



This is a repository copy of *Meningococcal core and accessory phasomes vary by clonal complex*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/162041/>

Version: Published Version

Article:

Wanford, J.J., Holmes, J.C., Bayliss, C.D. et al. (1 more author) (2020) Meningococcal core and accessory phasomes vary by clonal complex. *Microbial Genomics*, 6 (5). 367.

<https://doi.org/10.1099/mgen.0.000367>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Meningococcal core and accessory phasomes vary by clonal complex

Joseph J. Wanford¹, Jonathan C. Holmes¹, Christopher D. Bayliss¹ and Luke R. Green^{2,*}

Abstract

Neisseria meningitidis is a Gram-negative human commensal pathogen, with extensive phenotypic plasticity afforded by phase-variable (PV) gene expression. Phase variation is a stochastic switch in gene expression from an ON to an OFF state, mediated by localized hypermutation of simple sequence repeats (SSRs). Circulating *N. meningitidis* clones vary in propensity to cause disease, with some clonal complexes (ccs) classified as hypervirulent and others as carriage-associated. We examined the PV gene repertoires, or phasome, of these lineages in order to determine whether phase variation contributes to disease propensity. We analysed 3328 genomes representative of nine circulating meningococcal ccs with PhasomeIt, a tool that identifies PV genes by the presence of SSRs and homologous gene clusters. The presence, absence and functions of all identified PV gene clusters were confirmed by annotation or BLAST searches within the *Neisseria* PubMLST database. While no significant differences were detected in the number of PV genes or the core, conserved phasome content between hypervirulent and carriage lineages, individual ccs exhibited major variations in PV gene numbers. Phylogenetic clusters produced by phasome or core genome analyses were similar, indicating co-evolution of PV genes with the core genome. While conservation of PV clusters is high, with 76% present in all meningococcal isolates, maintenance of an SSR is variable, ranging from conserved in all isolates to present only in a single cc, indicating differing evolutionary trajectories for each lineage. Diverse functional groups of PV genes were present across the meningococcal lineages; however, the majority directly or indirectly influence bacterial surface antigens and could impact on future vaccine development. Finally, we observe that meningococci have open pan phasomes, indicating ongoing evolution of PV gene content and a significant potential for adaptive changes in this clinically relevant genus.

DATA SUMMARY

PhasomeIt analyses are included as a supplementary dataset on Zenodo (DOI: 10.5281/zenodo.3545849) (https://zenodo.org/record/3545849#.XoxM_HJKiM8). An Excel file is included [MGL Library (11.12.2018).xlsx] that describes the available isolates when the Meningococcal Genome Library (MGL) was interrogated. Further worksheets are provided to demonstrate which isolates were included for each clonal complex (cc) and the random choice of isolates. Compressed files of each PhasomeIt analysis grouped by cc are included, within these folders are three .html files; index.html, which summarizes all of the PhasomeIt analysis; core_phasome.html, describes the core phasome for the cc; tractLengths.html, describing each phase-variable gene tract length. Three folders are included that make up the data files for the described .html files.

Supplementary Material is available with the online version of this article.

INTRODUCTION

Neisseria meningitidis (the meningococcus) is a Gram-negative, obligate human commensal and pathogen. Normally, meningococci are found as asymptomatic colonizers of the human nasopharynx, but they can occasionally cross the mucosal barrier to cause septicaemia and meningitis [1]. However, the switch from carriage to an invasive state remains poorly understood, and whilst genomic correlates can be found with disease-causing potential, these do not explain the dual behaviour of this pathogen. One potential determinant of the disease to carriage switch is localized hypermutation and the consequent phase-variable (PV) expression of key surface molecules.

Received 11 December 2019; Accepted 27 March 2020; Published 29 April 2020

Author affiliations: ¹Department of Genetics and Genome Biology, University of Leicester, Leicester, UK; ²Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK.

***Correspondence:** Luke R. Green, l.r.green@sheffield.ac.uk

Keywords: clonal complex; carriage; hypervirulence; microevolution; *Neisseria*; phase variation.

Abbreviations: cc, clonal complex; IMD, invasive meningococcal disease; LOS, lipo-oligosaccharide; MGL, Meningococcal Genome Library; MLST, multilocus sequence typing; MRC, Medical Research Council; MRF, Meningitis Research Foundation; OMP, outer-membrane protein; PV, phase variable; SSR, simple sequence repeat; ST, sequence type; WGS, whole-genome sequence.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table and seven supplementary figures are available with the online version of this article.

000367 © 2020 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Phase variation is a high-frequency switch in gene expression, often conferred by localized hypermutation at simple sequence repeats (SSRs) in the genome [2]. Addition or deletion of repeats by slippage of the replicative polymerase at these loci can confer changes in gene expression from an ON to an OFF state through inducing frameshifts (if the SSR is within an ORF) [2], or can alter levels of gene expression between low, intermediate and high states (if the SSR is in a promoter or regulatory region) [3]. Importantly, PV expansion/retraction of SSRs occurs at a frequency vastly greater ($>1 \times 10^{-5}$ mutations per division) than macro-evolutionary events in the genome, such as SNPs or movement of transposable elements [4, 5]. Characterization of SSRs in genome sequences allows assessment of the propensity of a particular clone to vary surface phenotypes by phase variation. Furthermore, the protein products of many PV genes are outer-membrane proteins (OMPs), and selection for ON/OFF phase variation in these loci are often driven by recurrent host selection through bactericidal antibodies [6, 7]. Initial studies of meningococcal genome sequences identified ~50 PV genes with the potential to generate large amounts of genetic and phenotypic heterogeneity in otherwise isogenic bacterial populations [8]. Recently Aidley *et al.* [9] developed *PhasomeIt*, a new program for analysing whole-genome sequences (WGSs) for the presence of PV genes subject to SSR-mediated switches in gene expression. Wanford *et al.* utilized this program to analyse SSR-PV genes across the genus *Neisseria* and observed that there were significant differences in the number of PV genes between *Neisseria* spp., with the majority of commensal species having low numbers of PV genes, whilst *Neisseria lactamica* and the pathogenic *Neisseria* had ~40–60 PV genes per genome [10]. An intra-species analysis of phasome content across clonal complexes (ccs), which could specifically identify differential PV content across distinct lineages of meningococci known to differ in their disease-causing potential, has yet to be undertaken.

