Genomic diversity of *Streptococcus pneumoniae* serotype 1 across sub-Saharan Africa

Simon R. Harris on behalf of the Pneumococcal African Genomics Consortium*

Background

Streptococcus pneumoniae serotype 1 strains are responsible for a significant proportion of invasive disease in sub-Saharan Africa and, in some regions, are associated with lethal epidemics. Paradoxically, however, serotype 1 strains are rarely detected in carriage in the nasopharynx, suggesting a distinct lifestyle or niche compared to other serotypes. The propensity of serotype 1 for disease causation, particularly in Africa, could make it more likely for these strains to fill the disease niche left following the removal of vaccine serotypes and could further increase the risk of lethal epidemics. Thus there is a pressing need to understand the unique nature of invasive serotype 1 strains, their geographical distribution, and whether genotypes vary between carriage and disease manifestations. This will provide the basis for post-vaccine-introduction surveillance, allow us to begin to determine any future epidemic potential for serotype 1 post-vaccination implementation and identify strategies to minimize the impact.

Study goal

Determine the phenotypic and genotypic relationships of serotype 1 to establish a platform for post-PCV introduction molecular-based surveillance and for the identification and evaluation of broadly cross-protective pneumococcal protein vaccine antigens that will augment or replace PCV vaccines in Africa.

Research aims

- To define the genomic diversity of serotype 1 isolates and establish the relative contributions of sequence variation, gene content, and genome organization to this diversity.
- To establish the relationship of this diversity to clinical presentation, epidemic potential, disease versus carriage, and establish whether epidemic, disease and carriage characteristics (such as pneumonia, bacteremia or meningitis) are associated with particular genes or lineages.
- To establish the extent of regional serotype 1 temporal and geographic variation.

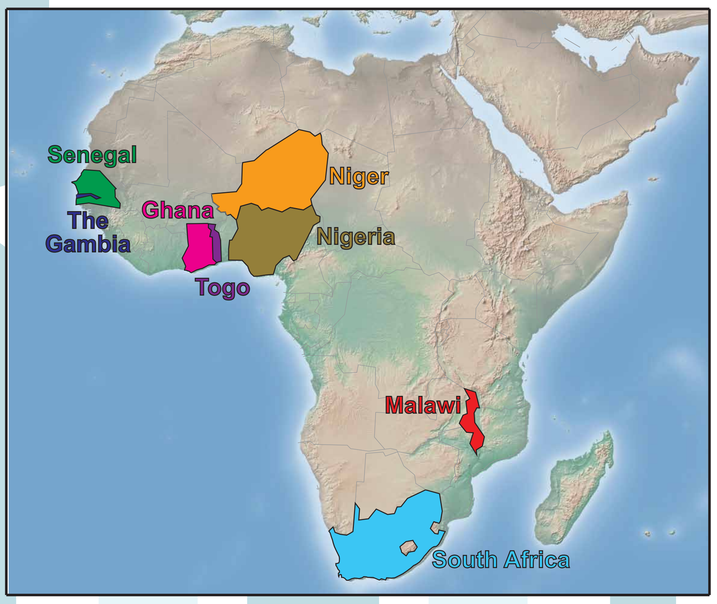
Methods

Samples were provided by four sites:
Malawi
Niger
South Africa
The Gambia

12-13 samples were selected from disease isolates at each site each year between 2005 and 2009. Blood and CSF samples were collected to reflect the relative prevalence of meningitis and bacteraemia. To reflect the relative abundance of paediatric serotype 1 strains, sampling was biased towards 75% paediatric (<15years old) and 25% adult samples.

