

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Clark, Stephen A; Doyle, Ronan; Lucidarme, Jay; Borrow, Ray; Breuer, Judith; (2018) Targeted DNA enrichment and whole genome sequencing of *Neisseria meningitidis* directly from clinical specimens. *INTERNATIONAL JOURNAL OF MEDICAL MICROBIOLOGY*, 308 (2). pp. 256-262. ISSN 1438-4221 DOI: <https://doi.org/10.1016/j.ijmm.2017.11.004>

Downloaded from: <http://researchonline.lshtm.ac.uk/4652525/>

DOI: <https://doi.org/10.1016/j.ijmm.2017.11.004>

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: Open Government License <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

<https://researchonline.lshtm.ac.uk>



Targeted DNA enrichment and whole genome sequencing of *Neisseria meningitidis* directly from clinical specimens

Stephen A. Clark^{a,*}, Ronan Doyle^b, Jay Lucidarme^a, Ray Borrow^a, Judith Breuer^b

^a Meningococcal Reference Unit, Public Health England, Manchester Royal Infirmary, Oxford Road, Manchester, UK

^b Division of Infection and Immunity, University College London, Gower Street, London, UK



ARTICLE INFO

Keywords:

Meningococcal
Whole genome sequencing
DNA enrichment
RNA baits
Clinical specimens

ABSTRACT

In England and Wales, approximately one half of all laboratory-confirmed meningococcal disease cases fail to yield a viable invasive isolate, primarily due to the use of antibiotics. Characterisation of non-culture meningococci has been restricted to the detection or sequencing of specific gene targets within clinical specimens. In this study we investigated the ability of the Agilent SureSelectXT kit to facilitate DNA enrichment and genome sequencing of meningococcal DNA within a small panel of blood and CSF specimens. A target-specific RNA oligonucleotide bait library was used to capture and enrich the bacterial DNA prior to next generation sequencing. A positive correlation between meningococcal DNA amount and genome coverage was observed with eight of the ten specimens producing genomes of acceptable quality. All commonly-used typing information derived from each acceptable non-culture genome matched those of an isolate from the same patient and the paired genomes showed a high level of congruence across indexed loci. We estimate that this technique could be used to perform whole genome sequencing on up to ~45% of the positive specimens received by the Public Health England's Meningococcal Reference Unit. Further optimisation of the extraction and/or enrichment processes may, however, increase the proportion of non-culture cases from which quality genomes can be obtained.

1. Introduction

Despite the widespread use of effective meningococcal vaccines, *Neisseria meningitidis* continues to cause endemic and epidemic invasive meningococcal disease (IMD) throughout the world (Ali et al., 2014; Pelton, 2016). A new generation of protein-based vaccines is being introduced with the aim of protecting against a greater proportion of the diverse strains circulating within target populations (Gasparini et al., 2014). Concurrently, due to developments in next generation sequencing technology, whole genome sequencing (WGS) of meningococcal isolates has become a routinely-used tool and provides invaluable information regarding genetic relatedness, antigenic distribution and metabolic characteristics (Brehony et al., 2015; Lucidarme et al., 2015; Watkins and Maiden, 2017).

Whilst WGS is more cost-effective than ever before, a major limitation is the requirement of viable bacteria from which sufficient amounts of DNA can be extracted. In England and Wales, the early administration of antibiotics prevents cultivation of meningococci from clinical specimens (e.g. blood and CSF) in approximately one half of laboratory-confirmed IMD cases. These cases are confirmed using real-

time PCR on DNA extracts from clinical specimens (Heinsbroek et al., 2013). Although sequencing of important gene targets can be carried out from these specimens (Clark et al., 2014), the low concentration of meningococcal DNA and presence of human host DNA prevents WGS using standard methods. The lack of genomic data in these cases limits the scope of vaccine strain coverage predictions and prevents high-resolution multilocus sequence typing (MLST) analyses, which can form the basis of the public health response to outbreaks and changes in IMD epidemiology (Chatt et al., 2014; Parikh et al., 2017).

To overcome these limitations, a number of commercial DNA enrichment strategies have been developed (Kozarewa et al., 2015). These techniques allow the sequencing of specific exons, genomic regions or whole genomes within mixed samples by purifying and concentrating the desired DNA fragments. One such approach involves using specific, biotinylated oligonucleotide baits that can hybridise to the target genomic fragments. Streptavidin-coated magnetic beads are then used to separate the hybridised fragments from the rest of the sample before the isolated DNA is amplified and sequenced (Bodi et al., 2013). The Agilent SureSelectXT Target Enrichment system, which uses a tailored RNA bait library to capture the target DNA has been successfully used

* Corresponding author.

E-mail address: stephen.clark@phe.gov.uk (S.A. Clark).

to enrich and sequence the genomes of herpesviruses, *Chlamydia trachomatis* and *Mycobacterium tuberculosis* directly from clinical specimens (Brown et al., 2015; Christiansen et al., 2014; Depledge et al., 2011). The aim of this work was to assess the effectiveness of the SureSelectXT system in facilitating WGS of diverse meningococcal strains directly from clinical specimens and calculate the proportion of non-culture cases confirmed by the Public Health England Meningococcal Reference Unit (MRU), that could be comprehensively characterised in this way.