Protection against invasive meningococcal disease (IMD) is achieved through multiple vaccines, both polysaccharide [11, 12] or protein-based [13], which elicit immunity against a subset of serogroups and lineages. Despite the efficacy of these vaccines, the circulation of non-vaccine-covered strains, serotype replacement and mutation-mediated vaccine escape remain a concern for public health [14, 15]. Vaccine design and clinical epidemiology have benefited from routine whole-genome sequencing of meningococcal disease isolates by allowing for detailed interrogation of the genomic population structure of circulating clones [16, 17]. The deposition of this genome data into the *Neisseria* PubMLST database [18] has provided an excellent resource for both clinical epidemiologists, and research into disease-associated genetic determinants [19]. Currently in the UK, most cases of disease are due to five major circulating ccs: cc11, cc213, cc269, cc41/44 and cc32. These lineages are a subset of the hypervirulent lineages that tend to have high disease to carriage ratios. Other ccs, such as cc22, cc23, cc162 and cc461, are infrequent causes of IMD and have low disease to carriage ratios; hence, they are referred to as carriage-associated. These ccs collectively

Impact Statement

Phase variation is a high frequency mutational mechanism by which bacteria switch expression of a gene ON or OFF, which allows them to rapidly adapt to new environments, such as upon transmission to a new host. The ability to phase vary genes can be detected from genome sequences by identifying the presence of simple sequence repeats in the ORF. In the UK, genomes from clinical isolates of the major human pathogen *Neisseria meningitidis* are routinely sequenced and deposited in the open PubMLST database. In this study, we performed a comparative analysis of the presence of phase-variable (PV) genes in 3328 genomes encompassing major carriage-associated and hypervirulent meningococcal clones. Whilst we detected no discernible difference in the number or functions of PV genes between carriage and hypervirulent clones, we found divergent repertoires of PV genes between related groups that may shed light on the selection pressures driving continual evolution of these clinically relevant clonal complexes. We anticipate our data will form a framework for understanding the implications of phase variation on carriage dynamics, and transmission throughout the population.

account for 87% of strains within the Meningitis Research Foundation (MRF) Meningococcal Genome Library (MGL) in *Neisseria* PubMLST. The WGS data deposited within the MGL provides an opportunity to interrogate whether phase variation influences how different meningococcal clones colonize their hosts and cause disease.

In this study, we have defined the collection of PV genes, the phasome, across nine currently circulating meningococcal ccs in the UK, using the novel tool *PhasomeIt* [9] to provide a conceptual framework for understanding colonization dynamics and pathogenesis. We hypothesize that differences in the phasomes between major ccs of meningococci have arisen due to specific variations in the within-host selection pressures experienced by these clones, such as the immune response to surface OMPs, and that these differences may have implications for rational vaccine design and evaluation.

METHODS

Genomic dataset

Genome contigs were extracted from the MRF MGL available through *Neisseria* PubMLST [18] in December 2018. From among the 3568 WGSs available, we selected the ccs represented by $>1.5\%$ of isolates (i.e. >54) for analysis (Supplementary Material, available with the online version of this article). This dataset covered the following ccs: cc11 ($n=815$), cc162 ($n=60$), cc213 ($n=224$), cc22 ($n=79$), cc23 ($n=499$), cc269 ($n=529$), cc32 ($n=165$), cc41/44 ($n=669$) and cc461 ($n=75$). A subset of 10.1% of the isolates did not have a defined cc, these

