EBioMedicine 2 (2015) 234–243

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Original Article

Genomic Analysis of the Evolution and Global Spread of Hyper-invasive Meningococcal Lineage 5

Odile B. Harrison [a,*], James E. Bray [a], Martin C.J. Maiden [a], Dominique A. Caugant [b,c]

[a] Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[b] WHO Collaborating Centre for Reference and Research on Meningococci, Norwegian Institute of Public health, P.O. Box 4404 Nydalen, NO-0403 Oslo, Norway
[c] Department of Community Medicine, Faculty of Medicine, University of Oslo, P.O. Box 1130 Blindern, NO-0318 Oslo, Norway

ARTICLE INFO

Article history:
Received 3 December 2014
Received in revised form 12 January 2015
Accepted 12 January 2015
Available online 13 January 2015

Keywords:
Neisseria meningitidis
Serogroup B
Genome sequencing
Evolution

ABSTRACT

Background: The predominant model for bacterial pandemics is the emergence of a virulent variant that diversifies as it spreads in human populations. We investigated a 40-year meningococcal disease pandemic caused by the hyper-invasive ET-5/ST-32 complex.
Methods: A global collection of *Neisseria meningitidis* isolates dating from 1969 to 2008 was whole genome sequenced (WGS) and analysed using a gene-by-gene approach at http://pubmlst.org/neisseria.
Findings: Analysis of WGS data identified a 'Lineage 5 pan genome' of 1940 genes, 1752 (92%) of which were present in all isolates (Lineage 5 'core genome'). Genetic diversity, which was mostly generated by horizontal gene transfer, was unevenly distributed in the genome; however, genealogical analysis of diverse and conserved core genes, accessory genes, and antigen encoding genes, robustly identified a star phylogeny with a number of sub-lineages. Most European and American isolates belonged to one of two closely related sub-lineages, which had diversified before the identification of the pandemic in the 1970s. A third, genetically more diverse sub-lineage, was associated with Asian isolates. Several isolates had acquired DNA from the related gonococcus.
Interpretation: These data were inconsistent with a single point of origin followed by pandemic spread, rather suggesting that the sub-lineages had diversified and spread by asymptomatic transmission, with multiple distinct strains causing localised hyperendemic outbreaks.

1. Introduction

Neisseria meningitidis, a Gram negative diplococcal bacterium, is normally a commensal resident of the oropharynx of a high percentage (10–30%) of the human population, very occasionally, causing life-threatening meningitis and septicaemia (Caugant and Maiden, 2009). The only well-established virulence factor of *N. meningitidis* is the polysaccharide capsule, which mediates resistance to complement-mediated lysis and opsonophagocytosis. Based on biochemical composition as well as genetic analysis, 12 serogroups have been described of which 6 (serogroups A, B, C, W, Y and, X) are associated with most disease worldwide (Harrison et al., 2013). Capsule polysaccharide conjugate vaccines have been successfully used to induce protective immunity against *N. meningitidis* serogroups A, C, W and, Y. However, due to similarities between the serogroup B polysaccharide and human glycoprotein structures, no such vaccine targeting this serogroup is available.

The genetic diversity and population structure of the species have been elucidated by the use of two related methods, multilocus enzyme electrophoresis (MLEE) starting in the early 1980s and multilocus sequencing typing (MLST) at the end of the 1990s (Caugant et al., 1986a; Maiden et al., 1998). Both methods assess genetic variation among isolates by indexing their whole genome through a small subset of representative housekeeping genes. With the availability of high-throughput Sanger DNA sequencing, MLEE was replaced by MLST, which presents the additional advantage of being fully portable through an Internet database (www.pubmlst.org/neisseria). Both methods produce equivalent data and the basic features of meningococcal populations first elucidated by MLEE were confirmed by MLST.

