Genetic Characterisation of *Streptococcus pneumoniae* Serotype 1 Isolates in Relation to Invasiveness

Richard Manuel Harvey, B.Sc. (Biomedical Science) (Hons), AMusA

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy from the University of Adelaide

August 2010

Discipline of Microbiology and Immunology
School of Molecular and Biomedical Sciences
The University of Adelaide
Adelaide, S.A., Australia

Abstract

Streptococcus pneumoniae (the pneumococcus) is one of the most significant causes of human mortality and morbidity, and is a leading cause of diseases such as pneumonia, invasive disease (including bacteraemia and meningitis [IPD]) and otitis media. However, the pneumococcus is more commonly carried asymptomatically within the nasopharynx. The likelihood of the pneumococcus progressing from asymptomatic carriage to IPD varies between strains, and is associated with certain serotypes and clones. In particular, serotype 1 strains have a high-attack rate as they readily progress from a state of transient carriage to IPD. Recently, a closely-related group of hypervirulent serotype 1 clones have been responsible for epidemics of IPD with unusually high mortality rates. In contrast, epidemic asymptomatic carriage of serotype 1 clones has been found in a number of remote indigenous communities in the Northern Territory of Australia. Such isolates of serotype 1 from asymptomatic carriage are unusual and provided a rare opportunity to perform genomic comparisons with invasive serotype 1 isolates in order to identify serotype-independent factors that contribute to differences in the invasive potential of the pneumococcus.

Preliminary work using the non-invasive serotype 1 isolates from the Northern Territory and a collection of invasive human isolates of both indigenous and non-indigenous origin identified three virulence profiles that were non-invasive, intermediately virulent, or highly virulent in mice. Subsequently, phenomic analyses did not identify differences in the amount of capsule or differences in the apparent molecular weight or relative expression of a selection of well-characterised protein virulence factors that correlated with a virulence phenotype. However, in preliminary genomic comparisons the chromosomal toxin-antitoxin (TA) system of the PPI-1

variable region (PezAT) was identified in only highly virulent serotype 1 isolates, but absent from intermediately virulent and non-invasive serotype 1 isolates.

Therefore, the broad objectives of this study where to determine the clonal relatedness of isolates representing all three virulence phenotypes, characterise the potential role of the PPI-1 variable region in IPD and identify additional variable regions of the pneumococcal genome that were associated with heightened virulence.

Interestingly, it was shown that the highly virulent strain 1861 was a one-locus variant of the sequence type 217 clone of lineage B, responsible for severe IPD in parts of Africa. Therefore, the highly virulent nature of strain 1861 (and strain 4496) in mice is likely to also be reflected in humans. In contrast, the non-invasive and intermediately virulent strains were of lineage A, which includes the most frequently detected clones in Europe and the United States. In addition, different organisations of the PPI-1 variable region correlated with certain lineages of serotype 1. For example, the lineage A isolates lacked *pezAT* and instead contained a transcriptionally active immunity system against the bacteriocin, mersacidin. Interestingly, following a survey of a variety of *S. pneumoniae* strains representing a broad array of serotypes, the mersacidin immunity system was identified as the most common feature of the PPI-1 variable region, and is also present in the pandemic carriage Spanish^{23F} ST81 clone. In contrast, the highly virulent isolates of lineages B and C encoded *pezAT* and a number of genes predicted to encode enzymes that catalyse the rate-limiting steps of pathways involved in the degradation and biosynthesis of some amino acids and the biosynthesis and conversion of UDP-sugars. Interestingly, key components of this region exhibited preferential expression in the lungs and blood when compared to the nasopharynx of infected mice. Subsequently, it was shown using replacement mutants of the PPI-1 variable region in a D39 background that the region from the highly virulent strains promotes greater competitive fitness within the blood, lungs and nasopharyngeal tissue, compared to the

equivalent region from the intermediately virulent and non-invasive strains in co-infected mice. Whilst the mechanism by which the PPI-1 variable region contributes to survival *in vivo* is not clear, a possibility is that centralised regulation of a number of metabolic pathways may enhance the survival of the pneumococcus in the lungs and blood.

