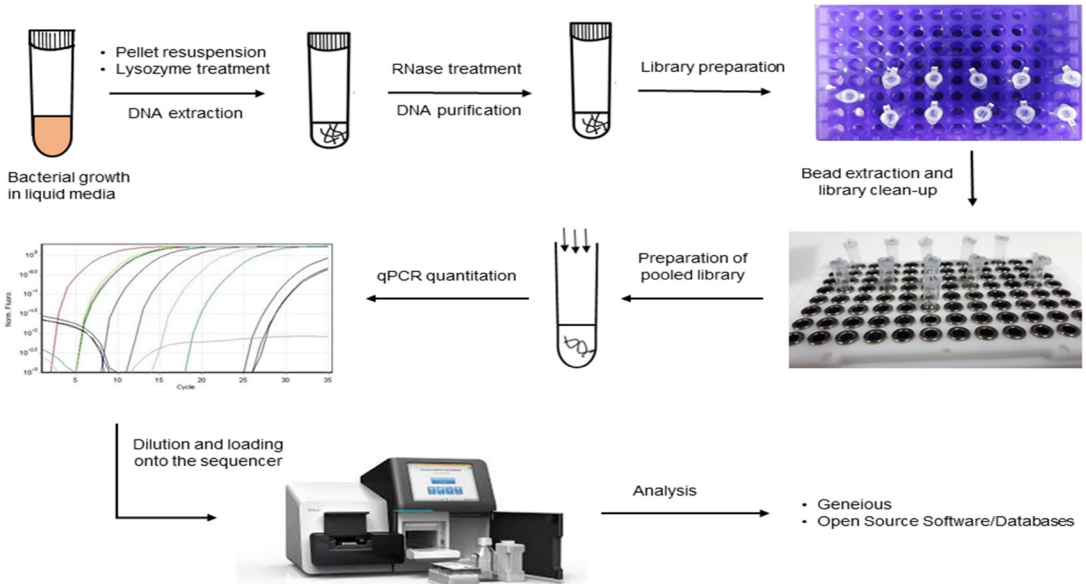


Figure 1. Graphical summary of the process of obtaining whole genome sequence data from a bacterial culture.

MATERIALS

Reagents

- ✓ Lysozyme (VWR, Australia, Cat.# 0663-10G)
- ✓ Ethanol, Pure (Sigma-Aldrich, Australia, Cat. # E7023)
- ✓ 2-Propanol (Sigma-Aldrich, Australia, Cat. # I8912)
- ✓ Phosphate Buffered Saline (Gibco™- Thermo Fisher Scientific, UK, Cat. # 10010023)
- ✓ Ultrapure™ DNase/RNase Free Distilled Water (Invitrogen, Australia Cat. # 10977-015)
- ✓ DNeasy® Blood and Tissue Kit (Qiagen, Germany, Cat. # 69504)
- ✓ High Pure PCR Template Preparation Kit (Roche, Germany, Cat. # 11796828001)
- ✓ Qubit™ dsDNA HS (High Sensitivity) Assay Kit (Invitrogen, Australia, Cat. # Q32851)
- ✓ Nextera® DNA Library Preparation Kit (Illumina, USA, Cat. # FC-121-1030)
- ✓ Nextera® XT Library Preparation Kit (Illumina, USA, Cat. # FC-131-1024)
- ✓ Nextera® XT Index Kit (Illumina, USA, Cat. # FC-131-1001)
- ✓ Miseq Reagent Kit v2 (300 cycles) (Illumina, USA, Cat. # MS-102-2002)
- ✓ KAPA Library Quantification Kit (Illumina, USA, Cat. # 07960140001)
- ✓ Agencourt® AMPure XP beads (Beckman Coulter, USA, Cat. # A63880)

Recipes

- ✓ Qubit working solution: dilute Qubit dsDNA HS Reagent 1:200 in Qubit dsDNA HS buffer. For *n* samples, prepare *n* × 200 µl working solution.
- ✓ 80% ethanol: add 2 ml absolute ethanol into 8 ml distilled water.
- ✓ 0.2 M NaOH: weigh 0.04 g of NaOH pellet and dissolve it into 5 ml distilled water.

Equipment

- ✓ Qubit™ assay tubes (Life-technologies, USA, Cat. # Q32856)
- ✓ PCR tubes (Molecular Bioproducts, USA, Cat. # MBP3412)
- ✓ Qubit® 2.0 Fluorometer (Invitrogen, Australia, Cat. # Q32866)
- ✓ Agencourt Magnetic stand (Beckman Coulter, USA, Cat. # A32782)
- ✓ Applied Biosystems® Veriti 96-Well thermal cycler (Thermo Fisher Scientific, USA)
- ✓ Rotor-Gene 6000 real-time thermocycler (Corbett Research, Australia)

PROCEDURE

Extraction of bacterial genomic DNA

1. Pellet the liquid culture media (200 µl) by centrifuging at 8000 *g* for 8 min in a sterile microfuge tube.

   CRITICAL STEP: All bacterial cultures should be treated as potentially pathogenic to the laboratory worker and colleagues. Therefore, the use of appropriate aseptic techniques, and the wearing of appropriate personal protective equipment are strongly recommended to maintain acceptable work health and safety standards and minimise exposure to harmful agents.

2. Resuspend the pellet in 600 µl phosphate-buffered saline (1×) until the absorbance at 600 nm (A_{600}) is between 1.0 and 2.0. Lyse the cells by adding 30 µl lysozyme (50 mg/ml), vortex, and incubate at 37°C for 1 h.
3. Follow the DNeasy® Blood and Tissue Kit Quick-start protocol to extract the DNA.
4. Elute the DNA in 100 µl volume and treat it with 2 µl RNase (100 mg/ml) (Qiagen, Hilden, Germany) and incubate at room temperature for 1 h.
5. Purify RNase-treated DNA using the High Pure PCR Template Preparation Kit.

   TIP: Perform only 4 DNA spin-wash steps instead of 9 recommended steps. Pre-incubate the elution buffer in a heat block set at 70°C.

6. Add 100 µl of binding buffer to RNase treated DNA and incubate at 70°C for 10 min.
7. Add 50 µl of 2-Propanol and transfer the content to a Roche spin column and spin at 8000 *g* for 1 min.
8. Discard the flow through and insert the spin column into a new collection tube.
9. Wash by adding 500 µl wash buffer and spin at 8000 *g* for 1 min.
10. Discard the flow through and insert the spin column into a new collection tube.
11. Perform a final spin at 8000 *g* for 1 min.
12. Finally, insert the column into a 1.5 ml sterile microfuge tube, add 50 µl of pre-heated elution buffer and spin at 8000 *g* for 1 min to elute the purified DNA for next generation sequencing.

    CRITICAL STEP: For next generation sequencing, contaminant-free, high-molecular weight DNA with an absorbance (260 nm/280 nm) ratio between 1.8 to 2.0 is considered a high-quality template DNA [15].

Quantification of bacterial genomic DNA

13. Dispense 190 µl and 198 µl of Qubit working solution in standard and sample tubes, respectively.
14. Add 10 µl standards (1 and 2) and 2 µl of sample in separate Qubit assay tubes.
15. Vortex the mixture for 3 s and incubate at room temperature for 2 min before taking the reading.
16. Adjust the DNA concentration of each sample to 0.2 ng/µl by diluting with a required volume of distilled water.

    CRITICAL STEP: The use of an accurate concentration of DNA is crucial for bacterial DNA genomic library preparation.

Tagmentation and PCR amplification of bacterial genomic DNA

TIP: For all of the methods below, the recommended 96-well TYC plate can be replaced with 0.2 ml thin wall clear, flat capped PCR tubes. In addition, multichannel pipettes and the high-speed micro plate shaker can be replaced with single channel pipettes and a bench top centrifuge, respectively.

Nextera XT tagment amplicon construction

17. In a PCR tube, add 5 µl tagmentation DNA buffer and 2.5 µl amplification tagmentation mix to 2.5 µl (0.2 ng/µl) input DNA.

18. Briefly vortexed the content and transfer to the thermocycler programmed for one step at 55°C for 5 min with heated lid, followed by a hold at 10°C for a volume of 10 µl.

Neutralization of Nextera XT tagment amplicon

19. Immediately after reaching the hold temperature of 10°C in the above step, neutralize NTA by adding 2.5 µl neutralization tagmentation buffer and incubate at room temperature for 5 min.

PCR amplification

20. For amplification, add 7.5 µl Nextera® PCR mastermix and 2.5 µl of each index primer, 1 and 2, to a tube containing neutralized NTA.

    CRITICAL STEP: Primer combinations, S502 with N705 /706 and S503 with N701/702 should be avoided. Avoid any repeated combinations and carefully note the primers used for each sample.

21. Gently pipette the content and perform a quick spin.

22. Proceed to amplification in a thermocycler programmed for a working volume of 25 µl with the following settings: heated lid, initial cycle at 72°C for 3 min followed by 95°C for 30 s and 12 cycles of (95°C for 10 s, 55°C for 30 s and 72°C for 30 s) with a final run at 72°C for 5 min followed by a hold temperature of 10°C. The amplified, tagmented library can be stored at 2 to 4°C overnight for PCR clean-up the next day.

Cleaning up the PCR product

NOTE: Bring AMPure XP beads to room temperature (for 20 min).

CRITICAL STEP: Prepare fresh 80% (v/v) ethanol and 0.2 M NaOH.

23. To 22.5 µl of PCR product, add 11.25 µl of vortexed (30 s) AMPure XP beads and mix by pipetting (10 times).

24. Incubate at room temperature for 5 min.

25. Place the tube on a magnetic stand for 2 min.

26. While leaving the PCR tubes on the magnetic stand, carefully aspirate the supernatant.

    CRITICAL STEP: Do not aspirate beads. If aspirated, redo steps 25 and 26.

27. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.

    CRITICAL STEP: Do not resuspend the beads.

28. Aspirate out the supernatant carefully.

29. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.

30. Aspirate out the supernatant carefully.

31. Remove the tube from magnetic stand and allow to air dry in a tube stand for approximately 5 min.

    CRITICAL STEP: Visually check for cracks as over drying the beads will significantly reduce elution efficiency.

32. Add 26.15 µl of resuspension buffer and gently pipette 20 times to mix.

33. Incubate the tubes at room temperature for 2 min and then place on a magnetic stand for 2 min (until the supernatant cleared).

34. Transfer the supernatant (25 µl) to a new PCR tube.

    NOTE: The final supernatant can be stored at −15°C to −20°C for up to 1 week but we recommend proceeding to library normalization immediately.

Library normalization

35. Perform the Qubit DNA quantification method as described above to determine the genomic DNA concentration in cleaned up product.

36. Pool the genomic DNA from all of the tubes.

    NOTE: Sample with the lowest DNA concentration can be used in a volume of 10 µl as the reference to prepare a library pool using the formula: Volume $_{required}$ (V2) = Concentration $_{original}$ (S1) × Volume $_{total}$ (V1 = 10 µl) / Concentration $_{required}$ (S2).

37. To × µl of library pool, add × µl of freshly prepared 0.2 molar NaOH (final concentration 0.1 molar) and incubate for 5 min at room temperature.

38. To the NaOH treated suspension add an equal volume (2× µl) of LNS1. Label the tube as pooled amplified library (PAL).

    TIPS: In this modified step, normalize the library by using LNS1 (Library Normalisation Storage Buffer 1) only.

39. Dilute the PAL to 1:1000 by adding 1 µl of NaOH-LNS1 treated suspension to 999 µl of ultrapure distilled water.

40. Use KAPA library quantification kit (No ROX) to check the concentration of diluted pooled library in a real time PCR system using the following set up: a hot start run at 95°C for 10 min followed by 40 cycles of (95°C for 10 s and 60°C for 30 s).

    NOTE: Include a set of six DNA standards (with concentrations ranging from 20 pM to 0.0002 pM), three sets of negative control (ultrapure distilled water), and three sets of the DNA library in the qPCR run.

41. Determine the concentration of DNA in the pooled library by the standard curve method and calculate concentration in picomolar (pM) for each tube.

    NOTE: To calculate the original concentration of the pooled library we applied the formula:

    [Average sample concentration (in pM) × insert size standards (452 bp) × dilution factor (1000)] divisible by the Insert size of pooled library (500 BP)

    For example, for a qPCR determined concentration of 2.36 pM in a 1:1000 dilution of the pooled library, the library DNA concentration will be: (2.36 pM × 452 bp × 1000) / 500 bp = 2133.44 pM

    The value obtained from the calculation represents the concentration of DNA in the pooled library.

42. To estimate the dilution factor required to achieve a final library concentration of 15 pM in a 600 µl volume use the formula:

$$\begin{aligned} \text{Volume}_{required} &= (\text{Concentration}_{required} \times \text{Volume}_{total}) / \text{Concentration}_{original} \\ &= (15\ \text{pM} \times 600\ \mu\text{l}) / 2133.44\ \text{pM} \\ &= 4.22\ \mu\text{l} \end{aligned}$$

    NOTE: Therefore, 4.22 µl is added to 595.78 µl of HT buffer to produce a final concentration of 15 pM, in a final volume of 600 µl. The diluted library is then ready to be heat denatured and loaded into the MiSeq reagent cartridge.

Preparing pooled library for loading onto MiSeq

43. Thaw the PAL at room temperature and mix by pipetting up and down (5 times) followed by brief centrifugation.

44. Based on the library concentration example above, transfer 595.78 µl of HT buffer to a 1.5 ml diluted amplified library (DAL) tube containing 4.22 µl PAL.

45. Mix using a pipette (5 times).

46. Vortex the DAL tube at top speed, centrifuge briefly, and incubate exactly for 2 min at 96°C ± 2°C.

47. Immediately transfer the DAL tube to ice for at least 5 min or until loading.

CRITICAL STEP: Put the Illumina MiSeq sequencer through a short wash cycle to avoid cross-contamination of the DAL from previous usage.

48. Thaw the MiSeq reagent cartridge at room temperature [16].

49. Generate a MiSeq sample sheet using the Illumina Experiment Manager. See step 21 to identify primer sets for each sample.

50. Use the following configuration to set up the Miseq machine. Generate FASTQ workflow; FASTQ Only application; NexteraXT assay; 151 insert reads; assignment of the samples with a unique identifier and index-pair combination.

CRITICAL STEP: Rinse the flow cell with MilliQ water and remove traces of water using soft tissue paper before inserting into the machine.

51. Transfer the entire 600 µl of DAL to the "Load here" well of the MiSeq reagent cartridge.

52. Following the setup procedure of the Illumina Experiment Manager, insert the cartridge into MiSeq instrument for sequencing to commence.

TIPS: The raw FastQ sequence reads from whole-genome sequencing can be stored on the local computer as well as on the Illumina BaseSpace server (https://basespace.illumina.com/) for further analysis.