2. Methods

2.1. Clinical specimens and DNA extraction

Clinical specimens were selected from group B IMD cases which also yielded a viable isolate (from a separate sample). The corresponding isolates had previously undergone WGS and the genomic data are available within the PubMLST database (<https://www.pubmlst.org/neisseria>) (Jolley and Maiden, 2010). The MLST data for these isolates were used to select ten non-culture clinical specimens (five blood and five CSFs) containing DNA from meningococci with distinct STs representing seven clonal complexes (and one singleton). The selected clinical specimens and corresponding isolates are listed in Table 1.

The real-time PCR cycle threshold (Ct) values of the specimens, derived using a *ctrA*-directed Taqman[®] assay (Heinsbroek et al., 2013), varied from 20 to 39 (Table 1). The concentration of meningococcal DNA within each specimen was estimated using previously-published Ct values obtained from quantified meningococcal DNA extracts (Clark et al., 2014). These values were used to generate a standard curve against which the Ct values of the ten selected specimens could be compared in order to estimate the meningococcal DNA concentration. The estimated DNA concentrations were calculated to be from 2 fg/μL to 153 pg/μL.

The available specimen volumes ranged from 40 to 300 μL. DNA extraction was carried out using the EZ1 DNA Blood Kit on an EZ1 Advanced XL (Qiagen, UK) as per manufacturer's instructions with an additional bead-beating bacterial cell lysis step. Sample volumes of 300 μL were used. Specimens with available volumes of < 300 μL were diluted in molecular grade water as required. The available specimen volumes were used along with the calculated DNA concentrations to estimate the total amount of meningococcal DNA available for extraction (Table 1).

2.2. In silico generation of 120mer RNA bait sequences

RNA baits of 120mer length were designed using sequences from 77 complete *N. meningitidis* genomes available on NCBI GenBank (www.ncbi.nlm.nih.gov/genbank) and 2898 *N. meningitidis* whole/draft genomes within PubMLST. These genomes represent a diverse array of capsular groups and clonal complexes. A previously-designed in-house Perl script was used to generate overlapping bait sequences to provide 2X coverage across all reference sequences (Depledge et al., 2011). The *in silico*-generated sequences were re-mapped on to both human and *N. meningitidis* genomes to check for specificity and major gaps in the bait set. Additional baits were generated separately to match any gaps identified and added to the overall bait set as required. Baits with specificity to the human genome were removed. The bait library sequences were uploaded to SureDesign and the oligonucleotides were manufactured by Agilent Technologies.

2.3. Library preparation and hybridisation

Preparation and enrichment of the Illumina sequencing libraries was performed using the SureSelectXT reagent kit for the MiSeq sequencing platform (Agilent, US) in accordance with the manufacturer's instructions.

Briefly, the Agilent SureSelectXT library preparation protocol

Table 1
Details of the selected non-culture specimen panel and corresponding isolate MLST data.

Specimen number	Specimen type	Ct value ^a	Estimated NM DNA concentration (pg/μL)	Specimen volume (μL)	Estimated amount of NM DNA extracted (pg)	Isolate ID	Isolate ST	Isolate CC
1	Blood	39	0.002	200	0.423	M15 240120	2100	ST-213 complex
2	Blood	32	0.130	120	15.654	M15 240017	485	ST-41/44 complex
3	CSF	27	2.480	40	99.198	M15 240178	11481	ST-35 complex
4	CSF	27	2.480	55	136.398	M15 240270	11483	none
5	CSF	26	4.469	55	245.814	M15 240103	1946	ST-461 complex
6	Blood	28	1.376	300	412.826	M15 240142	749	ST-32 complex
7	CSF	26	4.469	120	536.322	M15 240805	213	ST-213 complex
8	Blood	24	14.516	100	1451.587	M15 240650	1049	ST-269 complex
9	CSF	23	26.160	75	1962.022	M15 240240	1161	ST-269 complex
10	Blood	20	153.123	300	45936.960	M15 240180	162	ST-162 complex

^a determined using *ctrA*-specific real-time PCR. NM = *Neisseria meningitidis*.

involved fragmentation of the DNA using a Covaris ultrasonicator (Covaris, US) followed by purification, end-repair and ligation of Illumina adapters. The DNA was then amplified by adapter-mediated PCR prior to in-solution hybridisation using the meningococcal-specific RNA capture baits. Streptavidin-coated magnetic beads were used to isolate the biotinylated baits and hybridised DNA fragments. The captured DNA was then once again amplified and indexed prior to multiplex sequencing.

2.4. Sequencing and genome assembly

Sequencing of captured DNA libraries was performed on an Illumina MiSeq using the 500-cycle V2 kit (Illumina, US) in accordance with the manufacturer's protocol.

All reads were screened for human sequences by mapping to a human reference genome (Hg19, accession number: GCF.000001405.13) using BMap (version 37.00). Unmapped reads were corrected for predicted sequencing error using Lighter (Song et al., 2014) and overlapping read pairs were merged with FLASH (Magoč and Salzberg, 2011). All merged paired reads and unpaired reads were then retrieved and *de novo* assembled using SPAdes (Bankevich et al., 2012) without read error correction, assembler only, using k-mer values of 21, 33, 55, 77, 99 and 127.