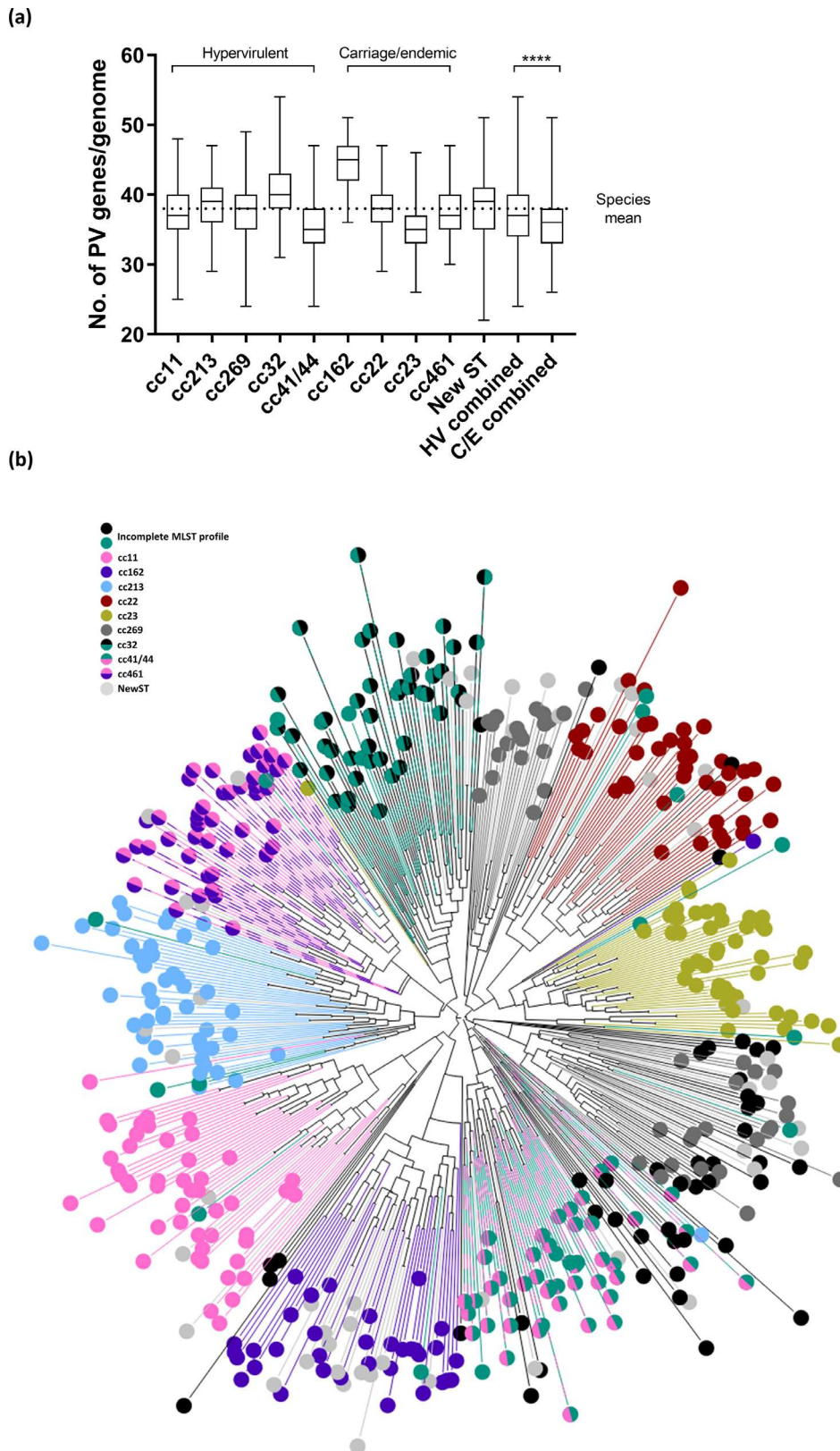


Fig. 1. Clustering of PV genes in meningococcal ccs. (a) Box plot depicting the range, median, and first and third quartiles of the number of PV genes detected for each cc. Statistical significance was analysed by two-way ANOVA. ****, $P < 0.0005$. The mean number of PV genes across all ccs is depicted with a dotted line. C/E, Carriage/endemic; HV, hypervirulent. (b) Neighbour-joining tree based on the presence or absence of PV homology groups across all ccs. Colours corresponding to each cc are shown in the key.

genomes were analysed using the PubMLST tool 'Search by combination of alleles' to identify partial multilocus sequence typing (MLST) profiles allowing for assignment to relevant ccs. The remaining 213 isolates were classified as 'new ST' (new sequence type). Contig files were converted to annotated GenBank files using PROKKA (v. 1.11) [20] with MC58 as a reference sequence. All genome sequences and their respective metadata are shown in the Supplementary Material. All PhasomeIt runs included a number of reference strains, such that loci could be rapidly identified by their reference strain number in the PubMLST allele database. These strains were: MC58, H44/76, NZ-05/33, α 710, FAM18, M25419, DE10444, M01-240355 and M04-240196.

Extraction of PV gene repertoire using PhasomeIt

PhasomeIt was run as described elsewhere [4, 5, 10, 21, 22] for each cc individually (containing 75–815 genomes per cc). An additional run was performed combining a random selection of 50 genomes from each cc. The cut-offs for the repeat numbers for each repeat tract type were as follows: 9 for poly-G repeat tracts, 11 for poly-T tracts, 8 for dinucleotide repeat tracts, 6 for trinucleotide tracts, 5 for tetranucleotide and pentanucleotide tracts, and 3 for hexanucleotide to nonanucleotide tracts. These cut-offs were developed based on available data highlighting the propensity of these SSRs to undergo indels for *Neisseria* [4, 5, 10, 21, 22]. The BioPython module Bio.Phylo was used to reconstruct neighbour-joining trees indicating the separation of isolates using a distance matrix that was derived from the Manhattan distance metric from binary lists of presence/absence of gene groups.

Downstream core-phasome analysis using PhasomeIt data

Core phasomes were derived for each cc individually using PhasomeIt. PV genes were considered core if they were present and contained an SSR in more than 50% of isolates. To facilitate comparison across different cc datasets, PhasomeIt functional groups were assigned NEIS numbers by querying the reference strain gene numbers against the allele definitions provided by the *Neisseria* PubMLST database; thus, grouping together all alleles of a given loci. Where a gene identified in the analysis was not present in any of the reference strains, the first identified sequence of this loci in PhasomeIt was directly queried against the PubMLST allele database. Core-phasome genes for each cc were subsequently cross-checked across all other cc datasets, to identify the percentage PV conservation of these loci. This comparative approach was required, as if a locus is PV in one cc but not in another, then it is considered part of the accessory phasome. The analysis pipeline used in this work is outlined in Fig. S1.

Presence/absence gene analysis

PhasomeIt will only identify a locus if a PV tract is present within one of the isolates from a cc [9, 10]. As each cc was analysed individually, it is possible that non-PV loci were not identified during the initial inquiry. Therefore, the compiled list of NEIS numbers corresponding to all core and accessory

PV genes were queried against all of the strains previously analysed using the 'Gene Presence' function in PubMLST. Gene Presence utilizes a BLASTN query (v1.4.16) [23] with 70% identity and 50% coverage to produce a '% present' matrix for all genes across each cc. To further confirm the presence or absence of loci of interest, reference sequences of each loci were queried using BLASTN with default parameters against a 50 genome subset from each cc within the PubMLST database. The two matrices, (i) % present in the phasome taken from PhasomeIt, and (ii) % present across the cc taken from the Gene Presence analysis, formed the basis of subsequent analysis.

Additional routine bioinformatic analysis

To confirm the BLAST-based sequence alignments employed by PhasomeIt were identifying multiple alleles of the same gene, and not loci with substantial regions of homology (e.g. comparison of *porA* and *porB* sequences), we interrogated a number of loci using the ClustalOmega sequence alignment tool (v1.2.4) [24]. Loci of interest were extracted from a subset of 25 genomes and aligned using ClustalOmega with default parameters. Alignments were analysed to confirm the repeat number described by PhasomeIt. Localization of SSRs and their context within the genome were visually interrogated using Artemis genome visualization software (v17.0.1) [25].

Statistical analysis

Statistical analysis of the numbers of PV genes in each cc was performed using a two-way ANOVA in GraphPad Prism software.