Bioinformatic analyses

NOTE: The selection of bioinformatics software for the analysis of WGS data will be determined by the objective of the study. Here, we used Geneious 9.1.8 (Biomatters Ltd.) [17], a desktop software to analyse our sequence data. Geneious was used to map the Fastq sequence reads to a publicly available reference genome for each species as follows:

53. Download Geneious from https://www.geneious.com/.

54. Go to File | Import | From File. Import raw-read files (Sample_xx_R1.fastq.gz and Sample_xx_R2.fastq.gz) into Geneious.

55. Download the Reference Genome from the NCBI database. For example, *Enterococcus faecium* NC_017960.

- ***55.1.*** In the Left panel | Go to NCBI | Nucleotide.
- ***55.2.*** Enter NC_017960 | Click Search.
- ***55.3.*** Once the genome has been found, click Download Full Sequence(s).
- ***55.4.*** Download the NC_017960 reference genome (The icon changes to a green circular genome when completed).
- ***55.5.*** Drag and drop the NC_017960 reference genome into the working folder.

56. Mapping the isolate sequence to the reference genome

- ***56.1.*** Hold CTRL and select both R1 and R2 raw read files (imported), and the reference genome (NC_017960) (downloaded).
- ***56.2.*** Click Align | Assemble|Map to Reference.
- ***56.3.*** Check the settings
  - ***56.3.1.*** Reference Sequence = NC_017960
  - ***56.3.2.*** Mapper = Bowtie2–fast and accurate read mapper
  - ***56.3.3.*** Trim Before Mapping = Do not trim
  - ***56.3.4.*** Results: Select all options
  - ***56.3.5.*** Results | Save consensus sequences | Options
  - ***56.3.6.*** Threshold = Highest Quality
  - ***56.3.7.*** Threshold for sequences without quality = 95%
  - ***56.3.8.*** No coverage call = '–'

56.4. When mapping to reference is complete, a new folder will be created containing four files:

56.4.1. Assembly Report

56.4.2. Consensus

56.4.3. Contig

56.4.4. Unused Reads

NOTE: Setting may vary depending on objective of analyses and quality of fastq reads.

NOTE: We also used open source databases, for example, TGS-TB [18], PhyResSe [19] and the Center for Genomic Epidemiology's [20] ResFinder and VirulenceFinder [21], to further analyse the whole genome sequence data of our selection of bacterial pathogens. These freely-available databases enable the acquisition of information on bacterial pathogens that included genotype and phylogeny, antibiotic-resistance mutations, and the presence of known virulence genes.

ANTICIPATED RESULTS

A consensus sequence was generated for each of the isolates analysed in Geneious. The Geneious report provided information on the percentage coverage of test sequence to the reference genome and the mean read depth (**Table 1**). Each contiguous sequence is viewable in Geneious and can be analysed for coverage with respect to the reference genome. Quality control checks of raw sequence data were also performed using FastQC [22]. This freely-available software provided information regarding per base sequence content and quality, per base and sequence GC content, and highlighted the parameters of the sequence quality.

Initial typing analysis

We used open source databases to analyze the sequence data. For example, Geneious mapped contiguous sequences were imported into PubMLST (https://pubmlst.org/) for sequence typing of *Haemophilus influenzae* and vancomycin-resistant *Enterococcus faecium*. This can also be achieved using raw fastq reads in the MLST profiling tool from the Center for Genomic Epidemiology (CGE) database (http://www.genomicepidemiology.org/). The Resfinder tool (https://cge.cbs.dtu.dk/services/ResFinder/) was used to identify acquired antimicrobial resistance genes from raw fastq files. For example, PubMLST typing classified NTHi 1 as sequence type 46 and Resfinder did not detect the presence of any antimicrobial resistance determining mutations.

Mycobacterium tuberculosis complex raw fastq.gz files were uploaded to the TGS-TB database (https://gph.niid.go.jp/tgs-tb/) to predict drug susceptibility, *in silico* spoligotype, lineage type, and phylogenetic classification. This database also enabled detection of IS6110 insertion sites, and 43 loci for variable number tandem repeat (VNTR) typing. The drug resistance profile of the MTBC isolates were further confirmed using PhyResSE database (http://phyresse.org/). For example, TGS-TB identified MTBC1 as a drug susceptible *Mycobacterium bovis* isolate.

Table 1. Percentage sequence coverage and mean read depth for each of the sequenced genomes with respect to reference strains.

Sample	Reference sequence coverage (%)	Mean read depth
VRE1	90.98	197.4
VRE2	89.4	178.4
VRE3	90.96	168
NtHi1	93.8	263.5
NtHi2	86.5	61.3
NtHi3	89.1	124.9
MTBC1	96.4	104.4
MTBC2	96.8	100.4
MTBC3	96.8	68.4

Coverage refers to the percentage of reference genome bases covered by mapped sequence reads. Mean read depth indicates the mean number of times each base is mapped by a sequence read. Reference genomes used were *E. faecium* ST18 DO (TX16) (accession number NC_017960), *Haemophilus influenzae* 86-028NP (nontypeable) (accession number NC_007146), and *Mycobacterium tuberculosis* H37Rv (accession number NC000962). VRE, vancomycin resistant *Enterococcus faecium*; NtHi, non-typeable Haemophilus influenza; MTBC, *Mycobacterium tuberculosis* complex.

TROUBLESHOOTING

Possible problems and their troubleshooting solutions are listed in **Table 2**. There are a number of limitations associated with the protocol that should be noted. These include: effective results with the protocol are reliant on the efficacy of the extraction procedure in producing a

sufficient quantity of genomic DNA; analysis of sequences generated on an Illumina platform can be affected by the presence of highly repetitive regions; and depending on the output information sought, genome assembly can be influenced by the reference genome selected for the mapping of reads. Nevertheless, the protocol was effective in generating high quality sequencing data for the range of bacterial species tested.

Table 2. Troubleshooting table.

Step #	Problems	Causes	Suggestions
23	Low concentration of AMPure XP bead captured purified products	Bead clean-up affects the quality and quantity of amplified libraries that will be included in downstream sequencing process	Make sure AMPure XP beads are held at room temperature for 20 min before starting the clean-up process. Furthermore, ensure that 80% ethanol is freshly prepared
35	Variation in the concentration of amplified library	The concentration of input DNA used for library preparation affects the final yield of genomic data	Measure the concentration of input DNA using Qubit fluorometer rather than a nanodrop and make the appropriate dilution for a DNA concentration of 0.2 ng/µl
56	Poor sequencing results	The choice of forward and reverse index primer set affects the sequencing of libraries prepared	Avoid primer combinations S502 with N705 /706, and S503 with N701/702
56	Poor sequencing results	Effective denaturation of pooled library not achieved after bead clean-up	Ensure that the NaOH is freshly prepared at the correct concentration
56	Poor sequencing results	Repeated thawing and freezing of the pooled library reduces the quality of sequence reads generated	Before preparing the pooled library for loading onto MiSeq, ensure that the machine has already been appropriately cleaned after the previous run and has sufficient storage space (at least 25 GB)

Acknowledgments

This research was supported by funding from the Royal Hobart Hospital Research Foundation (17-104) and the Tasmanian Community Fund (36Medium00014).

References

1. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, et al. (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 15: 1193-1202. doi: 10.1016/S1473-3099(15)00062-6. PMID: 26116186
2. Takiff HE, Feo O (2015) Clinical value of whole-genome sequencing of Mycobacterium tuberculosis. Lancet Infect Dis 15: 1077-1090. doi: 10.1016/S1473-3099(15)00071-7. PMID: 26277037
3. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, et al. (2012) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13: 130-136. doi: 10.1016/S1473-3099(12)70268-2. PMID: 23158674
4. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci USA 109: 3065-3070. doi: 10.1073/pnas.1121491109. PMID: 22315421
5. Llarena A, Taboada E, Rossi M (2017) Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections. J Clin Microbiol 55: 1269-1275. doi: 10.1128/JCM.00017-17. PMID: 28249998
6. Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, et al. (2013) A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak. BMJ Open 3: e002175. doi: 10.1136/bmjopen-2012-002175. PMID: 23306006
7. Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, et al. (2018) Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS One 13: e0192351. doi: 10.1371/journal.pone.0192351. PMID: 29466411
8. McGinnis J, Laplante J, Shudt M, George KS (2016) Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol 79: 44-50. doi: 10.1016/j.jcv.2016.03.005. PMID: 27085509
9. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SLM, et al. (2015) Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. Cell 161: 1516-1526. doi: 10.1016/j.cell.2015.06.007. PMID: 26091036
10. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, et al. (2017) Zika virus evolution and spread in the Americas. Nature 546: 411-415. doi: 10.1038/nature22402. PMID: 28538734
11. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 366: 2267-2275. doi: 10.1056/NEJMoa1109910. PMID: 22693998
12. Halachev MR, Chan JZ, Constantinidou CI, Cumley N, Bradley C, et al. (2014) Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. Genome Med 6: 70. doi: 10.1186/s13073-014-0070-x. PMID: 25414729
13. Davis RJ, Jensen SO, Van Hal S, Espedido B, Gordon A, et al. (2015) Whole Genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit. Infect Control Hosp Epidemiol 36: 1058-1064. doi: 10.1017/ice.2015.133. PMID: 26050101
14. Leong KWC, Cooley LA, Anderson TL, Gautam SS, McEwan B, et al. (2018) Emergence of Vancomycin-Resistant Enterococcus faecium at an Australian Hospital: A Whole Genome Sequencing Analysis. Sci Rep 8: 6274. doi: 10.1038/s41598-018-24614-6. PMID: 29674657
15. Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10: 21-8. doi: 10.1186/1746-4811-10-21. PMID: 25053969
16. Assad NA, Balmes J, Mehta S, Cheema U, Sood A (2015) Chronic obstructive pulmonary disease secondary to household air pollution. Semin Respir Crit Care Med 36: 408-421. doi: 10.1055/s-0035-1554846. PMID: 26024348
17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649. doi: 10.1093/bioinformatics/bts199. PMID: 22543367

18. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, et al. (2015) TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLoS One 10: 2015-10. doi: 10.1371/journal.pone.0142951. PMID: 26565975

19. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, et al. (2015) PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol 53: 1908-1914. doi: 10.1128/JCM.00025-15. PMID: 25854485

20. Thomsen MCF, Ahrenfeldt J, Cisneros JLB, Jurtz V, Larsen MV, et al. (2016) A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance. PLoS One 11: e0157718. doi: 10.1371/journal.pone.0157718. PMID: 27327771

21. Kleinheinz KA, Joensen KG, Larsen MV (2014) Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4: e27943. doi: 10.4161/bact.27943. PMID: 24575358

22. Babraham Bioinformatics (2010) FastQC: a quality control tool for high throughput sequence data. [cited October 4, 2018]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

in WSE that competitively inhibit the binding of NTHi to PAFR. Second, the utilization of other potential epithelial receptors besides PAFR in CDSE-induced NTHi adhesion (Avadhanula et al. 2006; Fink et al. 2002; Klaile et al. 2013; Novotny and Bakaletz 2016). PAFR has been reported to be the major receptor for NTHi binding via lipooligosaccharide glycoform-containing phosphorylcholine (Shukla et al. 2016a, b; Shukla et al. 2015; Swords et al. 2000). The temporal upregulation of PAFR expression in airway cells upon exposure to biomass smoke, air pollutants as well as cigarette smoke (KC et al. 2018a, b; Miyashita et al. 2017; Shukla et al. 2016a, b) appears to increase the susceptibility to NTHi infection (Shukla et al. 2015). Besides PAFR, NTHi has been reported to utilize the host extracellular matrix (ECM) proteins; fibronectin, laminin and collagen IV for colonization in the respiratory tract via the Hap adhesin (Fink et al. 2002). In addition, the NTHi Type IV pilus and the outer membrane protein P1 interacts with the intercellular adhesion molecule-1 (ICAM-1) and the carcinoembryonic antigen-related cell adhesion molecules (CEACAMs), respectively, to adhere to respiratory tract epithelial cells (Avadhanula et al. 2006; Klaile et al. 2013; Novotny and Bakaletz 2016). Thus, further work is needed to investigate the effect of CDSE and WSE on the expression of ECM proteins, ICAM-1 or CEACAMs, and how they relate to adherence of NTHi to the airway epithelium. While a number of studies have reported differential effects of wood and cigarette smoke exposure in terms of COPD phenotypes (Camp et al. 2014; Fernandes et al. 2017; Rivera et al. 2008), investigations are also needed to compare the effects of cow dung and cigarette smoke on the manifestation of COPD.

In terms of treatment of respiratory bacterial infections, the major challenges to anti-microbial therapy are the acquisition and spread of antibiotic resistance and the limited number of new antibiotics being developed and clinically approved for human use (Ventola 2015). Therefore, an alternative approach is needed to prevent and treat bacterial airway infections through interfering with bacterial/host tissue interfaces. A number of PAFR antagonists, including WEB-2086, Ginkgolide-B, CV-3988, PCA-4248, and CAS-99103–16–9 have been reported to block the attachment of bacterial pathogens to respiratory epithelium (Barbier et al. 2008; Grigg et al. 2012; Hergott et al. 2015; Negro Alvarez et al. 1997; Shukla et al. 2016a, b). Here, WEB-2086 and its structural analogue, C17, were found to reduce the adhesion of NTHi to CSE- and CDSE-exposed human bronchial epithelial cells in a dose-dependent manner. In addition, they both inhibited biomass and cigarette smoke-induced expression of pro-inflammatory cytokines IL-6 and IL-8, which are known features of COPD.

A limitation of this study is the use of a commercial airway epithelial cell line. Further work is now needed to evaluate the effects of biomass smoke exposure on PAFR expression and adherence of bacterial pathogens in vivo. Nevertheless, the results of the work reveal that biomass smoke exposure increases the adherence of a bacterial pathogen to bronchial epithelial cells. Furthermore, through the use of antagonists, we identified PAFR as a potential therapeutic target for alleviating the impact of biomass smoke exposure and reducing the risk of NTHi exacerbations. Translation of these findings would contribute to the development of new management strategies for biomass smoke-related obstructive airway disease.

Recommendation

This study has highlighted the need for global immediate action towards reducing the exposure to and impact of smoke generated from the combustion of biomass fuel. National policies, and community awareness and education on preventive interventions, including the use of efficient cooking stoves and cleaner fuels such as biogas, liquefied petroleum gas and electricity are necessary to substantially reduce the number of deaths and illnesses from the hazardous effects of biomass smoke exposure. Furthermore, therapeutic interventions, including the use of anti-oxidants, anti-inflammatory and anti-microbial drugs may be promising in limiting the effects of biomass exposure-related respiratory diseases, including COPD. This will thus, significantly contribute to achieving the primary goal target of the Sustainable Development Goal 3: Ensure healthy lives and promote well-being for all at all ages.