2.5. Estimation of specimen genome coverage and read depth

To assess genome coverage, organise the assembled contigs and remove possible contaminants (e.g. residual human sequences), the *de novo*-assembled contigs from each sample were aligned to the most closely-related complete meningococcal genome available in the PubMLST database (Fig. 1). For each strain, the most closely-related complete genome was identified through distance estimation based on shared K-mers between 237 complete genomes available on PubMLST and the corresponding isolate genome using Mash (Ondov et al., 2016). The following complete genomes: NZ-05/33, M04-240196, M01-

240355 (Budroni et al., 2011), α 710 (Joseph et al., 2010), MC58 (Tettelin et al., 2000), α 153 (Schoen et al., 2008), 12–176 (no ref, pubMLST ID 41342) and M22236 (Kretz et al., 2016) were used to create pseudo-references (Table 2). The contigs for each specimen genome were then aligned against the corresponding complete genome using the BWA-MEM algorithm (Li, 2013). The consensus sequences for the contigs aligned to the complete genomes were used as pseudo-references against which the screened sequencing reads were re-mapped to assess genome coverage and depth (Fig. 1).

2.6. Gene-by-gene comparison of specimen and isolate genomes

A gene-by-gene comparison of the specimen and corresponding isolate sequence data was performed by submitting the *de novo*-assembled contigs to the PubMLST database for automated gene annotation (Fig. 1). Commonly used typing data: the seven MLST loci, PorA VR1 and VR2, PorB, FetA VR, Factor H-Binding Protein (fHbp), Neisserial Heparin-Binding Antigen (NHBA) and *Neisseria* Adhesin A (NadA) were extracted for each of the genome pairs and compared.

To expand the analysis to a wider panel of genes, the PubMLST Genome Comparator tool was used to compare the paired genomes in terms of all catalogued NEIS loci ($n = 2652$, accessed on 08/02/2017), where present. Genes which could not be annotated due to missing data/incomplete assembly in one or both of the paired genomes were removed from the analysis and, among the remaining, characterised loci, those which had differing allele IDs were selected for further allelic comparison. The alleles were aligned and compared to quantify the number of discrepant bases. Single differences between two alleles were considered to be due to PCR error or single nucleotide polymorphisms. Genes with multiple differences were further analysed in order to better explain the differences. BLAST searches were performed within the available *N. meningitidis* genomes in NCBI Genbank in order to identify multiple genes copies (paralogy) and/or matching gene fragments.

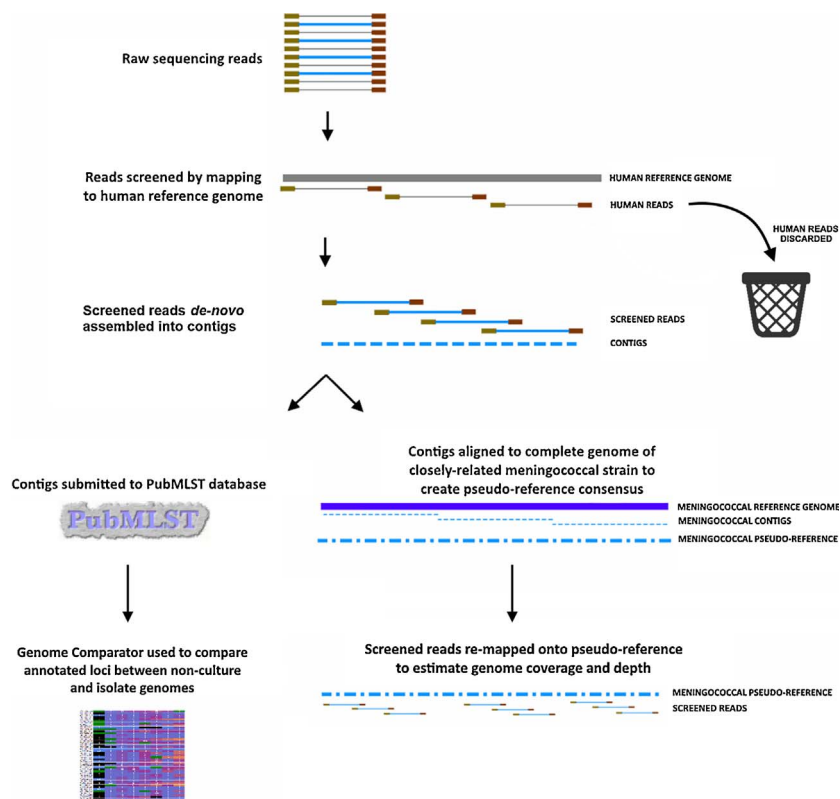


Fig. 1. Overview of analysis of meningococcal sequencing reads and assembled contigs. Raw reads were screened by aligning to human reference. Screened reads were *de-novo* assembled. Genome coverage and read depth was estimated by aligning against a pseudo-reference genome and the contigs were submitted to PubMLST to allow a gene-by-gene comparison of non-culture and corresponding isolate genomes.

Table 2

Statistics for read alignment to human reference and meningococcal pseudo-reference consensus for each clinical specimen.