Of the thousands of genotypes, distinguished by MLEE and MLST most are rarely, if ever, associated with disease, in contrast to the handful of clonal complexes responsible for epidemics or even pandemics (Yazdankhah et al., 2004). Many serogroup B outbreaks since the 1970s have been caused by *N. meningitidis* isolates belonging to the ST-32 clonal complex, previously designated electrophoretic type (ET)-5 complex. ET-5 was first identified from a case of serogroup B meningococcal disease in Norway in 1969 (Caugant et al., 1987). In the succeeding years, a hyper-endemic wave of serogroup B meningococcal disease started in Norway with an incidence reaching 8.7 per 100,000 population in 1983, subsequently decreasing over the years to less than 1.0 per 100,000 population in 2000. Similar or closely related

* Corresponding author at: University of Oxford, Department of Zoology, The Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK.
E-mail address: odile.harrison@zoo.ox.ac.uk (O.B. Harrison).

http://dx.doi.org/10.1016/j.ebiom.2015.01.004

clones expressing the same or different major antigenic outer membrane proteins were responsible for high incidence of serogroup B disease in several other European countries in the 1980s and 1990s, as well as outbreaks and epidemics in Latin America, including Cuba, Chile, Brazil, and Argentina (Bygraves et al., 1999; Caugant et al., 1986b; Cruz et al., 1990; Sacchi et al., 1992; Wedege et al., 1995). The ST-32/ET-5 complex has also caused a prolonged outbreak in the Pacific-Northwest of United States in 1993 to 2007 (Diermayer et al., 1999). A long-lasting outbreak in Normandy, France, in the past decade has also been caused by the ST-32 complex (Rouaud et al., 2006). While serogroup B disease is rare in Asia and Africa, the few available serogroup B isolates were also linked to the same clonal complex. Thus, the ST-32 complex caused disease globally over a 40-year period (Caugant et al., 1987).

Complete "finished" genomes for two ST-32 complex isolates, MC58 and H44/76, have been published, the former originating from the United Kingdom in the 1980s and the latter from a case in Norway in 1976 (Tettelin et al., 2000; Piet, 2011). Both isolates have been extensively used in serogroup B vaccine research with MC58 pivotal to the design of the 4CMenB vaccine (BexSero®) through a technique known as reverse vaccinology and H44/76 used in the design of several outer membrane vesicle-based vaccines (Serruto et al., 2012; van der Ley and Poolman, 1992). Serogroup B vaccine research has focussed on surface-expressed proteins which, while being immunogenic, might also be strongly under selection pressure. It is, therefore, essential to elucidate how these vaccine antigens might change over time and during worldwide spread of a hyper-invasive clone.

WGS provides a new means to elucidate genomic variation within a clonal complex of *N. meningitidis* and this paper presents a gene-by-gene description of WGS data from a global selection of isolates belonging to the ST-32 clonal complex. A pipeline for the population annotation of WGS has been developed (Fig. 1) combining the use of: i) the Bacterial Isolate Genome Sequence platform (BIGSdb) hosted on the www.pubmlst.org/neisseria database which currently enables the curation of over 2000 *Neisseria* genes and, ii) the prokaryotic annotation tool, PROKKA for novel gene discovery (Seemann, 2014). Through comparison with reference genomes, the Lineage 5 core genome (Lineage 5 cgMLST) was defined and compared between isolates revealing three distinct clusters of isolates grouping by PorA type within which small localised clusters were also visible. Novel gene discovery identified the Lineage 5 pan genome (Lineage 5 pgMLST) and included type IV secretion systems (T4SS), haptoglobin–haemoglobin receptors associated with iron acquisition, as well as a gonococcal conjugative plasmid.