Acknowledgements The authors acknowledge the technical support of Laboratory Manager, David Steele, at the Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania. They are also grateful to Conall O'Toole for assistance with the collection of wood and cow dung samples.

Funding RKC is a recipient of a Health Tasmania Graduate Research Scholarship through the University of Tasmania. PMH is funded by a Fellowship from the NHMRC (1079187).

Compliance with Ethical Standards

Conflict of interest The authors report no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors. The study was approved by the Tasmanian Health and Medical Human Research Ethics Committee (EC00337).

References

Amaral AFS et al (2018) Airflow obstruction and use of solid fuels for cooking or heating: BOLD results. Am J Respir Crit Care Med 197:595–610. https://doi.org/10.1164/rccm.201701-0205OC

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of an Isolate of Nontypeable *Haemophilus influenzae* from an Acute Exacerbation of Chronic Obstructive Pulmonary Disease in Tasmania

Rajendra KC,[a] Kelvin W. C. Leong,[b] Belinda McEwan,[c] Julia Lachowicz,[d] Nicholas M. Harkness,[d] Steve Petrovski,[e] Gunasegaran Karupiah,[a] Ronan F. O'Toole[a,b,f]

[a]Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia
[b]Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health, and Engineering, La Trobe University, Melbourne, Victoria, Australia
[c]Department of Microbiology and Infectious Diseases, Royal Hobart Hospital, Hobart, Tasmania, Australia
[d]Department of Respiratory and Sleep Medicine, Royal Hobart Hospital, Hobart, Tasmania, Australia
[e]Department of Physiology, Anatomy and Microbiology, School of Life Sciences, La Trobe University, Melbourne, Victoria, Australia
[f]Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin, Ireland

ABSTRACT Nontypeable *Haemophilus influenzae* (NTHi) is an important cause of human illness, including pneumonia and acute exacerbations of chronic obstructive pulmonary disease (COPD). We report here the draft genome of an isolate of NTHi collected from the sputum of a patient presenting with COPD in Tasmania, Australia.

Chronic obstructive pulmonary disease (COPD) is a serious, progressive condition characterized by a persistent reduction in lung airflow (1). It has emerged as the third leading cause of mortality, claiming more than 3 million lives worldwide in 2016 (2). In Australia, it is estimated that COPD affects 1.45 million people (3). COPD is an important disease in the state of Tasmania, where higher rates of smoking are observed with respect to the national rate (4). Nontypeable *Haemophilus influenzae* (NTHi) is a key pathogen that colonizes damaged airways in COPD patients and causes acute exacerbations that contribute to morbidity and mortality (5–7). Here, we present the draft assembled genome sequence of an NTHi strain isolated from a case of COPD in Tasmania.

NTHi strain RHH-38 was isolated in 2018 at the Royal Hobart Hospital in Tasmania from the sputum of a COPD patient presenting with an acute exacerbation. The sputum specimen was homogenized and cultured on chocolate blood agar plates at 35°C in a CO_2 atmosphere followed by storage at 2 to 8°C. Bacterial identification was performed using a Bruker matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometer. A single colony of *Haemophilus influenzae* was suspended in 200 μl phosphate-buffered saline (PBS), and then genomic DNA was extracted using the DNeasy blood and tissue kit (catalog number 69504; Qiagen, USA). The genomic DNA preparation was further purified using the High Pure PCR template preparation kit (catalog number 11796828001; Roche, Germany). DNA library preparation was carried out using a Nextera XT DNA library preparation kit (catalog number FC-131-1024; Illumina, USA) as described previously (8, 9). Sequencing was performed using an Illumina MiSeq platform with 150-bp paired-end sequencing. In total, 1,161,034 paired-end reads were generated, representing an average read depth of 88.83-fold. Reads were trimmed of adapters using Trimmomatic (10), and *de novo* assembly of reads was performed with SPAdes v3.12.0 (11). All parameters were set to default except for the size of k-mers, which were manually set to 21, 33, 43, 53, 63, and 75. This resulted in the generation of

Citation KC R, Leong KWC, McEwan B, Lachowicz J, Harkness NM, Petrovski S, Karupiah G, O'Toole RF. 2020. Draft genome sequence of an isolate of nontypeable *Haemophilus influenzae* from an acute exacerbation of chronic obstructive pulmonary disease in Tasmania. Microbiol Resour Announc 9:e00375-20. https://doi.org/10.1128/MRA.00375-20.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Address correspondence to Ronan F. O'Toole, r.otoole@latrobe.edu.au.

Received 10 April 2020
Accepted 11 April 2020
Published 7 May 2020

a 1,914,787-bp draft genome consisting of 68 contigs (≥200 bp) that covered 86.7% of the reference *H. influenzae* 86-028NP genome (12). The N_{50} value was 66,703 bp, and the overall GC content was 38.1%. The genome assembly quality, including completeness with respect to the reference genome, was determined using the QUAST quality assessment tool (13).

In silico MLST analysis, performed by submission of the draft genome to the *H. influenzae* multilocus sequence typing (MLST) website (https://pubmlst.org/hinfluenzae/) (14), assigned RHH-38 to sequence type 422 (ST422) based on seven housekeeping genes. The draft genome was annotated using RASTtk (15–17), which identified a total of 2,019 genes consisting of 1,960 coding sequences and 59 RNA genes. Default parameters were used for all software unless otherwise specified.

In conclusion, this study presents the published genome sequence assembly of an NTHi isolate from a case of COPD in Tasmania. The application of genome sequencing has the potential to provide insights into recurrent exacerbations of COPD due to NTHi and the ability to distinguish between relapse and reinfection.

This work was conducted in accordance with ethics approval number H0016214 from the Tasmanian Health and Medical Human Research Ethics Committee.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JAAECN000000000. The version described in this paper is version JAAECN010000000. The associated BioProject and BioSample accession numbers are PRJNA603840 and SAMN13942196, respectively.

ACKNOWLEDGMENT

R. KC is the recipient of a Health Tasmania Graduate Research Scholarship from the University of Tasmania.

REFERENCES

1. Petsonk EL, Hnizdo E, Attfield M. 2007. Definition of COPD GOLD stage I. Thorax 62:1107–1108.
2. World Health Organization. 2017. Disease burden and mortality estimates: cause-specific mortality, 2000–2016. World Health Organization, Geneva, Switzerland. http://www.who.int/healthinfo/global_burden_disease/estimates/en/.
3. Toelle BG, Xuan W, Bird TE, Abramson MJ, Atkinson DN, Burton DL, James AL, Jenkins CR, Johns DP, Maguire GP, Musk AW, Walters EH, Wood-Baker R, Hunter ML, Graham BJ, Southwell PJ, Vollmer WM, Buist AS, Marks GB. 2013. Respiratory symptoms and illness in older Australians: the Burden of Obstructive Lung Disease (BOLD) study. Med J Aust 198:144–148. https://doi.org/10.5694/mja11.11640.
4. Australian Bureau of Statistics. 2019. Higher rates of chronic health conditions in Tasmania. Australian Bureau of Statistics. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.001~2017-18~Media%20Release~Higher%20rates%20of%20chronic%20health%20conditions%20in%20Tasmania%20(Media%20Release)~10015.
5. Sethi S, Murphy TF. 2001. Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review. Clin Microbiol Rev 14:336–363. https://doi.org/10.1128/CMR.14.2.336-363.2001.
6. Sethi S, Evans N, Grant BJ, Murphy TF. 2002. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med 347:465–471. https://doi.org/10.1056/NEJMoa012561.
7. Hillas G, Perlikos F, Tzanakis N. 2016. Acute exacerbation of COPD: is it the "stroke of the lungs"? Int J Chron Obstruct Pulmon Dis 11:1579–1586. https://doi.org/10.2147/COPD.S106160.
8. Gautam SS, Kc R, Leong KW, Mac Aogáin M, O'Toole RF. 2019. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. J Biol Methods 6:e110. https://doi.org/10.14440/jbm.2019.276.
9. Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M, O'Toole RF. 2018. Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS One 13:e0192351. https://doi.org/10.1371/journal.pone.0192351.
10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
12. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, Carson MB, Zhong H, Gipson J, Gipson M, Johnson LS, Lewis L, Bakaletz LO, Munson RS, Jr. 2005. Genomic sequence of an otitis media isolate of nontypeable Haemophilus influenzae: comparative study with H. influenzae serotype d, strain KW20. J Bacteriol 187:4627–4636. https://doi.org/10.1128/JB.187.13.4627-4636.2005.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
14. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.
15. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
16. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.
17. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

Avadhanula V, Rodriguez CA, Ulett GC, Bakaletz LO, Adderson EE (2006) Nontypeable *Haemophilus influenzae* adheres to intercellular adhesion molecule 1 (ICAM-1) on respiratory epithelial cells and upregulates ICAM-1 expression. Infect Immun 74:830–838. https://doi.org/10.1128/iai.74.2.830-838.2006

Bakaletz LO, Novotny LA (2018) Nontypeable *Haemophilus influenzae* (NTHi). Trends Microbiol 26:727–728. https://doi.org/10.1016/j.tim.2018.05.001

Balmes JR, Eisen EA (2018) Household air pollution and chronic obstructive pulmonary disease. "A Riddle, Wrapped in a Mystery, Inside an Enigma". Am J Respir Crit Care Med 197:547–549. https://doi.org/10.1164/rccm.201801-0033ED

Barbier M, Oliver A, Rao J, Hanna SL, Goldberg JB, Alberti S (2008) Novel phosphorylcholine-containing protein of *Pseudomonas aeruginosa* chronic infection isolates interacts with airway epithelial cells. J Infect Dis 197:465–473. https://doi.org/10.1086/525048

Behera D, Jindal SK (1991) Respiratory symptoms in Indian women using domestic cooking fuels. Chest 100:385–388. https://doi.org/10.1378/chest.100.2.385

Camp PG et al (2014) COPD phenotypes in biomass smoke- versus tobacco smoke-exposed Mexican women. Eur Respir J 43:725–734. https://doi.org/10.1183/09031936.00206112

Capistrano SJ, van Reyk D, Chen H, Oliver BG (2017) Evidence of biomass smoke exposure as a causative factor for the development of COPD. Toxics. https://doi.org/10.3390/toxics5040036

Castell JV et al (1989) Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. FEBS Lett 242:237–239

Crystal RG, Randell SH, Engelhardt JF, Voynow J, Sunday ME (2008) Airway epithelial cells: current concepts and challenges. Proc Am Thorac Soc 5:772–777. https://doi.org/10.1513/pats.200805-041HR

da Silva LF, Saldiva SR, Saldiva PH, Dolhnikoff M (2012) Impaired lung function in individuals chronically exposed to biomass combustion. Environ Res 112:111–117. https://doi.org/10.1016/j.envres.2011.10.012

Dutta A, Ray MR, Banerjee A (2012) Systemic inflammatory changes and increased oxidative stress in rural Indian women cooking with biomass fuels. Toxicol Appl Pharmacol 261:255–262. https://doi.org/10.1016/j.taap.2012.04.004

Dutta A, Roychoudhury S, Chowdhury S, Ray MR (2013) Changes in sputum cytology, airway inflammation and oxidative stress due to chronic inhalation of biomass smoke during cooking in premenopausal rural Indian women. Int J Hyg Environ Health 216:301–308. https://doi.org/10.1016/j.ijheh.2012.05.005

Ehrlich RI, White N, Norman R, Laubscher R, Steyn K, Lombard C, Bradshaw D (2004) Predictors of chronic bronchitis in South African adults. Int J Tuberc Lung Dis 8:369–376

Faizan MA, Thakur R (2019) Association between solid cooking fuels and respiratory disease across socio-demographic groups in India. J Health Pollut 9:190911. https://doi.org/10.5696/2156-9614-9.23.190911

Fernandes L et al (2017) Small airway imaging phenotypes in biomass- and tobacco smoke-exposed patients with COPD. ERJ Open Res. https://doi.org/10.1183/23120541.00124-2016

Fink DL, Green BA, Geme JWS III (2002) The *Haemophilus influenzae* Hap autotransporter binds to fibronectin, laminin, and collagen IV. Infect Immun 70:4902–4907

Gabay C (2006) Interleukin-6 and chronic inflammation. Arthritis Res Ther 8(Suppl 2):S3. https://doi.org/10.1186/ar1917

Geelen S, Bhattacharyya C, Tuomanen E (1993) The cell wall mediates pneumococcal attachment to and cytopathology in human endothelial cells. Infect Immun 61:1538–1543

Grigg J et al (2012) Cigarette smoke and platelet-activating factor receptor dependent adhesion of *Streptococcus pneumoniae* to lower airway cells. Thorax 67:908–913. https://doi.org/10.1136/thoraxjnl-2011-200835

Gruenert DC, Finkbeiner WE, Widdicombe JH (1995) Culture and transformation of human airway epithelial cells. Am J Physiol 268:L347–360. https://doi.org/10.1152/ajplung.1995.268.3.L347

Hergott CB, Roche AM, Naidu NA, Mesaros C, Blair IA, Weiser JN (2015) Bacterial exploitation of phosphorylcholine mimicry suppresses inflammation to promote airway infection. J Clin Invest 125:3878–3890. https://doi.org/10.1172/jci81888

Hu G, Zhou Y, Tian J, Yao W, Li J, Li B, Ran P (2010) Risk of COPD from exposure to biomass smoke: a metaanalysis. Chest 138:20–31. https://doi.org/10.1378/chest.08-2114

Kc R, Shukla SD, Gautam SS, Hansbro PM, O'Toole RF (2018a) The role of environmental exposure to non-cigarette smoke in lung disease. Clin Transl Med 7:39. https://doi.org/10.1186/s40169-018-0217-2

Kc R, Zosky GR, Shukla SD, O'Toole RF (2018b) A cost-effective technique for generating preservable biomass smoke extract and measuring its effect on cell receptor expression in human bronchial epithelial cells. Biol Methods Protoc 3:bpy010

Kimura A, Kishimoto T (2010) IL-6: regulator of Treg/Th17 balance. Eur J Immunol 40:1830–1835. https://doi.org/10.1002/eji.201040391

Klaile E et al (2013) Carcinoembryonic antigen (CEA)-related cell adhesion molecules are co-expressed in the human lung and their expression can be modulated in bronchial epithelial cells by non-typable *Haemophilus influenzae*, *Moraxella catarrhalis*, TLR3, and type I and II interferons. Respir Res 14:85. https://doi.org/10.1186/1465-9921-14-85

Kurmi OP, Semple S, Simkhada P, Smith WC, Ayres JG (2010) COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. Thorax 65:221–228. https://doi.org/10.1136/thx.2009.124644

Leung JM et al (2017) The role of acute and chronic respiratory colonization and infections in the pathogenesis of COPD. Respirology (Carlton, Vic) 22:634–650. https://doi.org/10.1111/resp.13032