Sample	No. of read pairs	No. of reads mapped to human reference	% of reads mapped to human ref.	Closest-related complete genome	No. of reads mapped to NM pseudo-ref.	% of reads mapped to NM pseudo-ref.	% genome coverage $\geq 1 \times$	Median read depth
1	1,913,131	1,749,684	91.5	alpha710	1112	0.1	7.2	0
2	1,796,259	1,627,329	90.6	NZ-05/33	22,865	1.3	70.8	2
3	1,769,239	1,518,863	85.8	M04 240196	84,281	4.8	93.7	10
4	1,819,385	1,521,127	83.6	alpha153	124,710	6.9	96.4	14
5	1,932,925	1,635,041	84.6	M22236	134,565	7.0	92.4	17
6	1,802,516	1,562,163	86.7	MC58	91,072	5.1	93.8	10
7	1,959,202	1,070,035	54.6	M01 240355	730,805	37.3	98.1	89
8	2,358,677	964,492	40.9	M04 240196	1,250,017	53.0	98.6	149
9	1,594,558	898,513	56.3	M04 240196	563,338	35.3	96.7	73
10	3,276,069	944,227	28.8	12–176	2,098,332	64.1	95.6	254

NM = *Neisseria meningitidis*.

3. Results

3.1. Sequencing reads and screening of human reads

Table 2 lists the numbers of total read pairs generated from each sample and numbers and proportions of those read pairs that mapped onto the human reference and meningococcal pseudo-reference genomes. The total number of read pairs generated from each specimen varied from 1,594,558 to 3,276,069. The proportion of total read pairs aligning to the human reference ranged from 28.8% to 91.5%. For eight of the ten specimens, > 50% of the total reads were human derived. The proportion of reads aligning to the pseudo-reference meningococcal genomes ranged from 0.1% to 64.1% (Table 2) and correlated positively with the calculated meningococcal DNA amount.

3.2. Meningococcal genome coverage and read depth

The closest-related complete genome for each sample, the estimated genome coverage and sequencing read depth against the meningococcal pseudo-reference consensus are listed in Table 2. The genome coverage ranged from 7.2% to 98.6%, with eight of the ten genomes exhibiting > 92% coverage of their respective pseudo-references. Overall, a clear correlation was observed between the extracted meningococcal DNA load, the median read depth and estimated genome coverage (Fig. 2).

The two specimens with < 100 pg of DNA available (specimens 1 and 2) yielded low genome coverage (7% and 71%, respectively) and ‘shallow’ read depth (< 3 median depth). For the remaining specimens, there was a positive correlation between the DNA amount used and the

median read depth, however, above this 100 pg threshold, the genome coverage plateaued at 90–100% (Fig. 2).

3.3. Gene-by-gene, culture vs specimen genome comparison

One of the ten specimen genomes (specimen 1) was not accepted for inclusion into the PubMLST database due lack of matches to *Neisseria* loci. This sample had the highest Ct value (Ct 39), the lowest estimated DNA load (423 fg) and the lowest estimated genome coverage (7.2%).

Table 3 provides details on contig properties and the indexed genomic data extracted from PubMLST. Substantially more contigs were produced from assembly of the specimen reads when compared to the genomes of the isolates (specimen mean 1314 vs. 342 for isolates). Excepting the two poor quality specimen genomes (specimens 1 and 2), the cumulative contig length of the specimens was greater than the corresponding isolate genomes. This is likely to be the result of enduring host DNA, and/or DNA from environmental contaminants, within the assembly.

Following automated gene annotation, genotypic data for common typing targets were extracted from the nine specimen genomes and the corresponding isolates (Table 3). Comparisons of the annotated loci among specimen and isolate genomes revealed perfect concordance in terms of commonly-used typing information. Four of the specimen genomes yielded complete typing information, whilst the remaining five genomes failed to cover at least one antigenic and/or MLST genes. There was a clear correlation between the estimated amount of meningococcal DNA present in the specimens and the coverage of the typing loci. For all genomes with complete typing data, > 500 pg of DNA had been utilised to create the sequencing libraries, whilst those with < 500 pg available yielded incomplete data. The genomes with complete data also exhibited higher median read depths ($\geq 73 \times$).

To expand the analysis, genome comparisons in terms of all NEIS loci ($n = 2652$), where present, were carried out. The number of NEIS loci that were covered by both the specimen and isolate genomes for a given case (shared NEIS loci) varied from 945 to 1780 (Table 3). Following genome comparisons, 43 discrepant alleles were identified across all genome pairs (from 0.11% to 1.16% of shared NEIS loci; Table 3). The discrepant loci for each genome pair are listed in a Supplementary Table 1. Genes with discrepant alleles that differed at more than one residue (23/43) were checked for possible paralogous loci within NCBI Genbank genomes. All but one of the genes featuring > 1 bp polymorphisms were found to generate significant additional hits in complete genomes within Genbank (Table S1). The alleles of the remaining gene (NEIS0534) differed by 4 bp between specimen 4 and isolate M15 240270. All four differences were within a 28 bp region of the first half portion of the gene (all synonymous mutations). No similar sequences were found among human genomes on NCBI Genbank. This gene codes for a 30S ribosomal subunit (S16) and is highly conserved amongst invasive strains (data not shown). The cause of