RESULTS

The genus *Neisseria* exhibits significant variation in the numbers of PV genes, for example, *Neisseria flavescens* has three PV genes per genome, while meningococci have a mean of 39 per genome [10]. This species-to-species variation indicates that the number of PV genes may be influenced by niche specificity or alternate biological behaviours. As meningococcal ccs differ in their propensity to persist in carriers and to cause disease, we speculated that PV repertoires may vary between ccs. In order to test this hypothesis, we examined the PV signatures of nine ccs that were split into hypervirulent (cc11, cc41/44, cc213, cc32, cc269) [26] or carriage-associated/endemic (cc22, cc461, cc23, cc162) lineages [27]. The non-groupable isolates, not associated with a cc and identified as a new ST, were excluded from this classification system.

Detection and clustering of PV genes in different meningococcal ccs

PhasomeIt identifies PV genes in genome sequences based on the presence or absence of an SSR. The PV gene content was delineated for each of the major UK circulating ccs and a set of non-groupable strains (the new ST group) by grouping SSR-containing genes and non-PV homologues based on a 50% sequence homology/

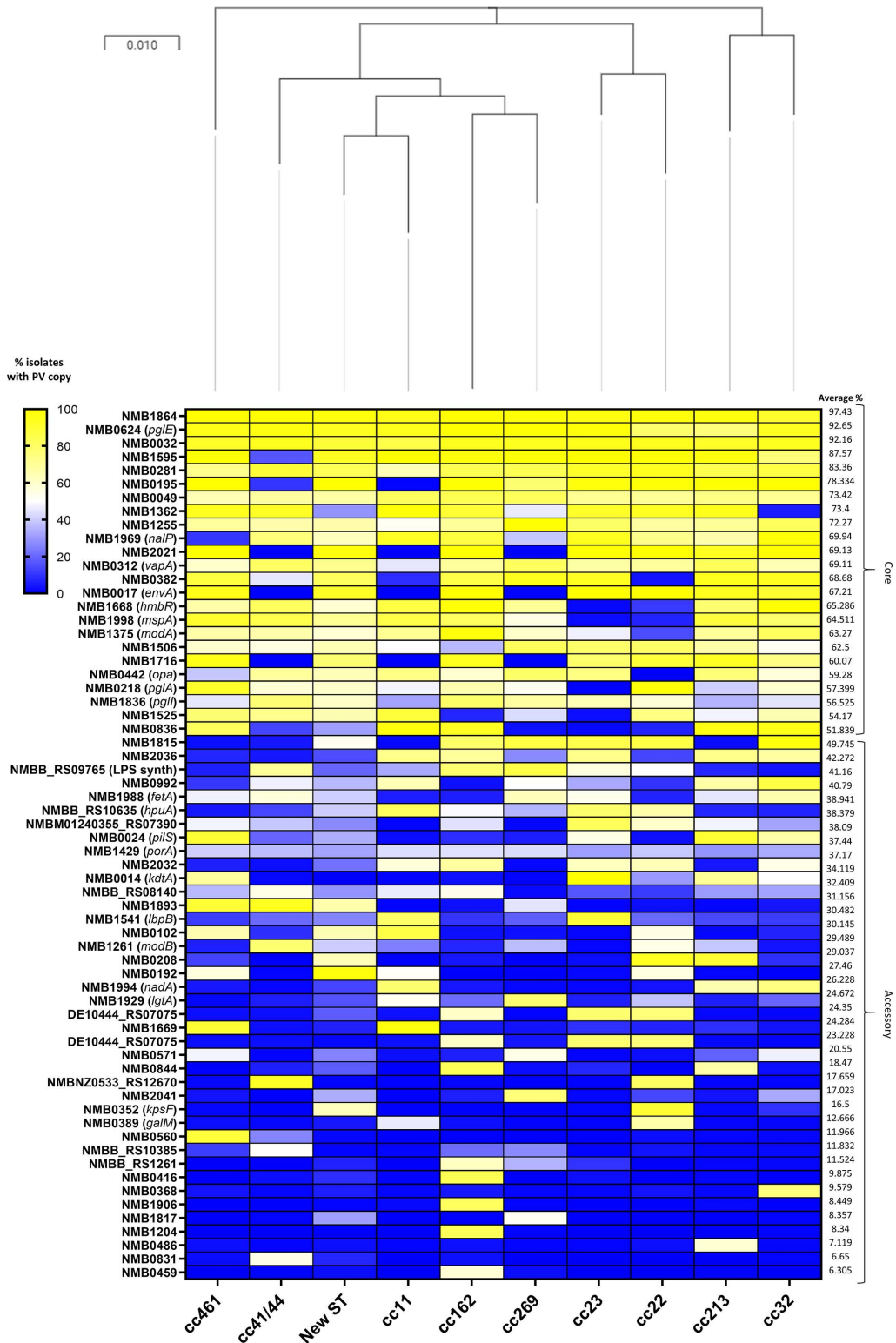


Fig. 2. Presence of PV genes across each cc. Core-phasome data were used to derive the per cent presence of PV genes above a 50% threshold for all isolates. Where genes were core in one cc, but not in another, the respective loci were queried against the 'non-core' cc dataset to provide the per cent conservation below the 50% core-phasome cut-off. Genes are sorted from top to bottom, on the highest mean conservation of phase variation across all ccs. The colour gradient represents the percentage of isolates with a PV copy of the respective gene ranging from blue (low conservation) to yellow (high conservation). To give an overview of population structure, an MLST tree was included based on the PubMLST typing scheme for meningococcus using one representative isolate from each cc. The MLST tree was drawn using iTOL software, through the PubMLST interface where the scale bar denotes single nucleotide polymorphisms.

coverage cut-off. We observed that each cc contained between 22 and 55 PV genes (Fig. 1a). The mean number of PV genes per genome was significantly lower in typically carriage-associated ccs compared to hypervirulent ccs (35 and 37 PV genes, respectively; $P < 0.0001$, two-way ANOVA). Of the hypervirulent ccs, cc32 encoded the largest number of PV genes (40.5 ± 4.1) when compared to cc11 (37.6 ± 3.8), cc41/44 (35.6 ± 3.9), cc213 (38.3 ± 3.8) and cc269 (37.6 ± 3.8). Of the typically carriage/endemic ccs, cc162 encoded the greatest mean number of PV genes (44.4 ± 3.4) compared with cc22 (37.9 ± 3.5), cc461 (37.7 ± 3.5) and cc23 (34.7 ± 3.1). Whilst strains designated new ST displayed an analogous number of PV genes (40) to other ccs, they had the largest range of detected PV genes (22–50). Statistical comparisons of the number of PV genes in each cc is detailed in the Table S1.