Magitta NF et al (2018) Prevalence, risk factors and clinical correlates of COPD in a rural setting in Tanzania. Eur Respir J. https://doi.org/10.1183/13993003.00182-2017

McCarthy CE, Duffney PF, Gelein R, Thatcher TH, Elder A, Phipps RP, Sime PJ (2016) Dung biomass smoke activates inflammatory signaling pathways in human small airway epithelial cells. Am J Physiol Lung Cell Mol Physiol 311:L1222–l1233. https://doi.org/10.1152/ajplung.00183.2016

Miyashita L, Padovan M, Suri R, Whitehouse A, Grigg J (2017) Expression of nasal epithelial platelet activating factor receptor (PAFR) and in vivo exposure to air pollution. J Allergy Clin Immunol 139:AB264

Montes de Oca M, Zabert G, Moreno D, Laucho-Contreras ME, Lopez Varela MV, Surmont F (2017) Smoke, biomass exposure, and COPD risk in the primary care setting: the PUMA study. Respir Care 62:1058–1066. https://doi.org/10.4187/respcare.05440

Mukaida N (2000) Interleukin-8: an expanding universe beyond neutrophil chemotaxis and activation. Int J Hematol 72:391–398

Mukaida N (2003) Pathophysiological roles of interleukin-8/CXCL8 in pulmonary diseases. Am J Physiol Lung Cell Mol Physiol 284:L566–577. https://doi.org/10.1152/ajplung.00233.2002

Negro Alvarez JM, Miralles Lopez JC, Ortiz Martinez JL, Abellan Aleman A, Rubio del Barrio R (1997) Platelet-activating factor antagonists. Allergol Immunopathol 25:249–258

Novotny LA, Bakaletz LO (2016) Intercellular adhesion molecule 1 serves as a primary cognate receptor for the Type IV pilus of nontypeable *Haemophilus influenzae*. Cell Microbiol 18:1043–1055. https://doi.org/10.1111/cmi.12575

Oishi K, Sonoda F, Kobayashi S, Iwagaki A, Nagatake T, Matsushima K, Matsumoto K (1994) Role of interleukin-8 (IL-8)

and an inhibitory effect of erythromycin on IL-8 release in the airways of patients with chronic airway diseases. Infect Immun 62:4145–4152

Park E, Lee K (2003) Particulate exposure and size distribution from wood burning stoves in Costa Rica. Indoor Air 13:253–259. https://doi.org/10.1034/j.1600-0668.2003.00194.x

Po JY, FitzGerald JM, Carlsten C (2011) Respiratory disease associated with solid biomass fuel exposure in rural women and children: systematic review and meta-analysis. Thorax 66:232–239. https://doi.org/10.1136/thx.2010.147884

Pope D et al (2015) Exposure to household air pollution from wood combustion and association with respiratory symptoms and lung function in nonsmoking women: results from the RESPIRE trial, Guatemala. Environ Health Perspect 123:285–292. https://doi.org/10.1289/ehp.1408200

Rivera RM, Cosio MG, Ghezzo H, Salazar M, Perez-Padilla R (2008) Comparison of lung morphology in COPD secondary to cigarette and biomass smoke. Int J Tuberc Lung Dis 12:972–977

Romieu I, Riojas-Rodriguez H, Marron-Mares AT, Schilmann A, Perez-Padilla R, Masera O (2009) Improved biomass stove intervention in rural Mexico: impact on the respiratory health of women. Am J Respir Crit Care Med 180:649–656

Sana A, Somda SMA, Meda N, Bouland C (2018) Chronic obstructive pulmonary disease associated with biomass fuel use in women: a systematic review and meta-analysis. BMJ Open Respir Res 5:e000246. https://doi.org/10.1136/bmjresp-2017-000246

Schröder J (1989) The monocyte-derived neutrophil activating peptide (NAP/interleukin 8) stimulates human neutrophil arachidonate-5-lipoxygenase, but not the release of cellular arachidonate. J Exp Med 170:847–863

Sethi S, Evans N, Grant BJ, Murphy TF (2002) New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med 347:465–471

Shukla SD, Fairbairn RL, Gell DA, Latham RD, Sohal SS, Walters EH, O'Toole RF (2016a) An antagonist of the platelet-activating factor receptor inhibits adherence of both nontypeable *Haemophilus influenzae* and *Streptococcus pneumoniae* to cultured human bronchial epithelial cells exposed to cigarette smoke. Int J Chron Obstruct Pulmon Dis 11:1647–1655. https://doi.org/10.2147/copd.S108698

Shukla SD, Muller HK, Latham R, Sohal SS, Walters EH (2016b) Platelet-activating factor receptor (PAFr) is upregulated in small airways and alveoli of smokers and COPD patients. Respirology (Carlton, Vic) 21:504–510. https://doi.org/10.1111/resp.12709

Shukla SD, Sohal SS, O'Toole RF, Eapen MS, Walters EH (2015) Platelet activating factor receptor: gateway for bacterial chronic airway infection in chronic obstructive pulmonary disease and potential therapeutic target. Expert Rev Respir Med 9:473–485. https://doi.org/10.1586/17476348.2015.1070673

Siddharthan T et al (2018) Association between household air pollution exposure and chronic obstructive pulmonary disease outcomes in 13 low- and middle-income country settings. Am J Respir Crit Care Med 197:611–620. https://doi.org/10.1164/rccm.201709-1861OC

Smith M et al (2014) Prevalence and correlates of airflow obstruction in approximately 317,000 never-smokers in China. Eur Respir J 44:66–77. https://doi.org/10.1183/09031936.00152413

Sussan TE et al (2014) Source of biomass cooking fuel determines pulmonary response to household air pollution. Am J Respir Cell Mol Biol 50:538–548. https://doi.org/10.1165/rcmb.2013-0201OC

Swords WE et al (2000) Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. Mol Microbiol 37:13–27

Tamimi A, Serdarevic D, Hanania NA (2012) The effects of cigarette smoke on airway inflammation in asthma and COPD: therapeutic implications. Respir Med 106:319–328. https://doi.org/10.1016/j.rmed.2011.11.003

Ventola CL (2015) The antibiotic resistance crisis: part 1: causes and threats P & T:a peer-reviewed. J Formulary Manag 40:277–283

WHO (2018a) Household air pollution and health. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health. Accessed 06 Aug 2018

WHO (2018b) The top 10 causes of death. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 28 July 2018

Yang Y, Mao J, Ye Z, Li J, Zhao H, Liu Y (2017) Risk factors of chronic obstructive pulmonary disease among adults in Chinese mainland: a systematic review and meta-analysis. Respir Med 131:158–165. https://doi.org/10.1016/j.rmed.2017.08.018

Zhou Y et al (2014) Lung function and incidence of chronic obstructive pulmonary disease after improved cooking fuels and kitchen ventilation: a 9-year prospective cohort study. PLoS Med 11:e1001621. https://doi.org/10.1371/journal.pmed.1001621

111. Singh A, Tuladhar B, Bajracharya K, Pillarisetti A (2012) Assessment of effectiveness of improved cook stoves in reducing indoor air pollution and improving health in Nepal. Energy Sustain Dev 16(4):406–414
112. Pilishvili T, Loo JD, Schrag S, Stanistreet D, Christensen B, Yip F et al (2016) Effectiveness of six improved cookstoves in reducing household air pollution and their acceptability in rural Western Kenya. PLoS ONE 11(11):e0165529
113. Zhou Y, Zou Y, Li X, Chen S, Zhao Z, He F et al (2014) Lung function and incidence of chronic obstructive pulmonary disease after improved cooking fuels and kitchen ventilation: a 9-year prospective cohort study. PLoS Med 11(3):e1001621
114. Sota CD, Lumbreras Martin J, Mazorra Aguiar J, Narros Sierra A, Fernández L, Borge García R. Effectiveness of improved cookstoves to reduce indoor air pollution in developing countries. The case of the cassamance natural subregion, western Africa. 2014
115. Oluwole O, Arinola GO, Ana GR, Wiskel T, Huo D, Olopade OI et al (2013) Relationship between household air pollution from biomass smoke exposure, and pulmonary dysfunction, oxidant-antioxidant imbalance and systemic inflammation in rural women and children in Nigeria. Glob J Health Sci 5(4):28–38
116. Rahman I (2012) Pharmacological antioxidant strategies as therapeutic interventions for COPD. Biochem Biophys Acta 1822(5):714–728
117. Mui TS, Man SP, Sin DD (2008) Developments in drugs for the treatment of chronic obstructive pulmonary disease. Expert Rev Clin Immunol 4(3):365–377
118. Grigg J, Walters H, Sohal SS, Wood-Baker R, Reid DW, Xu CB et al (2012) Cigarette smoke and platelet-activating factor receptor dependent adhesion of Streptococcus pneumoniae to lower airway cells. Thorax 67(10):908–913
119. Shukla SD, Fairbairn RL, Gell DA, Latham RD, Sohal SS, Walters EH et al (2016) An antagonist of the platelet-activating factor receptor inhibits adherence of both nontypeable *Haemophilus influenzae* and *Streptococcus pneumoniae* to cultured human bronchial epithelial cells exposed to cigarette smoke. Int J Chron Obstr Pulm Dis 11:1647–1655
120. Staunton DE, Merluzzi VJ, Rothlein R, Barton R, Marlin SD, Springer TA (1989) A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56(5):849–853
121. Avadhanula V, Rodriguez CA, Ulett GC, Bakaletz LO, Adderson EE (2006) Nontypeable Haemophilus influenzae adheres to intercellular adhesion molecule 1 (ICAM-1) on respiratory epithelial cells and upregulates ICAM-1 expression. Infect Immun 74(2):830–838
122. Heads of State commit to lead response to beat noncommunicable diseases, promote mental health. Geneva: World Health Organisation; 2018. http://www.who.int/news-room/detail/27-09-2018-heads-of-state-commit-to-lead-response-to-beat-noncommunicable-diseases-promote-mental-health. Accessed 27 Sept 2018
123. Dabadge A (2018) What has the Pradhan Mantri Ujjwala Yojana achieved so far?

MICROBIAL GENOMICS

RESEARCH ARTICLE
KC *et al.*, *Microbial Genomics*
DOI 10.1099/mgen.0.000405

MICROBIOLOGY SOCIETY
OPEN DATA OPEN MICROBIOLOGY

Whole-genome analyses reveal gene content differences between nontypeable *Haemophilus influenzae* isolates from chronic obstructive pulmonary disease compared to other clinical phenotypes

Rajendra KC[1], Kelvin W.C. Leong[2], Nicholas M. Harkness[3], Julia Lachowicz[3], Sanjay S. Gautam[1], Louise A. Cooley[4], Belinda McEwan[4], Steve Petrovski[5], Gunasegaran Karupiah[1] and Ronan F. O'Toole[2,*]

Abstract

Nontypeable *Haemophilus influenzae* (NTHi) colonizes human upper respiratory airways and plays a key role in the course and pathogenesis of acute exacerbations of chronic obstructive pulmonary disease (COPD). Currently, it is not possible to distinguish COPD isolates of NTHi from other clinical isolates of NTHi using conventional genotyping methods. Here, we analysed the core and accessory genome of 568 NTHi isolates, including 40 newly sequenced isolates, to look for genetic distinctions between NTHi isolates from COPD with respect to other illnesses, including otitis media, meningitis and pneumonia. Phylogenies based on polymorphic sites in the core-genome did not show discrimination between NTHi strains collected from different clinical phenotypes. However, pan-genome-wide association studies identified 79 unique NTHi accessory genes that were significantly associated with COPD. Furthermore, many of the COPD-related NTHi genes have known or predicted roles in virulence, transmembrane transport of metal ions and nutrients, cellular respiration and maintenance of redox homeostasis. This indicates that specific genes may be required by NTHi for its survival or virulence in the COPD lung. These results advance our understanding of the pathogenesis of NTHi infection in COPD lungs.

DATA SUMMARY

Sequence read files for all 40 isolates sequenced in this work have been deposited in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and are accessible through NCBI BioSample and SRA accession numbers SAMN13942196 and SRS6166046, respectively.

Two supplementary figures, five datasets and a fasta file of non-redundant pan-gene sequences are available on Figshare (https://doi.org/10.6084/m9.figshare.12545957.v1).

Dataset 1. This dataset contains information on the clinical source and geographical origin of the 568 isolates of nontypeable *Haemophilus influenza* (NTHi) included in this study. The dataset also contains the clade number and multilocus sequence typing allelic profile of each isolate. The 40 NTHi isolates newly sequenced for this study are boxed for easy identification.

Dataset 2. This dataset contains data on pan genes as identified by Roary from the analysis of 568 NTHi genomes. The dataset provides information regarding gene annotation and

Received 18 March 2020; Accepted 23 June 2020; Published 24 July 2020
Author affiliations: [1]Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Tasmania, Australia; [2]Department of Pharmacy and Biomedical Sciences, School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria, Australia; [3]Department of Respiratory and Sleep Medicine, Royal Hobart Hospital, Tasmania, Australia; [4]Department of Microbiology and Infectious Diseases, Royal Hobart Hospital, Tasmania, Australia; [5]Department of Physiology, Anatomy and Microbiology, School of Life Sciences, La Trobe University, Victoria, Australia.
***Correspondence:** Ronan F. O'Toole, r.otoole@latrobe.edu.au
Keywords: chronic obstructive pulmonary disease; nontypeable *Haemophilus influenzae*; whole-genome sequencing; pan-genome-wide association studies.
Abbreviations: CC, clonal complex; cgSNPs, core-genome SNP; COPD, chronic obstructive pulmonary disease; DAPC, discriminant analysis of principal components; FDR, false discovery rate; GO, gene ontology; LOS, lipooligosaccharide; MLST, multilocus sequence typing; MST, minimum spanning tree; NTHi, nontypeable *Haemophilus influenzae*; pan-GWAS, pan-genome-wide association study; PC, principal component; SNP, single nucleotide polymorphism; ST, sequence type.
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

the pan gene presence and absence profile of the 568 NTHi isolates.

Dataset 3. This dataset lists the names of genes that were enriched in isolates of a particular clade. The genes enriched in each clade are provided in separate worksheets.

Dataset 4. This dataset lists the genes associated with chronic obstructive pulmonary disease (COPD) strains of NTHi. This first worksheet lists the 680 genes that were found to be over- and under-represented in COPD strains. The second worksheet lists the 226 genes that were over-represented (odds ratio ≥2) in COPD strains. The third worksheet lists the 145 genes that were significantly associated with COPD strains. The fourth worksheet lists the 122 COPD-associated genes that were found to be significant after 1000 random permutations of the phenotypic data. The fifth worksheet lists the 79 genes that were unique (variants of the same genes were excluded) and were significantly associated with the COPD strains of NTHi.

