Genetic diversity and distribution of the pneumococcal surface lipoproteins and implications on potential protein-based vaccines

Ebrima Bojang
University of Cambridge
Wellcome Trust Sanger Institute

This dissertation is submitted for the degree of Master of Philosophy
August 2017

Hughes Hall College

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This thesis did not exceed the prescribed word limit by the Faculty of Biology.

Abstract

Streptococcus pneumoniae causes life-threatening diseases such as meningitis, sepsis and pneumonia. Over half a million children under 5 years die annually of pneumococcal disease. However, most of these deaths occur in resource-limited countries mostly in sub-Saharan Africa and Asia. Based on the antisera binding pattern of the capsules, the pneumococcus has almost 100 serotypes and the currently licensed vaccines are serotype specific and target only a subset of these serotypes. The 23-valent polysaccharide vaccine is not immunogenic in young children and the conjugate vaccines, which are immunogenic in young children cover only a small number of serotypes and are expensive to manufacture. Furthermore, there is serotype replacement with non-vaccine type serotypes in both carriage and disease.

Consequently, there has been much interest in finding alternative vaccine candidates that are serotype independent, less expensive to produce and most importantly, can induce sufficient immune response. Several pneumococcal proteins have been evaluated for their potential as vaccine candidates with mixed results.

Using reverse vaccinology, I have taken a holistic approach to look at the level of diversity and distribution of core (≥90% presence in my dataset) pneumococcal surface lipoproteins and predicted their immunogenicity. First, I screened all the genomes for surface exposed lipoproteins using established patterns. The candidate proteins also underwent immunogenicity screening and these proteins were ranked based on their potential as vaccine candidates.

The final candidate proteins include previously evaluated lipoproteins PsaA, AdcA, AdcAII, PiuA, PiaA as well as several new candidates that have not been evaluated in detail thus far, including YesO_2, TauA and PrsA.

Acknowledgement

First, I would like to thank Allah (SWT) for giving me life, health and the aptitude to undertake this dissertation.

Second, I am much grateful to my supervisor Professor Stephen Bentley who has been extremely helpful and supportive throughout this year. I would also like to thank Dr Simon Harris and Dr Lucy Weinert both of whom were on my thesis committee and helped guide this project by sharing with me their expertise. I would also like to say thank you to Chrispin Chaguza who took his time to introduce me to the tools available within the Infectious Genomics group. I would like to acknowledge everyone in E204 especially Rebecca Gladstone for always being there to answer my questions.

I would also like to thank my employer, MRC Unit, The Gambia for giving me the opportunity to undertake this prestigious degree. I am indebted to Dr Brenda Kwambana-Adams, Professor Martin Antonio and Dr Assan Jaye for seeing the potential in me and nominating me for this scholarship.

Thank you to the Wellcome Trust Sanger Institute for funding my studies and to the Graduate office for always assisting with all my needs

Finally, I would like to thank all my family and friends for their support and words. I dedicate this degree to my mum and dad. Both of them were instrumental in instilling discipline in me and making me realise that with hard-work, I can achieve all my dreams. Special thank you to my beautiful wife for the unconditional love and always being there for me. She has been my rock through-out this journey.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Full name**
BCG	Bacillus Calmette-Guérin
CASP	Critical assessment of protein structure prediction
CBP	Choline binding protein
I-TASSER	Iterative threading assembly refinement
IL	interleukin
IPD	Invasive pneumococcal disease
NVT	Non-vaccine type
PAF	Platelet-Activation Factor
PDB	Protein Data Bank
Phyre	Protein Homology/AnalogY Recognition Engine
PI	Protrusion Index
SNP	Single Nucleotide Polymorphism
STGG	Skim milk-Tryptone-Glucose-Glycerol
VT	Vaccine type
WGS	Whole genome sequencing