Phylogenetic relationships strongly influence gene content. To determine the similarities between the phasomes of the ccs, a neighbour-joining phylogenetic tree was derived based on the presence or absence of PV genes. ccs form discreet clusters that are delineated by specific PV gene content (Fig. 1b). Isolates of the new ST group exhibited ‘mixed clustering’ with respect to their PV gene content, suggesting that these isolates are closely related to the known lineages, but the association is masked by divergent or incomplete MLST profiles. In the majority of cases, clustering by phasome reflected the relatedness of meningococcal isolates by MLST and whole-genome phylogenetic analysis. For example, cc162 and cc269 share a root on the phasome neighbour-joining tree, and this mimics the relatedness of these isolates by MLST [28], in which these isolates are more closely related to each other than to other ccs. In contrast, cc213 and cc32 appear highly related on a phylogenetic tree based on the core genome (Fig. S2), but are separated on the phasome tree. These data indicate that PV gene repertoire mimics genomic phylogeny, but that there is a mobile gene pool with the potential to move PV genes between clonally unrelated ccs.

Repetitive DNA can cause problems in genome assembly, such as contig breaks, leading to highly fragmented genome assemblies [29]. Identification of functional gene groups by PhasomeIt relies on the ability to assign an SSR to an ORF [9]. To infer whether the differences detected in PV gene content between ccs was due to actual gene content or due to missing loci in the genome assembly, we analysed the correlation between the number of contigs in each cc data set (Fig. S3a) versus the number of PV genes detected by PhasomeIt (Fig. S3b). In agreement with previous work [10], we detected a negative correlation whereby the number of PV genes decreased as the number of contigs increased, which was significant for every cc except cc162 (Fig. S3b). This negative correlation is probably due to a small number of PV genes with longer repeat tracts that cause contig breaks and, hence, they are not detected by PhasomeIt. In general, the low correlation suggests there are minimal undetectable repeat tracts; therefore, they do not affect the conclusions of this work.

ccs contain a core phasome alongside a distinctive accessory phasome

To determine the specific differences in PV gene content driving the PhasomeIt clustering of ccs, we performed a comparative analysis of the percentage of isolates in each cc that contained specific ‘core phasome’ PV genes (Fig. 2). We found that 24 PV gene families were present in greater than 50% of genomes across all ccs, these genes are representative of the species core phasome as reported by Wanford et al. [10]. However, only three genes (NMB1864, *pglE* and NMB0032) were present in the core phasomes of all ccs (Fig. 2). Presence of a gene in the core phasome of a cc tended to show a binary pattern, whereby a gene was either present or absent from all isolates with very few genes showing an intermediate presence in the phasome. An exception was NMB1255, which had a 50% presence in the cc11 phasome. The core phasomes of each cc showed a variable amount of overlap, with 43 genes in three or more ccs, 7 genes in two, and 13 genes only in a single cc (e.g. NMB0416 is present in 100% of cc162 isolates, but not in any isolates of the other ccs). Lastly, the presence of specific genes in the phasome did not necessarily reflect the relatedness of isolates by MLST analysis. For example, cc213 and cc32 share a common root on the MLST tree, but have different core phasomes with 80–100% variation in the presence of several genes (Fig. 2).

PhasomeIt analysis includes non-PV homologues of PV genes identified through a networking approach (see the work by Aidley et al. [9]). However, PhasomeIt may not identify a specific PV gene in a cc if there are no PV homologues of that gene in any of the strains of that cc. To circumvent this issue, we utilized core-phasome locus tags (i.e. NEIS numbers) to query the PubMLST database and generate a presence/absence analysis for each cc irrespective of whether that gene contained an SSR (see the heat map in Fig. S4). This analysis indicated that despite sparse conservation of genes in the phasome (Fig. 2), 48 out of the 63 core-phasome genes were present in all isolates of all nine ccs. This indicates that there is variable conservation between ccs of the SSR within these 48 genes and not loss of these genes.

To examine the distribution of PV genes in more detail, we coded the PubMLST presence/absence data, and the per cent presence PhasomeIt data (Fig. 3) to produce a binary matrix, outlining whether a gene was present in the core phasome (>50% PubMLST; >50% PhasomeIt), present but without an SSR (>50% present in PubMLST; <50% PhasomeIt) or not present (<50% PubMLST; <50% PhasomeIt), as a proxy for the origin of PV loci in these ccs. Interestingly, the majority of gene groupings across all ccs were not present in the core phasome due to lack of presence of an SSR, with the remaining minority non-core-phasome genes lacking the gene altogether. Detailed analysis of conservation of specific functional groupings of genes will be described below.