Dataset 5. This dataset contains information on the functional annotation and classification of 79 COPD-associated genes. The third and fourth columns of the dataset list the corresponding protein IDs of the genes after querying their translated sequence against the UniProt database. The sixth and seventh columns contain information on the functional classification of genes based on Gene Ontology (GO) molecular function and biological process, respectively. The eighth and ninth columns contain information on the functional annotation of genes based on Clusters of Orthologous Groups of proteins (COG) analysis.

Impact Statement

Chronic obstructive pulmonary disease (COPD) is emerging as the third leading cause of human mortalities worldwide. Nontypeable *Haemophilus influenzae* (NTHi) is a major pathogen causing acute exacerbations resulting in diminished quality of life, hospitalization and increased risk of death in COPD patients. We analysed the core and accessory genome of 568 NTHi isolates, including 40 newly sequenced isolates, to genotypically distinguish between NTHi from COPD and other clinical phenotypes. This genome-wide analysis identified accessory gene content differences between COPD and non-COPD strains. It highlighted a set of virulence and metabolic functions that may be differentially required by COPD strains. This knowledge is important for developing improvements in the management of NTHi infections that can cause acute exacerbations in COPD patients.

INTRODUCTION

While nontypeable *Haemophilus influenzae* (NTHi) is a common commensal of the human nasopharynx, this bacterium is also associated with a spectrum of diseases including otitis media and sinusitis, as well as hospital- and community-acquired pneumonia [1]. In addition, NTHi is the most common bacterial cause of chronic obstructive pulmonary disease (COPD) exacerbations [2–4]. NTHi has developed mechanisms to thrive in the hostile environment of different anatomical regions, such as the middle ear, upper and lower respiratory tracts, blood and the meninges [5]. Evolutionary and ecological forces drive bacteria to adapt and grow in different niches [6–9] by utilizing the basic nutrients available and resisting toxic products present in its environment [10]. This evolutionary adaptation typically involves two fundamental processes. The first is through mutations in genes, such as SNPs or nucleotide insertions/deletions, which can potentially alter the antigenicity of surface proteins or change the activity of enzymes and transport proteins [11]. A related mechanism is phase variation in which loci susceptible to hypermutation can undergo slipped-strand mispairing, due to changes in simple sequence repeats, which can rapidly modulate the expression level of genes [7, 12]. The second process is the acquisition of entirely new genetic sequences via horizontal gene transfer, which can undergo homologous or non-homologous recombination into the recipient genome [13]. During homologous recombination, a chromosomal fragment of a genome is replaced with a homologous sequence from another genome, whereas non-homologous recombination results in gain and loss of genetic material [14]. These processes can contribute to phenotypic changes including increased virulence and antibiotic resistance, and adaptations to the host microenvironments such as immune evasion and greater metabolic capacity [7, 8, 15–19].

Conventional typing methods such as multilocus sequence typing (MLST) cannot distinguish between commensal and pathogenic strains of NTHi [20–22]. Furthermore, a previous study by De Chiara and colleagues reported that phylogeny provides insufficient resolution to discriminate between strains isolated from different clinical sources based on an analysis of 97 NTHi isolates [21]. Here, we hypothesized that NTHi associated with COPD may exhibit genetically encoded functional variances when compared to the isolates from non-COPD clinical illnesses. Therefore, we performed an analysis on a larger set of 568 NTHi genomes, which included 40 Australian isolates that were newly sequenced in this study. Our analyses involved the application of pan-genome-wide association studies (pan-GWASs) of genes to determine whether NTHi from COPD could be discriminated from isolates from other types of clinical disease.

METHODS

Bacterial strains collection, DNA extraction and genome sequencing

Forty NTHi isolates were retrieved from sputum samples of patients admitted to the Royal Hobart Hospital in Tasmania, Australia, from 2017 to 2018. Of these, 13 isolates were collected from COPD patients, with the remaining 27 collected from patients with other non-COPD disease presentations as shown in Dataset 1. DNA extraction and genome

sequencing were performed using the protocol that has been described in detail in our previous methods paper [23]. Briefly, for DNA preparation, pure cultures of NTHi isolates grown on chocolate agar were suspended in PBS. Genomic DNA was extracted using a DNeasy blood and tissue kit (Qiagen) and was treated with RNase for the removal of RNA. The genomic DNA was further purified using a High Pure PCR template preparation kit (Roche), as described previously [23]. The genomic DNA library of these isolates was then prepared using the Nextera XT library preparation kit (Illumina) and loaded into an Illumina MiSeq v2 (2×150 bp paired-end reads) cartridge for sequencing using an Illumina MiSeq platform at La Trobe University, Australia.

In addition to the 40 Australian NTHi isolates that were sequenced, 528 publicly available NTHi genomes [7, 21, 24–29] were also downloaded from the National Center for Biotechnology Information (NCBI) (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Haemophilus_influenzae/all_assembly_versions/) on September 10 2019 for analysis. The collection was composed of a heterogeneous group of isolates, in terms of geographical coverage. Based on the clinical source, the isolates were classified into two groups, COPD (*n*=373 isolates) and non-COPD (*n*=195 isolates).

De novo genome assembly, annotation and pan-genome analysis

Raw FASTQ files generated from the Illumina sequencer were uploaded to the Galaxy web platform [30], and the St Petersburg genome assembler (SPAdes) tool [31] was used for the *de novo* assembly of the sequence reads [30]. The default settings for all parameters were used, except for the size of k-mers, which were manually chosen as 21, 33, 43, 53, 63 and 75. The quality of genome assembly was evaluated using the Quality Assessment Tool (QUAST) [32]. Coverage of the reference genome was determined by aligning all of the sequenced genomes to the reference complete genome of the strain 86–028 NP [24]. The 40 isolates sequenced in this study had on average 120 contigs (>500 bp), with a mean genome size of 1906568 bp, reference genome coverage of 87.3% and a mean read depth of 118.8-fold.

The assembled FASTA/FNA files of the 40 Royal Hobart Hospital isolates and the 528 publicly available datasets were uploaded to the NeCTAR research cloud (http://cloud.nectar.org.au/) for subsequent analyses. We annotated relevant genomic features on the assembled/downloaded contigs using command line software tool Prokka version 1.12 with default parameters [33]. An *E* value threshold of 10^{-6} was used to determine the best match to known proteins in the databases, which included UniProt, Pfam and TIGRFAMs. If no matches were found, an ORF was labelled as a 'hypothetical protein'.

The annotated genome assembly outputs from Prokka (in GFF3 format) were aligned to build large-scale pan-genomes using a rapid large-scale prokaryote pan-genome analysis pipeline, Roary version 3.12 [34]. An additional argument '-e -mafft' was added to generate a multiFASTA alignment of core genes using MAFFT [35]. All other Roary parameters were used as default; minimum BLASTP identity of 90%; Markov clustering algorithm inflation value of 1.5. The genes identified within the genome alignment were classified with respect to their presence among the isolates: core (≥95%), shell (≥15 and<95%) and unique or cloud genes (<15%). The output files from Roary, *gene_presence_absence.csv* (Dataset 2) and *accessory_binary_genes.fa.newick*, were visualized using a Python script *roary_plots.py*, developed by Marco Galardini [36].

In silico MLST

The draft and complete genomes of 568 NTHi isolates were genotyped based on the allelic profile of seven housekeeping genes (*adk*, *atpG*, *frdB*, *fucK*, *mdh*, *pgi* and *recA*) hosted at https://pubmlst.org/hinfluenzae/ [37]. Each isolate was assigned with a sequence type (ST) using *in silico* MLST typing [38]. A minimum spanning tree (MST) was then generated based on the MLST profiles using the goeBurst algorithm [39] and the tree was visualized using Phyloviz [40].

Core-genome SNP (cgSNP) extraction and identification of genetic clusters

The core-genome alignment file was converted into a genlight object using the function *fasta2genlight* in the R package *adegenet* (version 2.1.1, RStudio version 1.0.143) [41]. The cgSNPs were extracted and analysed to determine clusters of genetically related isolates using the multivariate analysis method called discriminant analysis of principal components (DAPC) [42]. The cgSNPs raw data were initially transformed using principal component (PC) analysis, followed by the identification of genetic clusters using the *k-means* clustering algorithm. *k-means* determines a given number of groups (clusters) such that sequences belonging to the same cluster are more similar to each other than to sequences in other clusters. This was achieved using the function *find.clusters* with 150 PCs retained, accounting for >90% of the sample variability. The optimal number of clusters was then inferred using the Bayesian information criterion as eight genetic clusters, which were then efficiently described using the *dapc* function with 60 PCs retained. An optimal number of PCs was chosen to avoid both extremes of a good model: underfitting and overfitting of the model. The trade-off between power of discrimination and overfitting can be measured by the α-score, which is the difference between reassignment probabilities for the true cluster and randomly permuted clusters [42]. We calculated the α-score for a range of retained PCs by implementing *a.score* and selected 60 as an optimal number of retained PCs at which the α-score was maximum at 0.7. The first three eigenvalues (discriminant functions) were then selected for visualization and interpretation.

In addition, we performed DAPC on the presence and absence profile of accessory genes to assess whether the composition of the accessory genome supports the partitioning of the collection into the identified clades. The same predefined clusters (clade I–VIII) identified by the core-genome-based *k-means* clustering were used to group the isolates. The accessory genome DAPC was carried out using the same

methodology used for the core-gene DAPC except that the gene presence and absence matrix file was used instead of the core-genome alignment file. We retained the first 60 PCs and 3 eigenvalues to examine the genetic clusters. The mean α-score for 60 retained PCs was found to be 0.68.

Phylogenetic analysis

cgSNPs were further utilized for performing phylogenetic analyses using MEGAX software [43]. Evolutionary genetic distances between the strains were computed using the maximum composite likelihood method [44]. The evolutionary relationship was inferred using the neighbour-joining method [45]. Reliability of tree topology was tested using a bootstrap interior-branch test [46]. All the branches were supported by bootstrap values >90%.

Identification of COPD-associated genes

The distribution of COPD strains in the subpopulation (clades) of isolates was assessed. We also evaluated the genetic distinction between COPD and non-COPD strains using DAPC on cgSNPs and the presence/absence profile of accessory genes. A pan-GWAS approach was applied to determine whether any genes in the accessory genome were linked to a particular patho-phenotype including COPD using Scoary [47]. Scoary was implemented in Python using *gene_presence_absence.csv* (output from Roary) and *trait.csv* for the genotype and phenotype input files, respectively. Each candidate gene in the pan-genome was scored according to its apparent correlation to the clinical phenotype using a 2×2 contingency table of the presence/absence profile of each gene for the clinical phenotype. A Fisher's exact test was performed on each gene in a population-agonistic manner. The Benjamini–Hochberg false discovery rate (FDR) adjustment was applied to correct for multiple comparisons [48]. The cut-off for a significant association was a P value lower than 0.05. Furthermore, for casual inference, a pairwise comparison algorithm was implemented that controls for spurious associations dependent on the population structure [49]. The phylogenetic tree calculated internally by Scoary from the Hamming distances in the genotype matrix was used for the pairwise comparisons [47]. The causal association was scored as significant when the P values for both the best and the worst pairings were lower than 0.05. Finally, an additional label-switching permutation was implemented by running pan-GWASs on randomly permutated phenotypic values between individuals for 1000 times and retaining the 5% quantile, referred to as an empirical P value [50]. A minimum allele frequency threshold of 5% was used so that genes present either in more than 28 isolates or in less than 540 isolates were included in the pan-GWAS analysis to avoid assigning too high importance to very rare genes/variants.

Functional annotation and classification of candidate COPD genes

The nucleotide sequences of candidate genes were translated to their corresponding peptide sequences using EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). The peptide sequences were then queried against the UniProt database using the basic local alignment search tool with an E threshold of 0.001 (https://www.uniprot.org/blast/). In order to better understand the underlying biological processes, we performed functional classification of the COPD-associated genes using the Gene List Analysis Tool that is accessed through the web-based PANTHER version 14 classification system (http://pantherdb.org/) [51]. The complete Gene Ontology (GO) annotation system, which consists of three datasets, was used for mapping the functions of the genes of interest [52]. We classified our genes based on the GO molecular function and biological process.

RESULTS

Global collection of 568 NTHi strains, including 40 newly sequenced Australian isolates

For the initial assessment of the genetic diversity of the collection of NTHi isolates, we determined their ST using *in silico* MLST allelic profile. Based on the MLST profile, the 568 NTHi isolates were assigned to 174 unique STs [37]. Of which, 70 STs were associated with COPD, 34 STs contained both COPD and non-COPD strains, and the remaining 70 STs included non-COPD strains only (Fig. 1b). Sixty-one STs contained more than two NTHi isolates, of which twenty-seven STs were found to be associated with COPD (Fig. 1b). Some of the COPD-associated STs that contained five or more NTHi isolates were ST12 (n=11), ST98 (n=6), ST196 (n=7), ST349 (n=5), ST485 (n=5), ST1025 (n=13) and ST1812 (n=5).

A MST overview of all the isolates was generated by Phyloviz [40] (Fig. 1) using the goeBurst full MST algorithm based on MLST profile [39]. The MST overview was overlaid with the isolation data based on the geographical (Fig. 1a) and clinical (Fig. 1b) sources of the isolates. The distribution of the NTHi collection was uniform over the entire MST, in terms of geographical and clinical isolation (Fig. 1). There was no clustering specific to a geographical area. The USA, Spanish, Portuguese and Australian isolates were present in all groups and scattered throughout the entire tree. In addition, COPD isolates were also dispersed throughout the MST tree. By applying the traditional definition of clonal complex (CC) (genotypes which have allelic profiles that differ from the ST genotype at no more than one of the seven MLST loci, i.e. have at least six out of seven identical MLST genes) [53], we found 119 different CCs, of which 74 included a single ST. The largest CC consisted of only five STs and contained both COPD and non-COPD strains. Nine CCs included three or more STs; none of which were found to be specific to the COPD phenotype. This indicates a weak or no association between MLST genotype and COPD.

