Meningitis and Septicaemia 2019
November 5-6, 2019, British Museum, London.

Genomics: the power of WGS as a research and public health tool on a global scale and the GMGL initiative

Prof Martin C.J. Maiden, Department of Zoology, Peter Medawar Building, University of Oxford, South Parks Road, Oxford, OX1 3SY, United Kingdom.

More than ever, the control of infectious diseases is a global concern, requiring disease surveillance and public health responses to be international. The effective sharing of data is crucial to this enterprise, notwithstanding the fact that public health remains principally a responsibility of nation states. A fundamental requirement is unified nomenclatures that enable infectious agents to be accurately described and compared. Over the last 30 years the advent and development of nucleotide sequencing technologies has greatly facilitated this, by means of sequence-based typing, which is now possible up to the level of whole genome sequences (WGSs). Nucleotide sequences are definitive, reproducible, and widely comparable. As they are strings of letters, they are easily stored, transported, and manipulated with a wealth of analysis algorithms; however, as they are so large, with over two million letters for each meningococcus for example, they require interpretation to be useful and understandable, even by experts.

The open-access availability of these data has lagged behind their production for a variety of reasons. Barriers to effective sharing include imperatives of public health and privacy legislation (real and perceived) and concerns about commercial and research priorities (patents and publications). Large amounts of data are generated within local, reference, or research laboratories but are all-too-often ineffectively shared, even after publication. Frequently data are: (i) deposited in closed databases, which are only available to certain users, or (ii) placed open access in resources such as GenBank and the European Nucleotide Archive (ENA), but with little contextual or interpretative information, which limits their use. Those data collections that are available open-access are often contingent and not representative of global disease.

The Global Meningitis Genome Library (GMGL) Initiative aims to support the WHO global road map for defeating meningitis by 2030 by to addressing the issues of sequence data sharing for the four meningitis-associated pathogens: meningococcus; pneumococcus; group B streptococcus; and *Haemophilus influenzae*. It is inspired by, and builds on the on the success of, the Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL), which was founded by a collaboration among the Maiden and Tang Laboratories at the University of Oxford, Public Health England, and the Wellcome Trust Sanger Institute [1]. The MRF-MGL is hosted within the https://PubMLST.org databases, which was founded in 1998 for dissemination of sequence typing data. The GMGL will not be as a one-size-fits-all centralised database, but a collaborative initiative for data sharing, which will also act as a stimulus for the harmonisation of data collection, methodologies, nomenclature, and effective automated data sharing. It will also provide a single user interface that enables assess to a wide range of resources, including those hosted within https://PubMLST.org itself and other resources such as https://pubmed.gov, https://pathogen.watch/, https://microreact.org/ and many others. Active data curation and sophisticated data upload and sharing protocols will ensure data quality and enable the maintenance of confidentiality where this is essential.

1. Hill, D.M.C., J. Lucidarme, S.J. Gray, L.S. Newbold, R. Ure, C. Brehony, O.B. Harrison, J.E. Bray, K.A. Jolley, H.B. Bratcher, J. Parkhill, C.M. Tang, R. Borrow, and M.C.J. Maiden, *Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study.* Lancet Infect Dis, 2015. **15**(12): p. 1420-8.