Restriction modification systems

Consistent with previous work [30], our analysis identified PV copies of the three restriction modification systems encoded

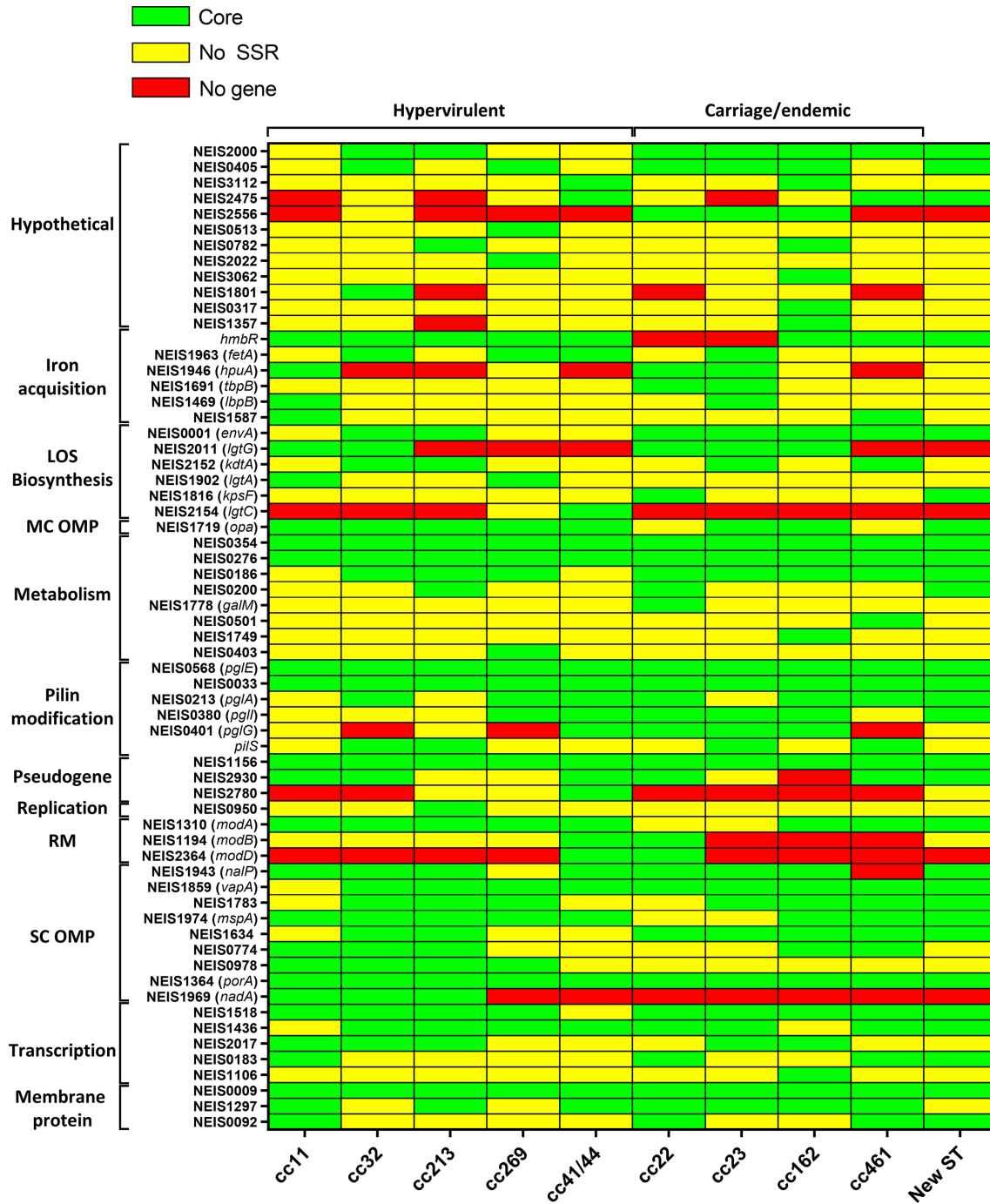


Fig. 3. Presence of modular PV genes and their cognate SSRs in hypervirulent and carriage-associated meningococci. The heatmap demonstrates whether genes were in the core phasome. Bars are depicted as follows: core – present in greater than 50% of isolates and phase variation in greater than 50% of isolates (green); no gene – absent from the core because of lack of the loci in 50% of isolates (red); or no SSR – presence of the gene but lacked an SSR in 50% of the isolates (yellow). Strains are grouped based on their predicted virulence (hypervirulent vs carriage), genes are arranged into modules by their known or predicted function. LOS, Lipo-oligosaccharide; MC, multi-copy; OMP, outer-membrane protein; RM, restriction modification; SC, single-copy.

by the *modA*, *modB* and *modD* loci. The PV *mod* alleles were detected by the presence of five or more repeats and were associated with GCCA, CCCAA and ACCGA repeats (Table 1). *modA* was highly conserved, being present in all ccs, and was subject to phase variation in all ccs except cc22 and cc23 (Fig. 3), which contained an SSR in less than 50% of isolates. *modB* was less conserved, being absent from cc461, cc162 and cc23 (Fig. 3). Where present, *modB* was subject to phase variation in only cc41/44 and cc22. Finally, *modD* was only present in cc41/44 and cc22, and was also PV in these ccs (Fig. 3).

Pilin glycosylation

Four pilin glycosylation loci, *pglA*, *pglI*, *pglE* and *pglG* [31], were identified as being associated with G, G, CAACAAA and C repeats, respectively (Table 1). The most conserved phase-variation-associated *pgl* gene was *pglE*, which was both present and highly prevalent as a PV locus across all ccs (92% phase-variation conservation across all genomes). Similarly, *pglI* and *pglA* were present in >50% of isolates across all ccs, but were not subject to phase variation in three and four ccs, respectively (*pglA* – cc11, cc23, cc213; *pglI* – cc461, cc11, cc213, cc32). *pglG* was absent from three ccs (cc461, cc269 and cc32) and only subject to phase variation in cc41/44, cc162, cc23 and cc22 (Fig. 3). Of note, we did not identify *pglH* in our phase-variation analysis, indicating that any repeat tract present in this gene either fell below our SSR cut-off or was missed in the fragmented assembly (see below).