Pan-genome analysis of 568 NTHi isolates

A pan-genome of 12249 genes was generated from the 568 draft NTHi genomes, including the 40 isolates newly sequenced in this study (Fig. 2). The nucleotide sequences of non-redundant pan genes, as identified by Roary from

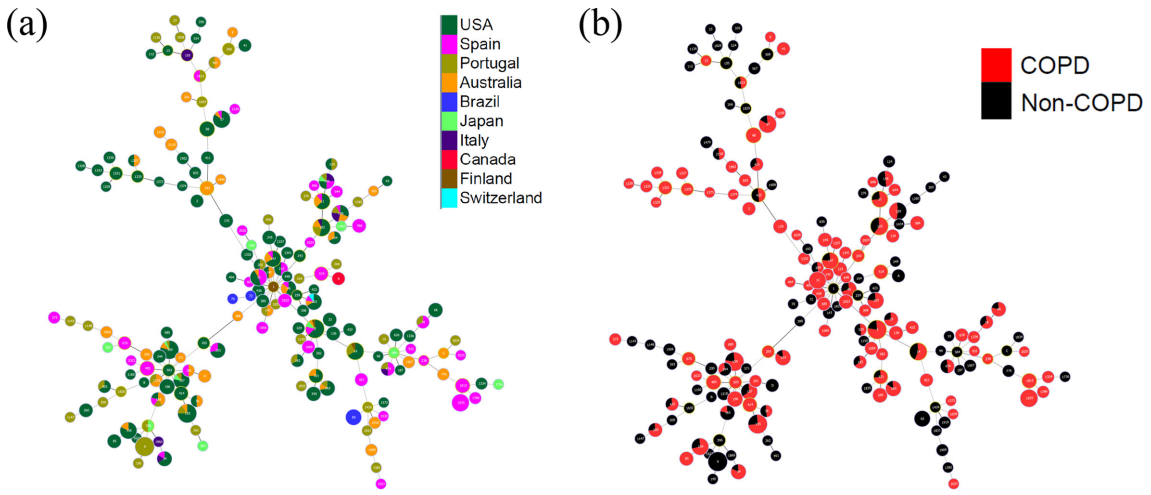


Fig. 1. MST overview of 568 NTHi isolates based on MLST, i.e. allelic profiles of seven housekeeping genes present in the PubMLST database. This was generated using the goeBurst full MST algorithm and was visualized with Phyloviz 2. Each node is a ST, and it is coloured according to the (a) geographical and (b) clinical sources of the isolates. The size of each node is proportional to the number of isolates. The larger STs containing more than 10 NTHi isolates are labelled in-text. There was no absolute separation of the strains according to geography. (a) Based on the MLST profiling, strains from the same STs were common to a wide range of geographical locations. (b) COPD-associated isolates were scattered over the MST indicating a weak or no association between MLST genotype and COPD. Of 174 unique STs in our collection, COPD isolates were found in 70 STs.

theanalysis of 568 NTHi genomes, is included in the Pangene_sequence file. The core-genome was represented by 853 genes that were present in at least 539 isolates. The core genes accounted for approximately 47% of the total number of genes (1821) present in the reference NTHi genome (86–028 NP) [24]. Whereas the core genome accounted for only 7% of the genes present in the pan-genome with the remaining 93% made up of the accessory pan-genome, which comprised the shell and cloud genes of the pan-genome (Fig. 2). The large accessory pan-genome confers diversity and high levels of genomic plasticity to NTHi. The majority of accessory pan genes (81% of pan genes) were rare and strain-specific, found in less than 15% of the NTHi collection. The distribution of genes in the population had a characteristic U-shape, as reported in previous studies in *H. influenzae* and other bacterial species [21, 54, 55].

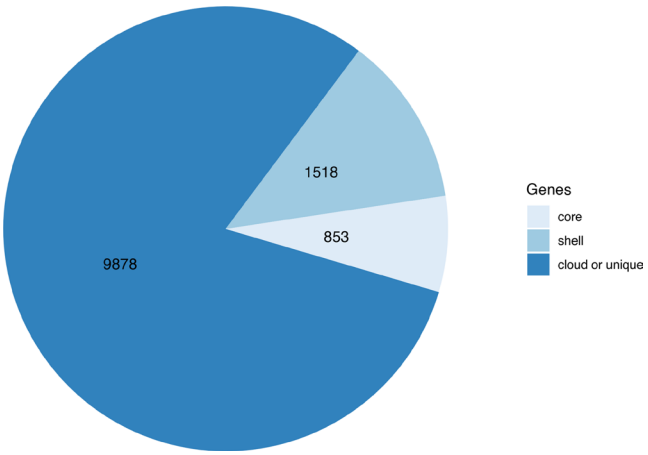


Fig. 2. Distribution of genes present in the pan-genome of the NTHi collection (*n*=568), which was constructed using Roary version 3.12. Of 12249 pan genes, the NTHi core-genome comprised 853 genes (present in at least 539 NTHi isolates). The remaining 11396 genes of the accessory genome were further classified into the shell (*n*=1518 genes, present in less than 539 and at least 85 NTHi isolates) and cloud or unique (*n*=9878 genes, present in less than 85 NTHi isolates). On average, 47% of each NTHi strain's gene set belonged to the core pan-genome. The remaining 53% of the strain's gene set belonged to the larger accessory pan-genome. This accessory pan-genome encompassing a large repertoire of genes confers diversity and high levels of plasticity to the NTHi genome.

Collection of 568 NTHi isolates exhibits a population structure defined by eight clades

The core-genome alignment generated by concatenation of individual core-gene alignments was used to infer the population structure. We extracted 97262 biallelic SNPs from the core-genome alignment of 696234 bp. To this dataset, we applied the DAPC to infer the number of clusters of genetically related isolates. Bayesian information criterion supported the partitioning of this collection of isolates into eight clusters or clades that were clearly separated from each other, except for a small overlap between clusters I and II (Fig. 3a).

We also tested whether the presence/absence profile of accessory genes supported the partitioning of the collection into the predefined eight clades. For this, we performed a DAPC on the dispensable genes using predefined grouping as identified by the cgSNPs DAPC. Isolates of clades III, IV, VII and VIII were clearly separated from each other, while isolates in clades I and II were found to be more closely related, which was

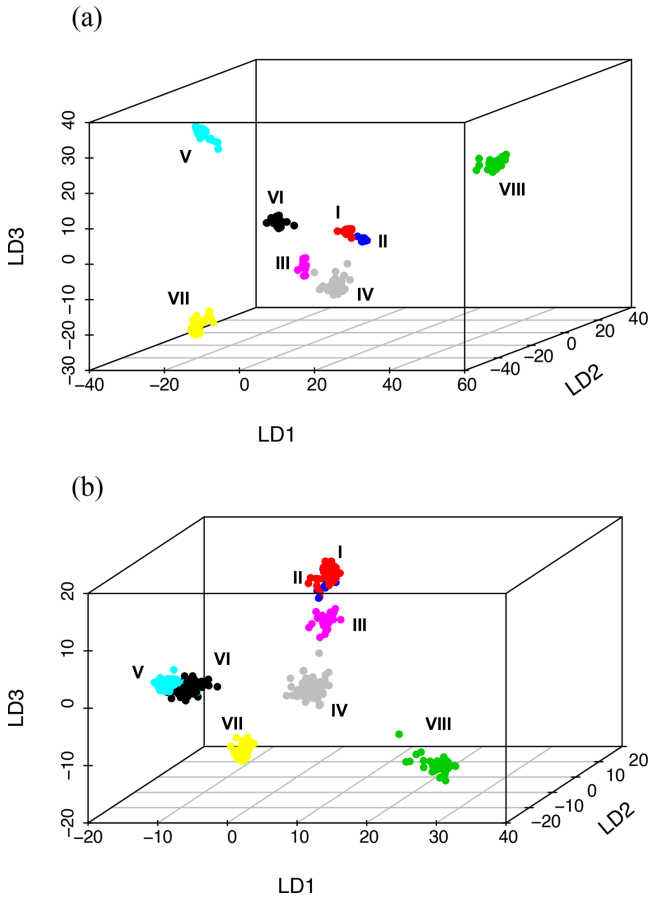


Fig. 3. (a) 3D scatterplot of the cgSNP-based DAPC of NTHi isolates (*n*=568). DAPC resolved the NTHi isolates into eight clusters (clades). Clades I and II were closely related to each other, whereas all other clusters were distinctly separated. (b) The accessory-genes-based DAPC correlated with the cgSNP-based DAPC with a distinct separation of clades III, IV, VIII and VIII, and a close relationship between clades I and II. The only discrepancy was with the isolates of clades V and VI, which were clearly separated on cgSNPs DAPC, whereas they overlapped on the accessory-gene-content DAPC. Each dot is an isolate, coloured according to the classification into one of the eight clusters/clades as assigned by the cgSNP-based DAPC.

consistent with the cgSNPs DAPC (Fig. 3a, b). This suggests that the isolates in clades I and II are evolutionarily related and have a common ancestor. The composition of accessory-genome DAPC further showed a close association between clades V and VI, which was not observed with the cgSNPs DAPC (Fig. 3a, b). This indicates that a set of accessory genes that are shared between isolates of clades V and VI could have been either inherited from a common ancestor or acquired through horizontal gene transfer.

Phylogeny separates isolates into groups that correlate with clades

We then compared the clade partitioning to standard phylogenetic analyses. We built a neighbour-joining phylogenetic tree using the cgSNPs (Fig. 4). Molecular phylogenetic analysis correlated perfectly with the clade partitioning as defined by the population genetics, supporting the clonal nature of the NTHi population characterized by eight distinct lineages. Isolates of clades I and II were evolutionarily related, as suggested by the cgSNPs and the accessory gene DAPC. Consistent with the accessory gene DAPC, phylogenetic analysis showed a close relationship between clades V and VI.

The clade partitioning, as defined by population genetics and evolutionary relationship based on cgSNPs, partially correlated with the general clustering from the MLST profile (Fig. 1) as shown in Fig. S1 (available with the online version of this article). Clades I, II and III, and clades V and VI clustered together in the MST (Fig. S1), which correlated with the phylogenetic tree topology (Fig. 4). However, isolates of clades IV, VII and VIII did not form a distinct cluster, and were uniformly dispersed over the MST (Fig. S1). Furthermore, core-genome SNPs-based phylogeny did not result in separation of the strains according to geographical location of sample isolation (Fig. S2).

Enrichment of genes in specific clades

Scoary, a pan-GWAS tool, was implemented to identify the genes that were enriched in a specific clade [47]. The FDR adjusted *P* value threshold of 0.001 was used, which identified 456, 167, 234, 532, 533, 599, 551 and 417 genes to be significantly overrepresented in clade I, II, III, IV, V, VI, VII and VIII, respectively (Dataset 3). Phylogenetically related clades shared a large proportion of clade-enriched genes. For example, clades I and II shared 29% (141 out of 489 genes), and clades V and VI shared 32.6% (278 out of 854 genes) of the genes enriched in the respective clades. Whereas, the evolutionarily distinct clades, such as clade VII and VIII, shared only 6.6% (60 out of 908 genes) of the genes over-represented in them.

Composition of the accessory genome but not the distribution of cgSNPs separates COPD from non-COPD strains

We then tested whether the 373 COPD strains could be associated with a particular clade(s). We overlaid the neighbour-joining phylogenetic tree with information on the clinical source of isolates as shown in Fig. 5. We found that COPD strains were distributed among all eight clades, with no absolute separation of the strains according to clinical source (Fig. 5). Expanding on our analysis, we next investigated whether the information on core-genome-wide polymorphic sites could be used to classify the collection into groups based on their clinical phenotypes. Discrimination analysis on cgSNPs showed significant overlap between COPD and non-COPD isolates, which is consistent with the previous results, suggesting poor resolution of cgSNPs for separating isolates according to their clinical source (Fig. 6a).

We then applied DAPC to the composition of the accessory genome, and found that, based on the first discriminant function, the presence/absence profile of dispensable genes could separate the COPD strains from the rest (Fig. 6b). The mean

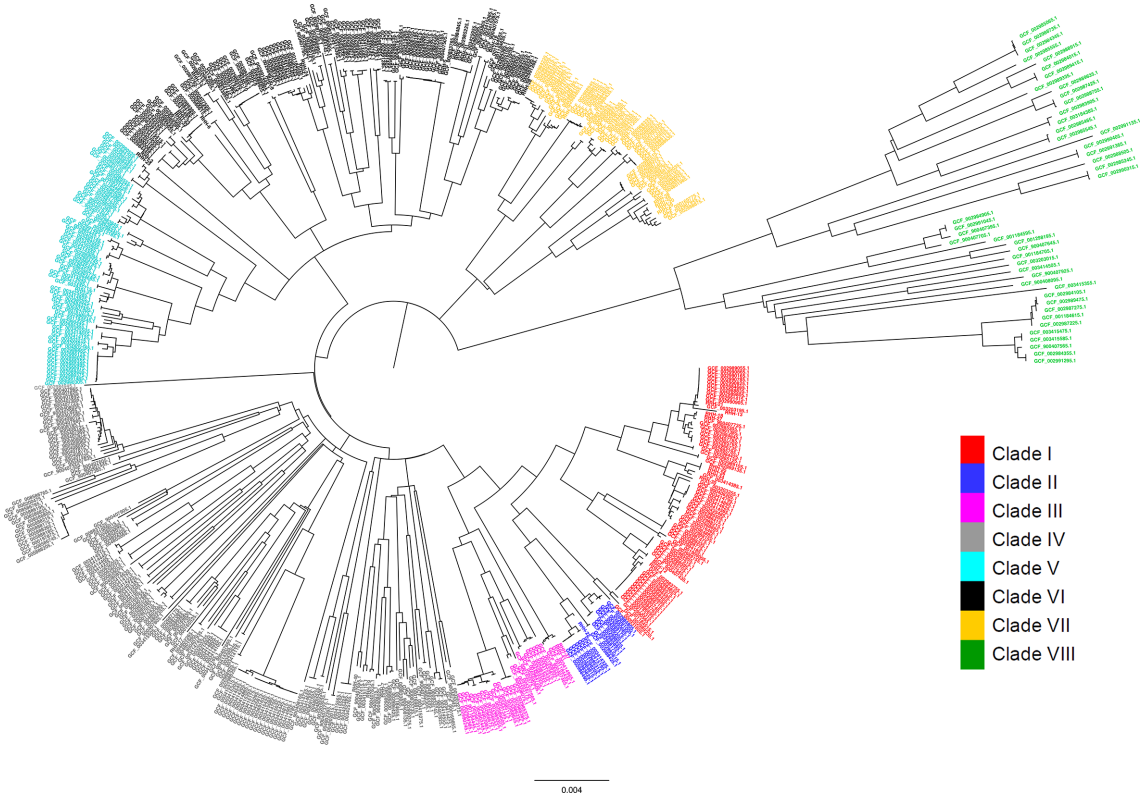


Fig. 4. Neighbour-joining phylogenetic tree of 568 NTHi isolates, based on the cgSNPs. The evolutionary distances were computed using the maximum composite likelihood method. There was a total of 664470 positions in the final dataset. Distinct sub-structuring of the NTHi population was evident with the cgSNP-based phylogeny. The phylogenetic analysis perfectly correlated with the DAPC-based classification that identified eight monophyletic clades. The close association between clades I and II, and between clades V and VI, as observed in accessory-genome-based DAPC, is consistent with the evolutionary relationship between them observed in the core-gene-sequence-based phylogeny obtained using the neighbour-joining maximum composite likelihood method. The tip labels are coloured according to the clades assigned by the cgSNP-based DAPC. Bar, number of base substitutions per site.

discriminant function separating the two populations was significantly different (P<0.0001, unpaired t-test with Welch's correction). This result indicates that there is a difference in the composition of the accessory genome in NTHi that could potentially separate COPD strains from those associated with other clinical illnesses.