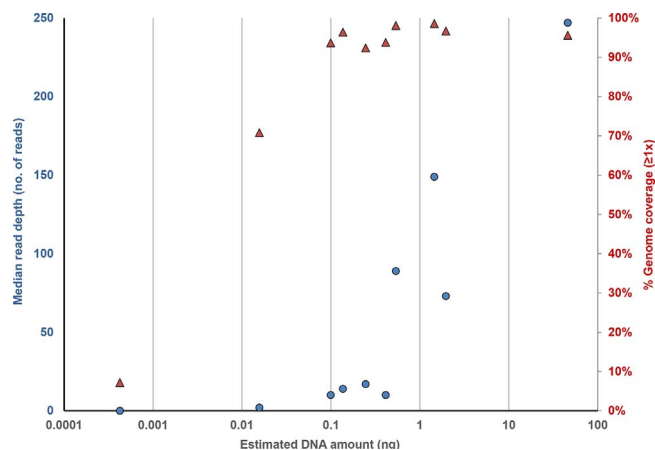


Fig. 2. Correlations between the estimated meningococcal DNA load in each specimen (X axis) and the median read depth and genome coverage. Circles: median read depth (left Y axis). Triangles: % genome coverage (right Y axis).

Table 3
Contig properties, extracted typing data and locus overview of ten non-culture genomes and corresponding isolate genomes.

Specimen/ isolate	DNA Source	No. of contigs	Cum. contig length	No. of tagged loci	No. of tagged NEIS loci	PorA VR1	PorA VR2	'PorB	FetA VR	fHbp peptide	NHBA peptide	NadA peptide	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	ST	Clonal complex	Discrepant alleles/shared NEIS loci (n)
1	Specimen	889	574,775																			
NOT ACCEPTED BY PUBMLST																						
M15 240120	Isolate	342	2,182,243	1907	1750	22	14	3-314	F3-6	45	18	0	7	5	1	13	9	53	9	2100	ST-213 complex	n/a
2	Specimen	1373	996,602	0	0																	
M15 240017	Isolate	362	2,101,025	2000	1806	12-1	16	3-817	F1-5	4	2	0	3	6	9	5	8	6	9	485	ST-41/44 complex	n/a
3	Specimen	1459	2,363,572	1193	1075	21-2	28	3-45	F5-2	24		0	4	10	1	9	3	10	12	11481	ST-35 complex	8/1052
M15 240178	Isolate	344	2,174,757	1938	1785	21-2	28	3-45	F5-2	24	29	0	4	10	1	9	3	10	12	11481	ST-35 complex	
4	Specimen	1224	2,421,445	1446	1313	22	9	3-818	F1-8	13	24	0		5	34	17	3	172	124			4/1287
M15 240270	Isolate	302	2,084,949	1848	1692	22	9	3-818	F1-8	13	24	0	12	5	34	17	3	172	124	11483	n/a	
5	Specimen	2449	2,602,857	1466	1345	19-2	13-9	3-14	F3-9	47		0	12		35	192	22	17				4/1318
M15 240103	Isolate	399	2,200,708	1923	1720	19-2	13-9	3-14	F3-9	47	118	0	12	5	12	35	192	22	17	1946	ST-461 complex	
6	Specimen	1509	2,271,736	1073	971	19	15	3-1	F5-1	1		1		10		4	6	3				11/945
M15 240142	Isolate	315	2,165,187	1963	1802	19	15	3-1	F5-1	1	120	1	8	10	77	4	6	3	8	749	ST-32 complex	
7	Specimen	1307	2,590,449	1918	1775	22	14	3-14	F5-5	185	18	0	7	5	1	13	36	53	15	213	ST-213 complex	8/1751
M15 240805	Isolate	315	2,178,381	1916	1764	22	14	3-14	F5-5	185	18	0	7	5	1	13	36	53	15	213	ST-213 complex	
8	Specimen	1062	2,469,552	1981	1838	19-1	15-11	3-423	F5-1	15	21	0	4	10	15	17	8	11	9	1049	ST-269 complex	2/1771
M15 240650	Isolate	420	2,187,886	1942	1787	19-1	15-11	3-423	F5-1	15	21	0	4	10	15	17	8	11	9	1049	ST-269 complex	
9	Specimen	1079	2,513,378	1965	1820	22	9	3-25	F4-1	13	17	0	4	10	34	5	38	11	9	1161	ST-269 complex	3/1780
M15 240240	Isolate	318	2,162,020	1956	1798	22	9	3-25	F4-1	13	17	0	4	10	34	5	38	11	9	1161	ST-269 complex	
10	Specimen	789	2,263,139	1900	1759	22	14	3-73	F5-9	700	20	0	1	5	13	53	26	41	3	162	ST-162 complex	3/1710
M15 240180	Isolate	305	2,110,985	1894	1741	22	14	3-73	F5-9	700	20	0	1	5	13	53	26	41	3	162	ST-162 complex	

Filled dark gray boxes represent missing typing data.

these discrepancies is unclear.