Whilst the PPI-1 variable region was important for the competitive fitness of D39 during disease, it was not clear whether this region was solely responsible for the differences observed in invasive potential between the highly virulent, intermediately virulent and non-invasive serotype 1 isolates. Therefore, comparative genomic hybridisation (CGH) and next generation genome sequencing were used to identify additional regions of the genome that are associated with the highly virulent isolates. It was found that genes homologous to the platelet-binding protein B (PblB) and a *Streptococcus mitis* lysogenic phage endolysin were present in the genome of only the highly virulent strains, and not in either the intermediately virulent and non-invasive strains. In addition, regions encoding a putative ABC transporter and enzymes predicted to be involved in the degradation of sialic acid, ZmpD, and a 64-kb Tn*5253*-like conjugative transposon that included a TA system that is highly homologous to *pezAT*, were found in only the highly virulent strains and not in the intermediately virulent or non-invasive isolates. Subsequent *in vivo* gene expression comparisons revealed that the phage-associated endolysin exhibited significantly greater expression in the lungs and blood of infected mice than the nasopharynx, which highlighted a potential mechanism for increased surface display of PblB in the lungs and blood. Whilst yet to be proven experimentally, it is thought that greater surface display of PblB could contribute to the rapid invasion of the blood that is characteristic of the highly virulent serotype 1 strains. In addition to PblB, greater expression of the sialic acid-associated ABC transporter was observed in the blood when compared to the lungs and nasopharynx of infected mice.

Therefore, whilst the role of the region remains to be determined, it might be possible that the region enables the utilisation of host-derived sialic acids as an energy source in the blood, thus promoting survival and growth.

However, a significant roadblock encountered in this study was the inability to genetically manipulate the highly virulent serotype 1 isolates. In order to confirm the importance of genes such as that in the PPI-1 variable region and *pblB* in virulence, mutagenesis of these regions was attempted. However, despite numerous attempts to optimise the transformation protocol, it is possible that some defect in the competence system that is linked to the over-expression of *comW* might be responsible for the inability to transform strains 1861 and 4496.

In this study a number of genomic regions were identified that via putative roles in metabolism, sugar acquisition and degradation and adherence to human platelets and their patterns of expression *in vivo* promote the invasion and survival of the pneumococcus in the blood and lungs. Such findings broaden the understanding of the progression to IPD from asymptomatic carriage and highlight strain-specific differences that could make some strains more virulent than others.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web via the University‟s digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

……...

Richard Manuel Harvey

Acknowledgements

Firstly I would like to thank my principal supervisor, Professor James Paton, for the time and effort that he has committed and the opportunities that he has provided to me over the time that I have been a student in his lab. Thank you also to my Co-supervisor, Dr. Uwe Stroeher, for his commitment both when he was in the lab and since he left. I am sure that both James and Uwe‟s guidance will have a lasting impact on my career. Thank you to Dr. David Ogyunniyi and Dr. Judy Morona for their ideas and help with various aspects of my project. Thank you to Dr. Lauren McAllister for helping me brainstorm ideas and for great company throughout the time of my project. I would also like to acknowledge other past and present members of the Paton lab, such as Dr. Adrienne Paton, Dr. Tony Focareta, Dr. Adam Potter, Dr. Claudia Trappetti, Dr. Sylvia Herold, James Byrne and Dr. Chris McDevitt for their assistance, which has made an important contribution to my project. Thank you also to Jan Cook and more recently Stephanie Philp for their effort in keeping the lab running smoothly. Thanks must especially go to Jan for helping to organise and perform the large-scale mouse experiments. I would also like to thank Dr. Heidi Smith-Vaughan and Dr. Amanda Leach from the Menzie‟s School of Health Research for providing indigenous isolates, and Andrew Lawrence for the isolates from the Women‟s and Children‟s Hospital in Adelaide. Thank you also to Dr. Andre Rickers and Rob King from Geneworks in Adelaide for their assistance with the genomic sequencing. I would also like to thank Professor Tim Mitchell from the University of Glasgow for his suggestions for my project and for his hospitality during my week in Glasgow. Thank you also to Charlie Plumptre, Nadine Verhoeven, Daphne Mermans, Dr. Kerrie Grabowicz, Dr. Trisha Rogers, Ursula Talbot, Dr. Hui Wang, Dr. Alistair Standish, Dr. Kim LeMessurier, Dr. Damien Chong and Melissa Chai for making the Paton lab such a great place to work. A

special thank you must also go to Brock Herdman for his continued company and entertainment since honours. Finally, I must thank my family. In particular, I would like to thank my parents for providing me with ongoing opportunities, support and encouragement. Thank you also to my daughter Lilly for spewing on some of the drafts of my thesis and for those beautiful smiles when I get home in the evening – Daddy loves you!! I would like to especially thank my wife Katie for not only her assistance in editing and formatting this thesis, or for doing the bulk of the housework while I've been writing up, but for her constant love and support, which makes life so enjoyable.