Genome-wide association studies identified a set of accessory genes that are significantly associated with COPD strains

We determined the accessory genes in the bacterial dataset that were associated with a COPD phenotype. Scoary was used to screen the genes for association with COPD [47]. After the FDR correction, Scoary predicted 680 genes present in the NTHi pan-genome to be significantly enriched in NTHi collection from the COPD airways (Dataset 4). Out of the 680 genes, 226 genes were found to be significantly overrepresented in the COPD strains as compared to non-COPD strains (odds ratio ≥2, P<0.05) (Dataset 4).

Additionally, Scoary identified 145 (out of 680) genes to be significantly associated with the COPD phenotype (P<0.05) after pairwise comparison (Dataset 4). The other remaining 535 genes identified as significant prior to population-aware analysis were found to be lineage-specific effects upon inspection of the population distribution by the pairwise comparisons test. Furthermore, 1000 random permutations of the phenotype data were implemented, and the associated test statistic was calculated for each permutation. After the permutations, only 122 genes were found to have a significant association with the COPD phenotype (Dataset 4). Out of 122 significant hits, 86 hits were found to be different alleles (variants) of the same gene, one positively and one negatively associated with the COPD phenotype. The two alleles of these 43 genes were different enough to not be clustered as the same by Roary. Finally, there were 79 unique genes likely to play a role in COPD (Dataset 4).

Functional classification of candidate COPD-associated genes

We next predicted the biological functions of the 79 unique COPD candidate genes. A total of 70 out of 79 genes mapped to UniProt proteins with a minimum identity threshold of

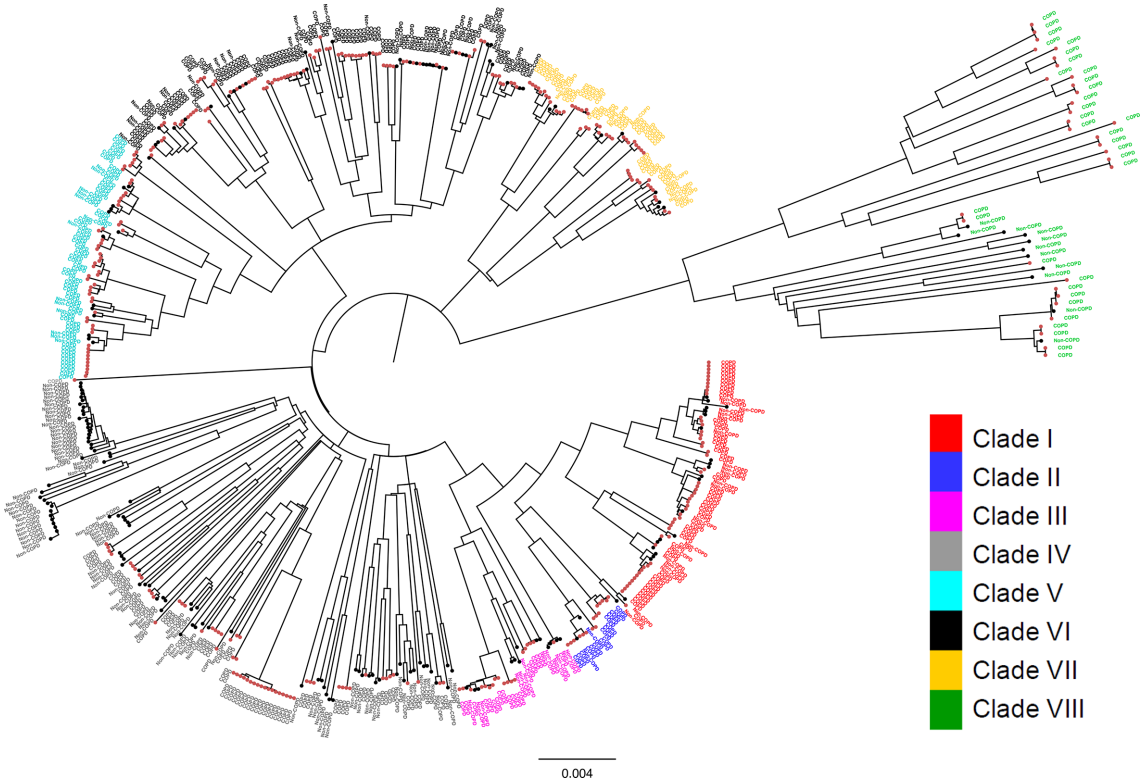


Fig. 5. Distribution of 568 NTHi isolates over the cgSNP-based neighbour-joining phylogenetic tree, which is annotated with the clinical source of isolation of the samples as COPD or non-COPD. Each evolutionary clade includes both COPD and non-COPD strains. Branch tips representing COPD strains are highlighted in red, whilst those representing non-COPD strains are shown in black. The tip labels are coloured according to the clades (I to VIII) as assigned by the cgSNP-based DAPC. Bar, number of base substitutions per site.

90% (Dataset 5). Furthermore, the candidate genes were functionally annotated by assigning their encoded proteins to Clusters of Orthologous Groups of proteins (COGs) categories [56]. The COG analysis, however, did not result in an increase in confident functional prediction for the candidate genes as compared to the original annotations, i.e. 70 genes were assigned to known orthologous groups when the identity threshold was maintained at 90% (Dataset 5).

Additionally, the functionally annotated genes that were associated with the COPD phenotype were classified based on their molecular functions and the biological processes they are involved in. Fifty-four and fifty-two of the seventy functionally annotated COPD-linked genes were assigned to GO molecular functions and biological processes, respectively (Dataset 5). A large number of these genes were found to be associated with transmembrane transporter activity (n=10), regulating the transport of inorganic phosphates, cations (Na^+ and K^+), metal ions such as copper and iron, lactate, dicarboxylate, carbohydrate, amino acids and proteins in and out of the bacterial cell (Table 1). Others were genes involved in cell redox homeostasis and cellular carbohydrate and protein metabolic processes, including the biosynthesis of aromatic and branched-chain amino acids. Consistent with previous studies [57, 58], a variant form of the IgA-peptidase-encoding gene has been found to be strongly associated with the COPD phenotype (odds ratio=4.4, P=0.0012). In addition, variants/alleles of genes encoding glycosyltransferases, such as *lex1* and *isgE*, and cytidylyltransferase-encoding *licC* that are involved in lipooligosaccharide (LOS) biosynthesis were also found to be associated with the COPD strains. With regard to other virulence genes such as *Haemophilus* adhesion and penetration protein (Hap), higher molecular weight proteins 1 and 2 (HMW1/2) and *H. influenzae* adhesin (Hia), they were found to be uniformly distributed among COPD and non-COPD strains.

H. influenzae is known to be naturally competent for transformation [59, 60]. *tfoX* (also called *sxy*) is a regulatory gene which is required for DNA uptake and transformation [61]. Its product, TfoX, interacts with cyclic-AMP receptor protein (CRP) and promotes the expression of genes of the competence regulon in *H. influenzae* [62, 63]. *H. influenzae* utilizes type IV pili for the transport of DNA across the membrane into the cytoplasm [64]. Variant forms of both *pilA* (encodes PilA, a major pilin subunit of type IV pili) [65] and *tfoX* are found to be associated with the COPD strains of NTHi (Table 1). RecJ is an exonuclease with 5′–3′ ssDNA-specific exonuclease activity that plays a crucial role in DNA repair and recombination pathways [66]. RecJ is associated with mismatch repair, and in addition, in conjunction with RecQ helicase, initiates recombination from a double-stranded

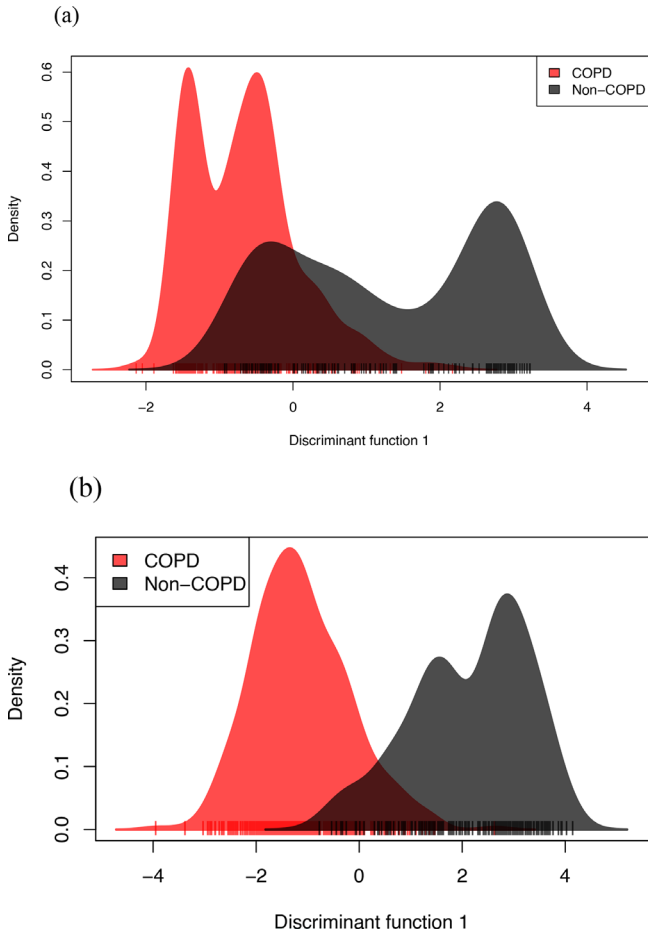


Fig. 6. DAPC of 568 NTHi genomes from COPD and other disease isolates. (a) The first discriminant function of the retained PCs based on cgSNPs leaves substantial overlap between COPD and non-COPD strains, 356 NTHi isolates (231 COPD and 125 non-COPD strains) were in the overlapping region. (b) DAPC on the presence/absence profile of accessory genes clearly provides a higher level of separation of COPD from non-COPD strains with 226 NTHi isolates (119 COPD and 107 non-COPD strains) in the overlapping region. Composition of accessory genes, but not the distribution of polymorphic sites in the core-genome, discriminates COPD strains from other clinical phenotype strains of NTHi. Each line is an isolate. COPD and non-COPD isolates are coloured in red and black, respectively.

break [67]. A variant of *recJ* has also been found to be associated with the COPD phenotype (odds ratio=2.5, P=0.019).

DISCUSSION

NTHi is associated with a wide range of diseases, including otitis media, meningitis and conjunctivitis, and is a major bacterial cause of exacerbations in COPD patients [3, 68–70]. The ability to predict a disease phenotype based on the genotype of a pathogen would be valuable in informing an appropriate prevention and treatment response. In some cases, phenotyping methods such as virulence factor profiling demonstrate clustering of the bacterial isolates into specific serotypes as in the case of *Streptococcus pneumoniae* [71]. While in others, MLST profiling correlates well with the disease induced by a pathogen, as in the case of *Enterococcus faecium*; for instance, a number of MLST STs, ST796, ST1421, ST233 and ST80, are associated with vancomycin resistance [72, 73]. In terms of infections with *H. influenzae*, capsulated strains have been reported to form serotype-specific clusters, in particular serotypes c, d, e and f that formed monophyletic clusters on a dendrogram reconstructed from the MLST allelic profile [21, 74]. However, the population of NTHi has been reported to be composed of a heterogenous group of isolates that form highly divergent clusters based on MLST [21]. Among the diseases caused by NTHi, Brazilian purpuric fever has been found to be caused by a well-defined NTHi clone (biogroup aegyptius) [75]; however, a correlation between genotypes of NTHi and COPD has not been established yet [21]. In this study, we analysed the whole-genome sequences of a large collection of NTHi strains that were isolated from different clinical sources to investigate a genetic basis of distinction between COPD and other phenotypic strains.

Our analyses indicate that conventional MLST typing exhibits low discriminatory power and is, thus, unsuitable for identifying COPD-specific clusters of NTHi. To increase the discriminatory power, we expanded the MLST scheme that comprises seven housekeeping genes and included 853 core genes in our analysis. We performed DAPC on cgSNPs, which grouped 568 NTHi isolates into distinct clades (Fig. 3a). Firstly, using this larger panel of genomes, the NTHi isolates resolved into eight clades compared to six clades in the previous analysis by De Chiara and colleagues in which 97 NTHi isolates were used [21]. NTHi isolates that fell within De Chiara's clade I and V further separated into two distinct clusters each; whereas, the other clades (II, III, IV and VI) did not show substructure with the analysis of 568 NTHi isolates. Each of the eight clades contained NTHi strains from diverse disease phenotypes. Moreover, phylogeny derived from an analysis of cgSNPs did not differentiate NTHi isolates based on their clinical source which is consistent with the study conducted by De Chiara *et al.* and a later study by Pettigrew and colleagues that analysed a collection of 403 NTHi genome sequences [7, 21]. The finding that NTHi strains with highly similar core-genome sequences can cause a wide range of diseases suggests that non-core accessory genes may to a large extent be responsible for the different disease phenotypes that result from NTHi infection, as has been observed in other pathogenic bacteria such as *Clostridium difficile* [76].

The eight distinct NTHi clades supported by population genetics correlated perfectly with molecular phylogenetic analysis based on the cgSNPs (Figs 3a and 4). Analysis of the composition of the accessory genome further correlated with phylogeny (Figs 3b and 4). This consistency highlights the clonal nature of the NTHi population. We also identified the genes that were enriched (over-represented) in each clade (Dataset 3) and found a set of clade-enriched genes distributed over the evolutionarily distant clades. For example, out of 456, 551 and 417 genes that were specifically overrepresented in clades I, VII and VIII, respectively, (FDR adjusted P value <0.001), 11 genes were found to be common in these

Table 1. List of virulence and transformation competency associated genes that significantly correlated with COPD strains of NTHi

Gene	Gene name	OR	*P*	Pairwise *P*	Emp *P*	Function
igA	IgA1 protease autotransporter	4.5	0.0012	1.4×10^{-7}–1.53×10^{-5}	0.0009	Virulence
isgE	*N*-Acetyl-glucosamine-transferase	5.5	0.0008	1.4×10^{-6}–6.6×10^{-5}	0.0009	LOS synthesis
lex1	LOS biosynthesis protein lex-1	2.8	0.0274	3.6×10^{-5}–2.27×10^{-2}	0.0009	LOS synthesis
licC	2-C-Methyl-D-erythritol 4-phosphate cytidylyltransferase	6.4	4.9×10^{-9}	4.0×10^{-8}–2.3×10^{-7}	0.0019	LOS synthesis
pilA	Type IV pilin subunit protein PilA	5.1	0.0002	3.1×10^{-6}–3.1×10^{-6}	0.003	Adhesion; transformation
tfoX	DNA transformation protein	8.6	0.0003	3.7×10^{-8}–4.2×10^{-7}	0.001	Transformation

OR, odds ratio; *P*, false rate discovery adjusted *P* value; Pairwise *P*, range of *P* values from the pairwise comparisons; Emp *P*, empirical *P* values after 1000 permutations.

phylogenetically distinct clusters. This underscores the role of horizontal gene transfer prevalent in the NTHi population that accounts for genetic diversity among the species [77].