3.4. Potential utilisation of WGS for non-culture cases in England and Wales

The genome quality results produced from the specimen panel during this study were used to estimate the proportion of non-culture cases confirmed by the MRU that could be successfully sequenced using this technique. The results of this study suggest that ≥ 100 pg is required to produce a genome of sufficient quality to provide typing data and extended MLST analyses. To put this in context, the meningococcal DNA loads of hypothetical clinical specimens were estimated based on Ct values and specimen volumes (Supplementary Table 2). The estimated DNA loads indicate that samples with Ct values of > 30 contain less than 100 pg (at any volume) and would, therefore, be unlikely to produce acceptable genomic data. From January to December 2016, 54.5% of the non-culture specimens confirmed by the MRU produced a Ct value of > 30 . Conversely, samples with Ct ≤ 26 (23.4% of 2016 specimens) are predicted to produce acceptable genomes even if only 25 μ L of volume is available for extraction. For specimens with Ct values between 26 and 31, the volume of specimen available is likely to be the key determinant of successful WGS (Supplementary Table 2).

4. Discussion

The inability to perform standard WGS from non-culture specimens limits the extent of strain characterisation for around half of IMD cases in England and Wales. Genome sequencing from these samples would support the enhanced surveillance of vaccine antigens and investigations into potential vaccine failures. Recent studies have indicated that genomic data can be used to characterise antigenic promoter regions and, to some degree, predict the expression level of important vaccine antigens (Biagini et al., 2016). Genomic data have also been important in identifying newly-emerging invasive sub-lineages and differentiating potential outbreak strains (Chatt et al., 2014; Lucidarme et al., 2015; Tzeng et al., 2017). The MRU has been collecting and storing clinical specimens from confirmed IMD cases since 1996, and in recent years, the vast majority of residual non-culture material has been retained. The DNA within these specimens represents a genetic 'snap shot' mid-infection. The ability to perform WGS on these materials may provide opportunities to improve our understanding of meningococcal virulence as well as allowing us to paint a more accurate picture of the rapidly-changing epidemiological landscape.

In this study we have demonstrated the ability to enrich and sequence diverse group B meningococcal genomes directly from blood and CSF specimens using RNA baits to capture the meningococcal DNA fragments. Eight of the ten specimens tested produced draft genomes of acceptable genomic coverage and read depth. A lack of genomic data from the remaining two specimens was likely to be due to the low amount of meningococcal DNA within the specimens.

To assess the accuracy of this enrichment and sequencing technique, basic comparisons between the specimen genomes and the genomes of corresponding isolates (i.e. from the same patient) were performed. The genomes sequenced from purified isolate DNA contained fewer contigs than those from clinical specimens, and in most cases featured more annotated genes. Despite screening for human reads prior to assembly, the longer cumulative contig length among the specimen genomes suggests human sequences (or other contaminating DNA) were included in the genomic data. All of the commonly-used typing information (MLST, vaccine antigens, etc.) matched between the paired genomes, however, there were some allelic discrepancies amongst the indexed NEIS loci (as defined by PubMLST). Almost all of these are likely to be the result of PCR error and/or artefacts of read mis-assembly within paralogous genes. One non-paralogous locus, *rpsP* (NEIS0534), encoding a highly-conserved ribosomal protein, featured four nucleotide differences between the paired genomes. An explanation for these discrepancies is not apparent, however, it could possibly be indicative of contaminating DNA within the assembly. Other than these differences, all genome pairs were highly congruent demonstrating the accuracy of the direct WGS protocol. One limitation of these comparisons was the lack of standardisation of assembly/sequencing methods between the culture and specimen genomes. The influence of DNA target enrichment on sequence quality is, therefore, hard to assess accurately. Future studies controlling for such variables would provide a more detailed comparison of genome quality from the different DNA sources.

In the SureSelectXT system, RNA bait sequences are generated using genomic data from the target organism. *N. meningitidis* exhibits high homologous recombination rates, which may hinder the cross-reactivity of the RNA capture baits (Vos and Didelot, 2009). In this study, data from 2975 meningococcal genomes were used to generate the bait sequences. This panel was highly biased towards the major hyper-virulent lineages with $> 75\%$ of the strains belonging to one of only seven clonal complexes (data not shown). Whilst this does raise questions about the ability of the baits to enrich DNA belonging to less common invasive lineages and/or carriage strains, the Perl script used to generate the baits is designed to remove redundancy and prevent bias

towards predominant sequences. In this case, the baits appeared to be effective against diverse group B strains, however, the variation in DNA amount across the specimen panel made it difficult to identify any differences in the efficiency of the hybridisation between the different lineages. This represents another aspect of *N. meningitidis* DNA enrichment that requires further evaluation.

From these preliminary results we conclude that at least ~100 pg of meningococcal DNA is required to generate genomes of sufficient quality, whilst using > 500 pg in the library preparation results in a greater proportion of on-target reads and yields genomes with fewer interrupted/incomplete genes. Consequently, along with the Ct value, the volume of extract/specimen available could be a major factor in the likelihood of obtaining a genome of sufficient quality. The estimated meningococcal DNA concentrations within the clinical specimens were calculated using a standard curve of Ct values and quantified DNA concentrations of meningococcal isolate extracts (Clark et al., 2014). Whether Ct values from purified isolate extracts correlate with the Ct values from clinical specimens, which contain non-meningococcal DNA and possible PCR inhibitors, is unclear. Nonetheless, we believe these values are likely to provide a reasonable approximation.