Lipo-oligosaccharide (LOS) biosynthesis

PV LOS modifying genes, *lgtG*, *lgtA* and *lgtC* [32], were identified in association with tracts of nine or more G, G and C repeats, respectively (Table 1). *lgtA* was the only one of these three genes to be present in all ccs, but was only PV in two hypervirulent clones, cc11 and cc269 (Fig. 3). *lgtC* was not present in any of the carriage/endemic-associated ccs but was present in two hypervirulent clones, cc269 and cc41/44, and only PV in the latter cc. *lgtG* was not present in the carriage-associated ccs cc461 and the new ST, nor in the hypervirulent clones cc213, cc269 and cc41/44 (Fig. 3). In the remaining ccs, this gene was present and subject to phase variation. For some isolates, in the gene grouping encompassing *lgtG*, we observed a second homologous loci, although none of these were subject to phase variation. This can be due to other genes in the locus with high levels of homology to *lgtG*. We found no evidence of grouping of loci for the other *lgt* genes. In addition, we identified three novel PV loci with roles in synthesis/modification of the LOS. *kdtA* is responsible for addition of a Kdo moiety to lipid A in the early stages of LOS biosynthesis [33]. This gene was present in all ccs, and contained an in-frame PV poly-G SSR at the 3' end of the gene, with a tract length range of 9–10 repeats (Table 1) in two hypervirulent ccs (cc32 and cc213) and two carriage-associated clones (cc23 and cc461).

We also observed that *envA* can be subject to phase variation due to a GC repeat tract. This SSR was present in all carriage-associated ccs, but only in cc32 and cc213

hypervirulent clones (Fig. 3). This gene encodes an UDP-3-O-acetyl N-acetylglucosamine deacetylase that is putatively involved in LOS biosynthesis (Table 1) [34]. Another ubiquitously distributed gene, *kpsF*, encoding a rabinose-5-phosphate isomerase [35], was also found to contain a GC tract above the eight repeat unit cut-off. A PV version of *kpsF* was present in some isolates of cc22, and the new ST isolates (Fig. 3). A key future test is to demonstrate phase variation of one of these two genes, as dinucleotide tracts are not known to mediate phase variation in *Neisseria*.

Iron acquisition

We identified the three known PV iron-acquisition systems, *hmbR*, *fetA* and *hpuA* [36], which were associated with polyG, C and G repeats (Table 1), respectively. A PV copy of *hmbR* was absent from the carriage-associated cc22 and cc23 lineages, but present in >50% of isolates across all other ccs (Fig. 3). The *fetA* gene was present in all ccs; however, only showed >50% phase variation within cc32, cc269, cc41/44 and the carriage-associated cc23 (Fig. 3). Similarly, the *hpuA* gene was present within two hypervirulent lineages (cc11 and cc269) and three carriage-associated lineages (cc162, cc22 and cc23), but only generally PV within cc11, cc22 and cc23 (Fig. 3).

In an extension of this module, SSR tracts were identified in three iron-acquisition genes not typically considered as subject to phase variation; *tbpB* (encoding transferrin-binding protein) [37], *lbpB* (encoding lactoferrin-binding protein) [38] and *pigA* (a *Pseudomonas* homologue involved in haem acquisition) [39] (Table 1). These genes were associated with ATAACAAA, AAGCTG and GAAGCC repeats with phase-variation thresholds of six repeats. The former of these tracts would be subject to ON/OFF phase variation, whereas the latter two would be subject to in-frame phase variation. All ccs encoded *tbpB*, but only the carriage-associated cc22 and cc23 lineages encoded PV copies in >50% of isolates (Fig. 3). Similarly, *lbpB* and NEIS1587 (*pigA*) were ubiquitous, but only contained PV SSRs in hypervirulent cc11 and the carriage-associated cc23 isolates (Fig. 3).

Other OMPs

Our results also confirm and extend observations on the distributions of several other PV single copy OMPs [40–45]. Both the *porA* and *nadA* genes were always present as PV loci, with the former present in all ccs (Fig. 3), and the latter only in the cc11, cc32 and cc213 hypervirulent lineages. Analyses of the two PV autotransporters indicated that *mspA* was universally present in all lineages, but only contained SSRs in hypervirulent lineages and two carriage-associated lineages, cc162 and cc461 (Fig. 3). Similarly, *nalP* was encoded by all hypervirulent and carriage-associated ccs except cc461, and was PV in all except cc269 (Fig. 3).

Due to the homology cut-off employed for generating homology groups in PhasomeIt, we were unable to separate highly homologous, multi-copy genes into individual loci as expected for the *pilC1/pilC2* and the four *opa* loci. Analysis

Table 1. Functional analysis of core and accessory PV genes of meningococcus

Functions of genes were either discerned from published literature where available or from predicted functional domains using a BLASTP search. Predominant SSRs are those that make up more than 50% of the loci across the dataset. Evidence for phase variation was split into the following: known, functional studies in the literature; alignment, SSR present and differential tract length observed at homologous loci; and SSR present, SSR identified in the ORF but no alignment-based evidence for phase variation.