We then applied discriminant analysis on the composition of accessory genes, which demonstrated a clear separation between strains associated with COPD and other clinical illnesses. Moreover, using a pan-GWAS approach, we identified a subset of NTHi accessory genes associated with the COPD phenotype. Some key COPD-associated genes likely to be involved in pathogenesis are listed in Table 1. NTHi utilizes cells surface structures to interact with host cells and pave the way for colonization and invasion [78]. We identified a variant form of *pilA*, encoding type IV pilin subunit protein, and LOS biosynthetic genes (*licC, lex1* and *isgE*) as being associated with the COPD strains. PilA and LOS have been demonstrated to play an important role in biofilm formation and colonization of the respiratory tract [79, 80]. Evasion of host immune defence is another probable mechanism by which NTHi strains thrive in the COPD airways. We found an association of the gene encoding IgA1 protease, which cleaves immunoglobulin A (IgA), with COPD isolates of NTHi (Table 1). IgA1 protease has previously been reported to be important to *H. influenzae* in the lower airways of COPD patients [81].

Pettigrew's and Molere's groups recently investigated large prospectively collected NTHi genomes to give insight into molecular changes during persistence in the COPD lung [7, 25]. They found genetic changes in multiple genes that regulate expression of virulence functions, such as adherence, nutrient uptake and immune evasion, which are likely to be involved in NTHi survival in the COPD lung. This suggests that in comparison to other ecological niches, such as the middle ear, sinuses, eye, meninges and the upper respiratory tract, NTHi in COPD airways are exposed to different microenvironments defined by distinct nutrient availability, pH, oxidizing potential and/or immune response. NTHi, therefore, exhibits genomic changes which appear to aid survival and adaptation in the hostile environment of COPD airways. Consistent with these findings, we found that COPD isolates of NTHi encoded different metabolic activities compared to strains associated with other clinical phenotypes. This suggests that metabolic capacity, in part, plays an important role in enabling NTHi to contribute to COPD pathogenesis and further supports the concept of nutritional virulence as an important determinant of pathogenic capability in NTHi [82].

In conclusion, our study indicates that the virulence and survival of NTHi in COPD is influenced by genes outside of the core-genome. The set of accessory genes associated with COPD strains may assist in successfully establishing a niche in COPD airways through acquisition of nutrients, evasion of the immune response, and enhancement of adhesion and colonization of airways. In addition, the presence of competence and recombination genes may enable NTHi strains to acquire genes that confer a competitive advantage in the COPD airways. Further work will examine how our finding that COPD strains of NTHi possess a distinct gene content could be translated into improvements in the management of NTHi infections in COPD.

Funding information
Funding for this work was provided from La Trobe University. R.K. was the recipient of a Health Tasmania Graduate Research Scholarship through the University of Tasmania.

Acknowledgements
We thank James Marthick for assisting in the laboratory during genome sequencing. This research was supported by use of the NeCTAR Research Cloud (https://nectar.org.au/research-cloud/), a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement

This work was conducted in accordance with Ethics Approval no. H0016214 from the Tasmanian Health and Medical Human Research Ethics Committee.

References

1. Agrawal A, Murphy TF. *Haemophilus influenzae* infections in the *H. influenzae* type B conjugate vaccine era. *J Clin Microbiol* 2011;49:3728–3732.
2. Moghaddam SJ, Ochoa CE, Sethi S, Dickey BF. Nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease and lung cancer. *Int J Chron Obstruct Pulmon Dis* 2011;6:113–123.
3. Murphy TF, Brauer AL, Schiffmacher AT, Sethi S. Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004;170:266–272.
4. Sriram KB, Cox AJ, Clancy RL, Slack MPE, Cripps AW. Nontypeable *Haemophilus influenzae* and chronic obstructive pulmonary disease: a review for clinicians. *Crit Rev Microbiol* 2018;44:125–142.
5. Wong SMS, Akerley BJ. Genome-scale approaches to identify genes essential for *Haemophilus influenzae* pathogenesis. *Front Cell Infect Microbiol* 2012;2:23.
6. Messer JS, Liechty ER, Vogel OA, Chang EB. Evolutionary and ecological forces that shape the bacterial communities of the human gut. *Mucosal Immunol* 2017;10:567–579.
7. Pettigrew MM, Ahearn CP, Gent JF, Kong Y, Gallo MC *et al*. *Haemophilus influenzae* genome evolution during persistence in the human airways in chronic obstructive pulmonary disease. *Proc Natl Acad Sci USA* 2018;115:E3256–E3265.
8. Harrison A, Hardison RL, Fullen AR, Wallace RM, Gordon DM *et al*. Continuous microevolution accelerates disease progression during sequential episodes of infection. *Cell Rep* 2020;30:2978–2988.
9. Elhenawy W, Tsai CN, Coombes BK. Host-specific adaptive diversification of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Cell Host Microbe* 2019;25:301–312.
10. Clementi CF, Murphy TF. Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front Cell Infect Microbiol* 2011;1:1.
11. LeClerc JE, Li B, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 1996;274:1208–1211.
12. Torres-Cruz J, van der Woude MW. Slipped-strand mispairing can function as a phase variation mechanism in *Escherichia coli*. *J Bacteriol* 2003;185:6990–6994.
13. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;405:299–304.
14. Didelot X, Méric G, Falush D, Darling AE. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 2012;13:256.
15. Wilson BA, Garud NR, Feder AF, Assaf ZJ, Pennings PS. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. *Mol Ecol* 2016;25:42–66.
16. Leimbach A, Hacker J, Dobrindt U. *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr Top Microbiol Immunol* 2013;358:3–32.
17. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK *et al*. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 2009;461:1243–1247.
18. Lieberman TD, Michel J-B, Aingaran M, Potter-Bynoe G, Roux D *et al*. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* 2011;43:1275–1280.
19. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 2016;14:150–162.
20. LaCross NC, Marrs CF, Gilsdorf JR. Population structure in nontypeable *Haemophilus influenzae*. *Infect Genet Evol* 2013;14:125–136.
21. De Chiara M, Hood D, Muzzi A, Pickard DJ, Perkins T *et al*. Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci USA* 2014;111:5439–5444.
22. Erwin AL, Sandstedt SA, Bonthuis PJ, Geelhood JL, Nelson KL *et al*. Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. *J Bacteriol* 2008;190:1473–1483.
23. Gautam SS, Kc R, Leong KW, Mac Aogáin M, O'Toole RF. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *J Biol Methods* 2019;6:e110.
24. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R *et al*. Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype D, strain KW20. *J Bacteriol* 2005;187:4627–4636.
25. Moleres J, Fernández-Calvet A, Ehrlich RL, Martí S, Pérez-Regidor L *et al*. Antagonistic pleiotropy in the bifunctional surface protein FadL (OmpP1) during adaptation of *Haemophilus influenzae* to chronic lung infection associated with chronic obstructive pulmonary disease. *mBio* 2018;9:e01176-18.
26. Kappler U, Dhouib R, Nair RP, McEwan AG. Draft genome sequences of three nontypeable strains of *Haemophilus influenzae*, C188, R535, and 1200, isolated from different types of disease. *Genome Announc* 2017;5:e00035-17.
27. Aziz A, Sarovich DS, Nosworthy E, Beissbarth J, Chang AB *et al*. Molecular signatures of non-typeable *Haemophilus influenzae* lung adaptation in pediatric chronic lung disease. *Front Microbiol* 2019;10:1622.
28. Atack JM, Murphy TF, Bakaletz LO, Seib KL, Jennings MP. Closed complete genome sequences of two nontypeable *Haemophilus influenzae* strains containing novel *modA* alleles from the sputum of patients with chronic obstructive pulmonary disease. *Microbiol Resour Announc* 2018;7:e00821-18.
29. Kc R, Leong KWC, McEwan B, Lachowicz J, Harkness NM *et al*. Draft genome sequence of an isolate of nontypeable *Haemophilus influenzae* from an acute exacerbation of chronic obstructive pulmonary disease in Tasmania. *Microbiol Resour Announc* 2020;9:e00375-20.
30. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D *et al*. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 2018;46:W537–W544.
31. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
32. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
33. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
34. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
35. Katoh K, Misawa K, Kuma K-ichi, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
36. Galardini M. Roary_plots Github; 2013. https://github.com/sanger-pathogens/Roary/tree/master/contrib/roary_plots
37. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
38. Seeman T. MLST Github; 2014. https://github.com/tseemann/mlst
39. Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 2009;10:152.
40. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M *et al*. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 2012;13:87.
41. Jombart T. *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;24:1403–1405.

42. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 2010;11:94.
43. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
44. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 2004;101:11030–11035.
45. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
46. Sitnikova T. Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol* 1996;13:605–611.
47. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 2016;17:238.
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57:289–300.
49. Maddison WP. Testing character correlation using pairwise comparisons on a phylogeny. *J Theor Biol* 2000;202:195–204.
50. North BV, Curtis D, Sham PC. A note on calculation of empirical P values from Monte Carlo procedure. *Am J Hum Genet* 2003;72:498–499.
51. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 2019;47:D419–D426.
52. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25:25–29.
53. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 2004;186:1518–1530.
54. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ *et al*. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 2010;11:R107.
55. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S *et al*. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 2009;5:e1000344.
56. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT *et al*. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 2001;29:22–28.
57. Fernaays MM, Lesse AJ, Sethi S, Cai X, Murphy TF. Differential genome contents of nontypeable *Haemophilus influenzae* strains from adults with chronic obstructive pulmonary disease. *Infect Immun* 2006;74:3366–3374.
58. Murphy TF, Lesse AJ, Kirkham C, Zhong H, Sethi S *et al*. A clonal group of nontypeable *Haemophilus influenzae* with two IgA proteases is adapted to infection in chronic obstructive pulmonary disease. *PLoS One* 2011;6:e25923.
59. Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* 1994;58:563–602.
60. Sisco KL, Smith HO. Sequence-specific DNA uptake in *Haemophilus* transformation. *Proc Natl Acad Sci USA* 1979;76:972–976.
61. Redfield RJ. sxy-1, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. *J Bacteriol* 1991;173:5612–5618.
62. Karudapuram S, Barcak GJ. The *Haemophilus influenzae* dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. *J Bacteriol* 1997;179:4815–4820.
63. Redfield RJ, Cameron ADS, Qian Q, Hinds J, Ali TR *et al*. A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. *J Mol Biol* 2005;347:735–747.
64. Chen I, Dubnau D. DNA uptake during bacterial transformation. *Nat Rev Microbiol* 2004;2:241–249.
65. Carruthers MD, Tracy EN, Dickson AC, Ganser KB, Munson RS *et al*. Biological roles of nontypeable *Haemophilus influenzae* type IV pilus proteins encoded by the pil and com operons. *J Bacteriol* 2012;194:1927–1933.
66. Sharma M, Majumdar PK. Occupational lifestyle diseases: an emerging issue. *Indian J Occup Environ Med* 2009;13:109–112.
67. Burdett V, Baitinger C, Viswanathan M, Lovett ST, Modrich P. *In vivo* requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci USA* 2001;98:6765–6770.
68. Thanavala Y, Lugade AA. Role of nontypeable *Haemophilus influenzae* in otitis media and chronic obstructive pulmonary disease. *Adv Otorhinolaryngol* 2011;72:170–175.
69. van Wessel K, Rodenburg GD, Veenhoven RH, Spanjaard L, van der Ende A *et al*. Nontypeable *Haemophilus influenzae* invasive disease in the Netherlands: a retrospective surveillance study 2001–2008. *Clin Infect Dis* 2011;53:e1–7.
70. Bodor FF, Marchant CD, Shurin PA, Barenkamp SJ. Bacterial etiology of conjunctivitis-otitis media syndrome. *Pediatrics* 1985;76:26–28.
71. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. *Lancet Infect Dis* 2005;5:83–93.
72. Leong KWC, Cooley LA, Anderson TL, Gautam SS, McEwan B *et al*. Emergence of vancomycin-resistant *Enterococcus faecium* at an Australian hospital: a whole genome sequencing analysis. *Sci Rep* 2018;8:6274.
73. Leong KWC, Kalukottege R, Cooley LA, Anderson TL, Wells A *et al*. State-wide genomic and epidemiological analyses of vancomycin-resistant *Enterococcus faecium* in Tasmania's public hospitals. *Front Microbiol* 2019;10:2940.
74. Meats E, Feil EJ, Stringer S, Cody AJ, Goldstein R *et al*. Characterization of encapsulated and noncapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol* 2003;41:1623–1636.
75. Brenner DJ, Mayer LW, Carlone GM, Harrison LH, Bibb WF *et al*. Biochemical, genetic, and epidemiologic characterization of *Haemophilus influenzae* biogroup aegyptius (*Haemophilus aegyptius*) strains associated with Brazilian purpuric fever. *J Clin Microbiol* 1988;26:1524–1534.
76. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y *et al*. Pathogenicity locus, core genome, and accessory gene contributions to *Clostridium difficile* virulence. *mBio* 2017;8:e00885-17.
77. Power PM, Bentley SD, Parkhill J, Moxon ER, Hood DW. Investigations into genome diversity of *Haemophilus influenza*e using whole genome sequencing of clinical isolates and laboratory transformants. *BMC Microbiol* 2012;12:273.
78. Foxwell AR, Kyd JM, Cripps AW. Nontypeable *Haemophilus influenzae*: pathogenesis and prevention. *Microbiol Mol Biol Rev* 1998;62:294–308.
79. Jurcisek JA, Bookwalter JE, Baker BD, Fernandez S, Novotny LA *et al*. The PilA protein of non-typeable *Haemophilus influenzae* plays a role in biofilm formation, adherence to epithelial cells and colonization of the mammalian upper respiratory tract. *Mol Microbiol* 2007;65:1288–1299.
80. Swords WE, Moore ML, Godzicki L, Bukofzer G, Mitten MJ *et al*. Sialylation of lipooligosaccharides promotes biofilm formation by nontypeable *Haemophilus influenzae*. *Infect Immun* 2004;72:106–113.
81. Lomholt H, van Alphen L, Kilian M. Antigenic variation of immunoglobulin A1 proteases among sequential isolates of *Haemophilus influenzae* from healthy children and patients with chronic obstructive pulmonary disease. *Infect Immun* 1993;61:4575–4581.
82. Schoen C, Kischkies L, Elias J, Ampattu BJ. Metabolism and virulence in *Neisseria meningitidis*. *Front Cell Infect Microbiol* 2014;4:114.