A key disadvantage of the SureSelectXT Target Enrichment system is the significantly greater financial cost in relation to WGS from isolates. Whilst this system is likely to be more efficient than multiple nested PCR sequencing assays in terms of the number of loci characterised and the laboratory time required, the current reagent costs may be too high to justify routine WGS of all non-culture cases. Analysis of specific samples of interest (e.g. those from outbreaks or potential vaccine failures) would, however, be feasible and could provide pivotal information to guide public health responses.

It was estimated that up to 45% of recent non-culture specimens could be successfully genome sequenced using this technique, and ~25% of specimens should yield genomes with complete typing data. Further measures to optimise DNA extraction, deplete non-target DNA prior to enrichment and/or redesign the RNA bait library have previously been shown to improve the quality of the pathogen genome obtained (Brown et al., 2015). Future optimisation of the extraction/enrichment process may increase the proportion of non-culture IMD cases from which genomes can be sequenced.

5. Conclusions

The generation of draft meningococcal genomes directly from clinical specimens was achieved using the Agilent SureSelectXT kit. Estimations suggest that almost half of the non-culture IMD specimens received by the MRU may yield acceptable genomic data, however, further method optimisation may help to increase this proportion. The current high financial cost associated with the enrichment process may preclude routine WGS of all non-culture invasive strains.

Conflicts of interest

SAC has performed contract research on behalf of Public Health England for Pfizer and GlaxoSmithKline. JL has performed contract research on behalf of Public Health England for GlaxoSmithKline, Pfizer and Sanofi Pasteur. RB has performed contract research on behalf of Public Health England for Baxter, GlaxoSmithKline, Novartis Vaccines, Pfizer and Sanofi Pasteur. RD and JB have no interests to declare.

Funding

This study was supported by GlaxoSmithKline.

Acknowledgements

This publication made use of the *Neisseria* Multi Locus Sequence Typing website (<https://pubmlst.org/neisseria/>), developed by Keith

Jolley and sited at the University of Oxford (Jolley and Maiden, 2010), and the Meningitis Research Foundation Meningococcus Genome Library (<http://www.meningitis.org/research/genome>) developed by Public Health England, the Wellcome Trust Sanger Institute and the University of Oxford as a collaboration. The development of this site has been funded by the Wellcome Trust and European Union.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ijmm.2017.11.004>.

References

- Ali, A., Jafri, R.Z., Messonnier, N., Tevi-Benissan, C., Durrheim, D., Eskola, J., Fermon, F., Klugman, K.P., Ramsay, M., Sow, S., Zhujun, S., Bhutta, Z., Abramson, J., 2014. Global practices of meningococcal vaccine use and impact on invasive disease. *Pathog. Glob. Health* 108, 11–20.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477.
- Biagini, M., Spinsanti, M., De Angelis, G., Tomei, S., Ferlenghi, I., Scarselli, M., Rigat, F., Messuti, N., Biolchi, A., Muzzi, A., Anderloni, G., Brunelli, B., Cartocci, E., Buricchi, F., Tani, C., Stella, M., Moschioni, M., Del Tordello, E., Colaprico, A., Savino, S., Giuliani, M.M., Delany, I., Pizza, M., Costantino, P., Norais, N., Rappuoli, R., Masignani, V., 2016. Expression of factor H binding protein in meningococcal strains can vary at least 15-fold and is genetically determined. *Proc. Natl. Acad. Sci.* 113, 2714–2719.
- Bodi, K., Perera, A.G., Adams, P.S., Bintzler, D., Dewar, K., Grove, D.S., Kieleczawa, J., Lyons, R.H., Neubert, T.A., Noll, A.C., Singh, S., Steen, R., Zianni, M., 2013. Comparison of commercially available target enrichment methods for next-generation sequencing. *J. Biomol. Technol.* 24, 73–86.
- Brehony, C., Hill, D.M., Lucidarme, J., Borrow, R., Maiden, M.C., 2015. Meningococcal vaccine antigen diversity in global databases. *Eurosurveillance* 20, 1–9.
- Brown, A.C., Bryant, J.M., Einer-Jensen, K., Holdstock, J., Houniet, D.T., Chan, J.Z.M., Depledge, D.P., Nikolayevskyy, V., Broda, A., Stone, M.J., Christiansen, M.T., Williams, R., McAndrew, M.B., Tutill, H., Brown, J., Melzer, M., Rosmarin, C., McHugh, T.D., Shorten, R.J., Drobniowski, F., Speight, G., Breuer, J., 2015. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. *J. Clin. Microbiol.* 53, 2230–2237.
- Budroni, S., Siena, E., Dunning Hotopp, J.C., Seib, K.L., Serruto, D., Nofroni, C., Comanducci, M., Riley, D.R., Daugherty, S.C., Angiuoli, S.V., Covacci, A., Pizza, M., Rappuoli, R., Moxon, E.R., Tettelin, H., Medini, D., 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4494–4499.
- Chatt, C., Gajraj, R., Hawker, J., Neal, K., Tahir, M., Lawrence, M., Gray, S.J., Lucidarme, J., Carr, A.D., Clark, S.A., Fowler, T., 2014. Four-month outbreak of invasive meningococcal disease caused by a rare serogroup B strain, identified through the use of molecular PorA subtyping, England, 2013. *Eurosurveillance* 19.
- Christiansen, M.T., Brown, A.C., Kundu, S., Tutill, H.J., Williams, R., Brown, J.R., Holdstock, J., Holland, M.J., Stevenson, S., Dave, J., Tong, C.Y.W., Einer-Jensen, K., Depledge, D.P., Breuer, J., 2014. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. *BMC Infect. Dis.* 14.
- Clark, S.A., Lucidarme, J., Newbold, L.S., Borrow, R., 2014. Genotypic analysis of meningococcal factor H-binding protein from non-culture clinical specimens. *PLoS One* 9, e89921.
- Depledge, D.P., Palser, A.L., Watson, S.J., Lai, I.Y.C., Gray, E.R., Grant, P., Kanda, R.K., Leproust, E., Kellam, P., Breuer, J., 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One* 6, e27805.
- Gasparini, R., Amicizia, D., Domnich, A., Lai, P.L., Panatto, D., 2014. *Neisseria meningitidis* B vaccines: recent advances and possible immunization policies. *Exp. Rev. Vaccines* 13, 345–364.
- Heinsbroek, E., Ladhani, S., Gray, S., Guiver, M., Kaczmarski, E., Borrow, R., Ramsay, M., 2013. Added value of PCR-testing for confirmation of invasive meningococcal disease in England. *J. Infect.* 67, 385–390.
- Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* 11, 595.
- Joseph, B., Schneiker-Bekel, S., Schramm-Gluck, A., Blom, J., Claus, H., Linke, B., Schwarz, R.F., Becker, A., Goesmann, A., Froesch, M., Schoen, C., 2010. Comparative genome biology of a serogroup B carriage and disease strain supports a polygenic nature of meningococcal virulence. *J. Bacteriol.* 192, 5363–5377.
- Kozarewa, I., Armisen, J., Gardner, A.F., Slatko, B.E., Hendrickson, C.L., 2015. Overview of target enrichment strategies. *Curr. Protoc. Mol. Biol.* 112, 7.21–23.
- Kretz, C.B., Retchless, A.C., Sidikou, F., Issaka, B., Ousmane, S., Schwartz, S., Tate, A.H., Pana, A., Njanpop-Lafourcade, B.M., Nzeyimana, I., Obama Nse, R., Deghmane, A.E., Hong, E., Brynildsrud, O.B., Novak, R.T., Meyer, S.A., Oukem-Boyer, O.O.M., Ronveaux, O., Caugant, D.A., Taha, M.K., Wang, X., 2016. Whole-genome characterization of epidemic *Neisseria meningitidis* serogroup C and resurgence of serogroup W, Niger, 2015. *Emerg. Infect. Dis.* 22, 1762–1768.

- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr doi:arXiv:1303.3997 [q-bio.GN].
- Lucidarme, J., Hill, D.M.C., Bratcher, H.B., Gray, S.J., du Plessis, M., Tsang, R.S.W., Vazquez, J.A., Taha, M.K., Ceyhan, M., Efron, A.M., Gorla, M.C., Findlow, J., Jolley, K.A., Maiden, M.C.J., Borrow, R., 2015. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. *J. Infect.* 71, 544–552.
- Magoč, T., Salzberg, S.L., 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132.
- Parikh, S.R., Newbold, L., Slater, S., Stella, M., Moschioni, M., Lucidarme, J., De Paola, R., Giuliani, M., Serino, L., Gray, S.J., Clark, S.A., Findlow, J., Pizza, M., Ramsay, M.E., Ladhani, S.N., Borrow, R., 2017. Meningococcal serogroup B strain coverage of the multicomponent 4CMenB vaccine with corresponding regional distribution and clinical characteristics in England, Wales, and Northern Ireland, 2007–08 and 2014–15: a qualitative and quantitative assessment. *Lancet Infect. Dis.* 17, 754–762.
- Pelton, S.I., 2016. The global evolution of meningococcal epidemiology following the introduction of meningococcal vaccines. *J. Adolesc. Heal.* 59, S3–S11.
- Schoen, C., Blom, J., Claus, H., Schramm-Gluck, A., Brandt, P., Müller, T., Goesmann, A., Joseph, B., Konietzny, S., Kurzai, O., Schmitt, C., Friedrich, T., Linke, B., Vogel, U., Frosch, M., 2008. Whole genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3473–3478.
- Song, L., Florea, L., Langmead, B., 2014. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol.* 15, 509.
- Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., Deboy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citti, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., Venter, J.C., 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287, 1809–1815.
- Tzeng, Y.-L., Bazan, J.A., Turner, A.N., Wang, X., Retchless, A.C., Read, T.D., Toh, E., Nelson, D.E., Del Rio, C., Stephens, D.S., 2017. Emergence of a new *Neisseria meningitidis* clonal complex 11 lineage 11.2 clade as an effective urogenital pathogen. *Proc. Natl. Acad. Sci.* 114, 4237–4242.
- Vos, M., Didelot, X., 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* 3, 199–208.
- Watkins, E.R., Maiden, M.C.J., 2017. Metabolic shift in the emergence of hyperinvasive pandemic meningococcal lineages. *Sci. Rep.* 7, 41126.