Abbreviations

Abbreviations acceptable to the American Society for Microbiology are used without definition in this thesis. Additional abbreviations are defined when first used in the text, and are listed below.

3HIBDH	3-hydroxyisobutyrate dehydrogenase
ACT	Artemis comparison tool
aorE	Shikimate dehydrogenase
ARs	Accessory regions
BA	Blood agar
BCAAs	Branched-chain amino acids
BgaA	β-galactosidase
BHI	Brain heart infusion broth
BSA	Bovine serum albumin
cCAT	complete-CAT medium
CcpA	Catabolite control protein A
CCR	Carbon catabolite repression
CD	Conserved domain
CGH	Comparative genomic hybridisation
ChoP	Phosphorylcholine
CI	Competitive index
Cml	Chloramphenicol
CSOM	Chronic suppurative otitis media
CSP	Competence stimulating peptide
CTM	cCAT medium supplemented with BSA

DC	Dendritic cell
Ddl	D-alanine-D-alanine dehydrogenase
Dgk	Diacylglycerol kinase
dNTPs	Deoxyribonucleoside triphosphates
DEPC	Diethyl pyrocarbonate
DOC	Sodium deoxycholate
Erm	Erythromycin
GAPDH	Glycerolaldehyde-3-phosphate dehydrogenase
GalE	UDP-glucose 4-epimerase
GDH	Glucose-6-phosphate dehydrogenase
Gen	Gentamycin
Gki	Glucose kinase
HMM	Hidden Markov Model
HylA	Hyaluronate lyase
ICE	Integrative conjugative element
IFN-γ	Interferon γ
IL-1	Interleukin-1
i.n.	Intranasal
i.p.	Intraperitoneal
IPD	Invasive pneumococcal disease
IR	Input ratio
KEGG	Kyoto Encyclopaedia for Genes and Genomes
LD	Limit of detection
LTA	Lipoteichoic acid
LytA	N-acetylmuramoyl-L-alanine amidase
MLST	Multi-locus sequence typing

MQ	MilliQ
MSHR	Menzie‟s School of Health Research
NAL	N-acetylneuraminate lyase
NanA	Neuraminidase A
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NET	Neutrophil extracellular trap
$NmlR_{sp}$	MerR-like regulator
Nov	Novobiocin
NplT	Neopullulanase
OM	Otitis media
O/N	Overnight
OR	Ouput ratio
ORF	Open reading frame
PavA	Pneumococcal adherence and virulence factor A
PblB	Platelet-binding protein B
PBS	Phosphate buffered saline
PBP	Penicillin-binding protein
PCV7	7-valent pneumococcal conjugate vaccine
PezAT	PezA-PezT TA system
PFGE	Pulsed-field gel electrophoresis
Pht	Pneumococcal histidine triad protein
Pit	Pneumococcal iron transport
Ply	Pneumolysin
PPI-1	Pneumococcal pathogenicity island 1
PPSV23	23 valent pneumococcal polysaccharide vaccine

PsaA	Pneumococcal surface adhesion A
PspA	Pneumococcal surface protein A
PspC	Pneumococcal surface protein C
PsrP	Pneumococcal serine rich protein
PTS	Phosphotransferase system
RBS	Ribosome-binding sites
RecP	Transketolase
Rel_{sp}	RelA/SpoT homologue
rPAF	Platelet-activating factor receptor
RT	Room temperature
SB	Serum broth
SD	Standard deviation
SDg	Shine-Dalgarno
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
SNPs	Single nucleotide polymorphisms
Spe	Spectinomycin
Spi	Signal peptidase I
SpxB	Pyruvate oxidase
ST	Sequence type
Strep	Streptomycin
StrH	β-N-acetylglucosaminidase
TA	Toxin-antitoxin
TBE	Tris borate and EDTA
TE	Tris EDTA
Tet	Tetracycline

THY	Todd-Hewitt broth supplemented with yeast extract
TLR-4	Toll-like receptor 4
TMP	Tympanic membrane perforation
TNF	Tumour necrosis factor
TSB	Tryptic soy broth
WCH	Women‟s and Children‟s Hospital
WHO	World Health Organisation
Xpt	Xanthine phosphoribosyltransferase
ZmpB	Zinc metalloproteinase B