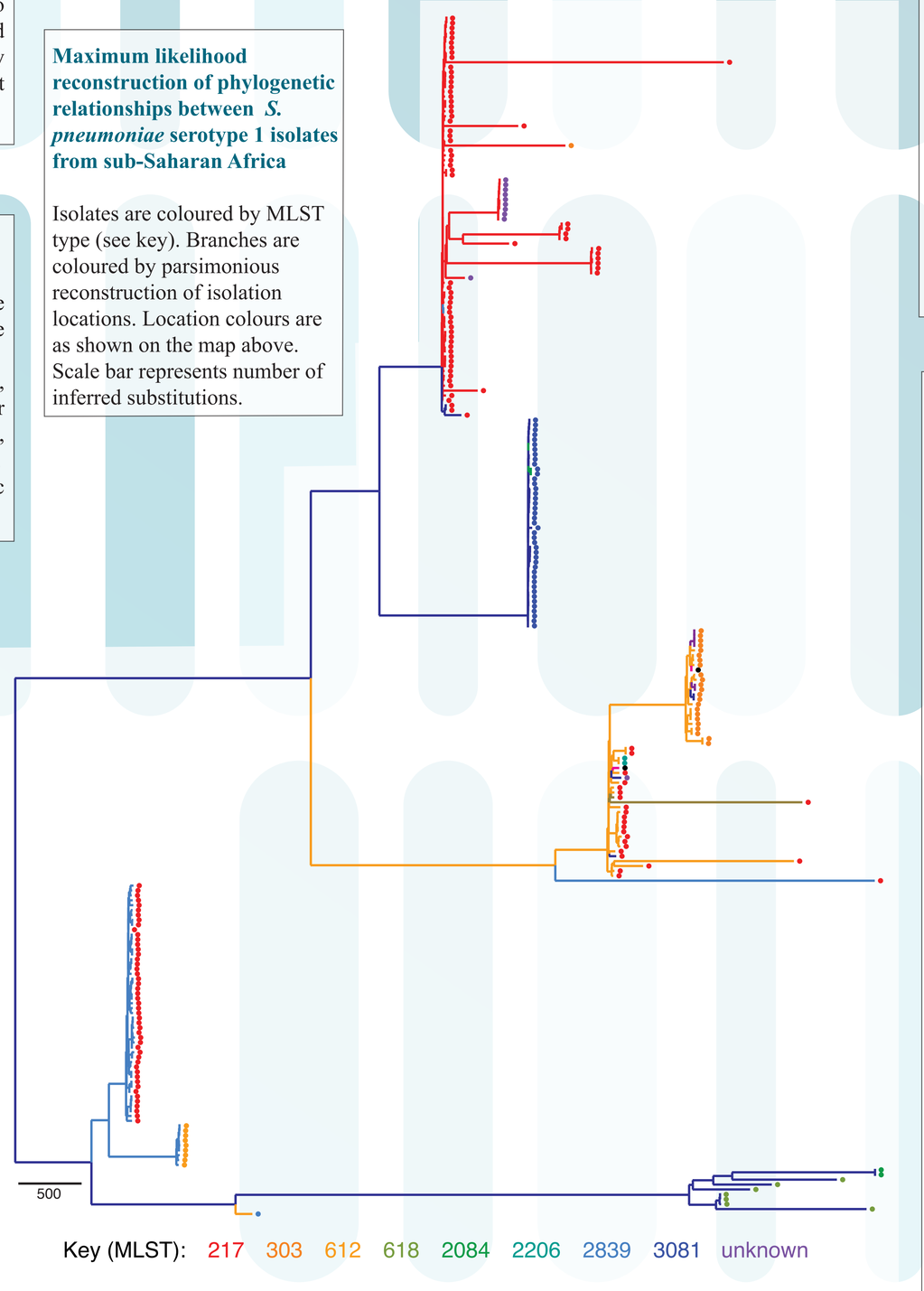
All samples were <SOMETHING ABOUT HOW THEY WERE GROWN/PREPPED?> and sequenced on the Illumina Genome Analyser II or Illumina HiSeq 2000.

Sequence reads were mapped against a complete serotype I reference genome from Ghana using SMALT[1] and variant sites identified in the core genome were used to reconstruct a phylogenetic tree of African sertoype 1 strains using RAxML[2].