2. Methods

2.1. Isolate Collection, WGS and Assembly

Forty-three *N. meningitidis* isolates belonging to clonal complex ST-32, were selected from the strain collection at the WHO Collaborating Centre for Reference and Research on Meningococci in Oslo, Norway. This collection of isolates was chosen to be representative of the 40 year global pandemic forming a baseline dataset for the analysis of this lineage and others. Isolates were retrieved from storage, inoculated onto Columbia horse blood agar and incubated for 24 h at 37 °C in a 5% CO_2 atmosphere. Prior to DNA extraction, colonies were visually checked for purity. A number of quality controls are carried out following Illumina sequencing to confirm the purity of the samples and to check for contamination. Firstly, assembled genome sizes are verified to ensure they are within the expected *Neisseria* genome range (Table 1) as contaminated samples will contain a distinctly larger genome. Secondly, all samples are scanned for the 53 ribosomal gene proteins which are genus, species as well as clonal complex specific (Jolley et al., 2012a; Jolley and Maiden, 2013). Should a sample be contaminated, conflicting alleles for the ribosomal genes will be apparent. *Neisseria* species can be further identified using the 50 S ribosomal protein L6 (*rplF*) gene (Bennett et al., 2014) and once again, contaminated DNA samples will contain conflicting *rplF* sequence data. The genomes of MC58 (accession numbers in brackets AE002098) and H44/76 (AEQZ00000000), from the UK and Norway respectively, were included as well as isolate, CU385 (AEQJ01000000) (Tettelin et al., 2000; Piet,

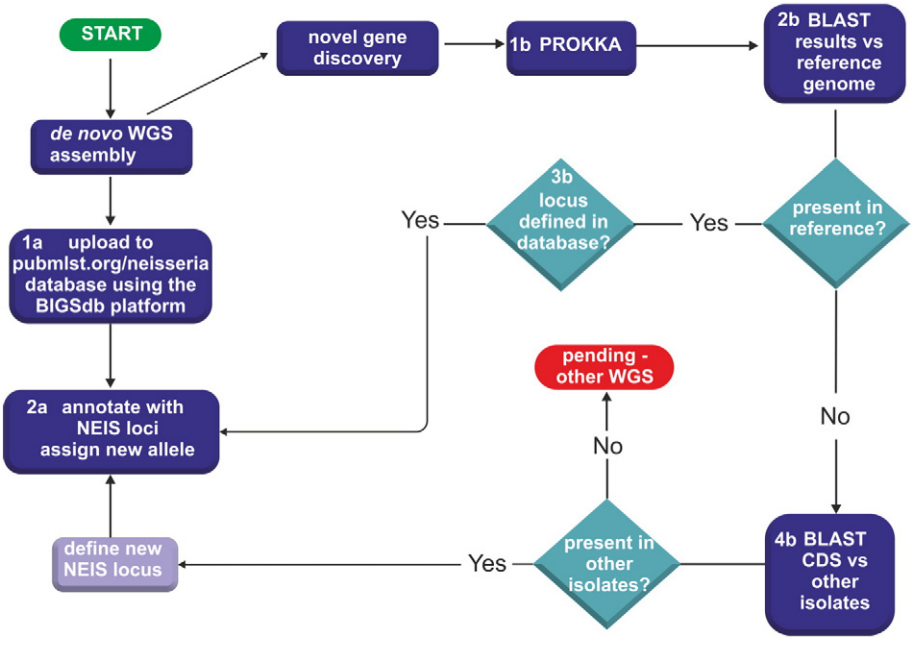

Fig. 1. Population annotation pipeline. This pipeline provides a generalizable approach to the curation and annotation of WGS which can be applied to other lineages. It combines the use of: i) the Bacterial Isolate Genome Sequence database (BIGSdb) and, ii) the prokaryotic annotation tool PROKKA for novel gene discovery. At the time of writing, over 2000 *Neisseria* genes (NEIS loci) had been defined and, once deposited in the database (step 1a), WGS data were automatically annotated with NEIS loci (step 2a). Novel gene discovery used PROKKA (step 1b) and new genes were screened against reference genomes (2b) as well as NEIS loci defined in the database (3b), thereby eliminating genes which were already being curated. The remaining genes were then checked against WGS data belonging to other isolates enabling the distribution of novel genes to be determined (4b). Before new loci were defined in the database, these were checked in the genome annotation tool, Artemis.