Maximum likelihood reconstruction of phylogenetic relationships between *S. pneumoniae* serotype 1 isolates from sub-Saharan Africa

Isolates are coloured by MLST type (see key). Branches are coloured by parsimonious reconstruction of isolation locations. Location colours are as shown on the map above. Scale bar represents number of inferred substitutions.



Preliminary findings

Unlike large scale analyses carried out on other *S. pneumoniae* serotypes, the phylogenetic reconstruction of African serotype 1 strains showed strong geographical clustering (branch colours on tree) with each of the four major collection sites forming at least one distinct clade separated by large numbers of SNPs. This suggests the separation of these clusters in not recent.

The strong geographical clustering allows potential identification of geographical transmission events. For example, two isolates from South Africa cluster within the Malawian clade. Further research has shown that one of these isolates is from an individual from Mozambique.

Mapping MLST information (isolate colours on tree) onto the phylogeny shows good congruence between MLST types and the whole genome tree. ST217 is represented in the Malawian, South African and Niger clades, suggesting that 217 is the ancestral serotype 1 ST in Africa. The presence of ST217 in three divergent clades highlights the limitations of MLST as a typing tools. A previously reported replacement of one ST with another in Gambia is supported in the tree, which shows that the two Gambian clusters are highly divergent. Within the Malawian clade a lineage containing a novel MLST type has emerged.

Isolates from carriage and different disease groups did not form separate clusters on the tree, suggesting a similar genetic background in carriage and disease.

Similarly, isolates from adults and childern (<15) did not cluster separately which suggests that it is the same genotype that is circulating between children and adults.

* The Pneumococcal African Genomics Consortium is a Bill & Melinda Gates funded group including members from the following institutions:

Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW) Malawi
Medical Research Council (MRC) The Gambia
Centre de Recherche Médicale et Sanitaire (CERMES) Niger
Respiratory and Meningeal Pathogens Research Unit (RMPRU), National Institute for Communicable Disease (NICD) of the National Health Laboratory Service (NHLS) South Africa
Wellcome Trust Sanger Institute (WTSI) UK
University of Liverpool (UoL) Department of Clinical Infection Microbiology and Immunity, Institute of Infection and Global Health UK
Liverpool School of Tropical Medicine (LSTM) UK
The Rollins School of Public Health Emory University USA

Consortium members

MLW (Malawi): Dean Everett (UoL), Rob Heyderman (LSTM) & Neil French (UoL)
CERMES (Niger): Jean-Marc Collard
NHLS (SA): Anne von Gottberg & Keith Klugman (Emory)
MRC (The Gambia): Martin Antonio
WTSI (UK): Stephen Bentley and Julian Parkhill

Advisory group

Brian Greenwood, LSHTM
Richard Adegbola, Gates Foundation
Mark Alderson, PATH
Bill Hanage, Harvard
Orin Levine, JHU

Extending the sampling

As the Consortium develops in its analytical capacity and expertise, we will increase and broaden the partnership to other countries in the region. The data that emerges from this unique Consortium will provide the basis for an African-based initiative aimed at sampling and analysing genetic diversity across important carriage and disease-causing pneumococcal strains, providing molecular epidemiology based surveillance and identifying and thus enabling the development of appropriate and effective vaccine candidates for use in an African setting

We are asking for any interested group with serotype 1 isolates that they are able to provide to extend this project to contact the consortium to discuss possible collaboration. Please send all enquiries to dean.everett@liverpool.ac.uk.

References

1 http://www.sanger.ac.uk/resources/software/smalt
2 Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688-90.

PAGe
PNEUMOCOCCAL AFRICAN GENOMICS CONSORTIUM
BILL & MELINDA GATES foundation
NATIONAL HEALTH LABORATORY SERVICE
NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
MRC The Gambia Unit
wellcome trust sanger institute
CERMES Niger