NEIS number (gene)	Function/functional domain	Predominant SSR	ccs gene conserved and subject to phase variation	Evidence for variation
LOS biosynthesis				
NEIS0001 (<i>envA</i>)	UDP-3-O-acyl N-acetylglucosamine deacetylase	GC	32, 213, 22, 23, 162, 461, new ST	SSR present
NEIS2011 (<i>lgtG</i>)	LOS glucosyltransferase	C	11, 32, 22, 23, 162	Known [6]
NEIS2152 (<i>kdtA</i>)	Addition of Kdo moiety to lipid A	G	32, 213, 23, 461	SSR present
NEIS1902 (<i>lgtA</i>)	N-Acetylglucosaminyltransferase	C	11, 269	Known [59]
NEIS1816 (<i>kpsF</i>)	Rabinose-5-phosphate isomerase	GC	22, new ST	SSR present
NEIS2154 (<i>lgtC</i>)	Galactosyltransferase	G	41/44	Known [59]
Pilin modification				
NEIS0568 (<i>pglE</i>)	Glycosylation of pilus	CAACAAA	All	Known [60]
NEIS0033 (<i>pilC2</i>)	Retraction of pilus	C	All	Known [61]
NEIS0213 (<i>pglA</i>)	Glycosylation of pilus	G	32, 269, 41/44, 22, 162, 461, new ST	Known [60]
NEIS0380 (<i>pglI</i>)	Glycosylation of pilus	G	269, 41/44, 22, 162, new ST	Known [60]
NEIS0401 (<i>pglG</i>)	Glycosylation of pilus	C	41/44, 22, 23, 162	Known [60]
<i>pilS</i>	Silent loci which recombines with <i>pilE</i>	GCCGAC	32, 213, 23, 461	SSR present
Restriction modification				
NEIS1310 (<i>modA</i>)	Methylation and foreign DNA restriction	GCCA	11, 32, 213, 269, 41/44, 162, 461, new ST	Known [30]
NEIS1194 (<i>modB</i>)	Methylation and foreign DNA restriction	CCCAA	41/44, 22	Known [30]
NEIS2364 (<i>modD</i>)	Methylation and foreign DNA restriction	ACCGA	41/44, 22	Known [30]
Single-copy OMPs				
NEIS1943 (<i>nalP</i>)	Autotransporter	C	11, 32, 213, 41/44, 22, 23, 162, new ST	Known [62]
NEIS1859 (<i>vapA</i>)	Contains autotransporter domain (UniProt)	GCAA	32, 213, 269, 41/44, 22, 23, 162, 461, new ST	SSR present
NEIS1783	Outer membrane protein class 4	GAACCC	32, 213, 269, 23, 162, 461, new ST	SSR present
NEIS1974 (<i>mvpA</i>)	Autotransporter	C	11, 32, 213, 269, 41/44, 162, 461, new ST	Known [62]
NEIS1634	Membrane fusion protein/efflux pump	GC	32, 213, 22, 23, 162, 461, new ST	SSR present
NEIS0774 (<i>clpA</i>)	ATP-dependent Clp protease	TGAAGA	11, 32, 213, 162, 461	SSR present
NEIS0978	Adhesin	C	11, 32, 213, 269	SSR present
NEIS1364 (<i>porA</i>)	Porin	G	All	Known [63]
NEIS1969 (<i>nadA</i>)	Adhesin/invasin	TAAA	11, 32, 213	Known [3]
Iron acquisition				
hmbR (<i>hmbR</i>)	Haemoglobin receptor	G	11, 32, 213, 269, 41/44, 162, 461, new ST	Known [64]
NEIS1963 (<i>fetA</i>)	Ferric enterobactin transporter	C	32, 269, 41/44, 23	Known [65]
NEIS1946 (<i>hpuA</i>)	Haemoglobin receptor	G	11, 22, 23	Known [66]
NEIS1691 (<i>tbpB</i>)	Transferrin-binding protein	ATAACAAA	22, 23	SSR present
NEIS1469 (<i>lbpB</i>)	Lactoferrin receptor	AAGCTG	11, 23	SSR present

Continued

Table 1. Continued

NEIS number (gene)	Function/functional domain	Predominant SSR	ccs gene conserved and subject to phase variation	Evidence for variation
NEIS1587 (<i>pigA</i>)	Haem utilization protein (haem oxygenase)	GAAGCC	11, 461	SSR present
Metabolism				
NEIS0354 (<i>hemL</i>)	Glutamate-1-semialdehyde aminotransferase	CGGTTG	All	SSR present
NEIS0276	Peptidyl-prolyl cis-trans isomerase	GCCAAAGCT	All	SSR present
NEIS0186 (<i>pdxA</i>)	4-Hydroxythreonine-4-phosphate dehydrogenase	GC	32, 213, 269, 22, 23, 162, 461, new ST	SSR present
NEIS0200	Ferredoxin	GC	213, 22, new ST	SSR present
NEIS1778 (<i>galM</i>)	Aldose 1-epimerase; carbohydrate metabolism	CCGCTACCC	22	SSR present
NEIS0501	Serine acetyltransferase	CCGCGG	461	SSR present
NEIS1749	UDP-MurNAc-pentapeptide synthetase	CGC	162	SSR present
NEIS0403 (<i>ribD</i>)	Riboflavin biosynthesis protein RibD	CGG	269	SSR present
Transcription/translation				
NEIS1518 (<i>alaS</i>)	Alanyl-tRNA synthetase	ACGCGC	11, 32, 213, 269, 22, 23, 162, 461, new ST	SSR present
NEIS1436 (<i>argS</i>)	Arginyl-tRNA synthetase	ACGCGC	32, 213, 269, 41/44, 22, 23, 461, new ST	SSR present
NEIS2017 (<i>truA</i>)	tRNA pseudouridine synthase A	C	11, 32, 213, 23, 162	SSR present
NEIS0183	Ribonuclease HII	T	11, 22, 461, new ST	SSR present
NEIS1106	Transcriptional regulator	TG	162	SSR present
Hypothetical				
NEIS2000	NA	CG	32, 213, 22, 23, 162, 461, new ST	SSR present
NEIS0405	NA	GCCAAAGCT	32, 269, 22, 23, 162, new ST	SSR present
NEIS3112	NA	CAAG	41/44, 162	SSR present
NEIS2475	NA	CCTT	41/44, 461, new ST	SSR present
NEIS2556	NA	AT	22, 23, 162	SSR present
NEIS0513	NA	G	269	SSR present
NEIS0782	NA	ACGGAT	213, 162	SSR present
NEIS2022	NA	CCTGTTT	269	SSR present
NEIS3062	NA	ATTATC	162	SSR present
NEIS1801	NA	A	32	SSR present
NEIS0317	NA	GACACG	162	SSR present
NEIS1357	NA	C	162	SSR present
Multi-copy OMP				
<i>opa</i>	Adhesion to epithelial cells	CTTCT	NA	Known [45]
Pseudogene				
NEIS1156	Homology to <i>IgtA</i>	G	All	SSR present
NEIS2930 (<i>icsA</i>)	Autotransporter precursor	AGCA	11, 32, 41/44, 22, 461, new ST	SSR present
NEIS2780	Type-1 restriction enzyme specificity MPN_089	G	41/44	SSR present

Continued

Table 1. Continued

NEIS number (gene)	Function/functional domain	Predominant SSR	ccs gene conserved and subject to phase variation	Evidence for variation
Other				
NEIS0950	Replication initiation	G	213	SSR present
NEIS0009	Uncharacterized membrane protein	A	All	SSR present
NEIS1297	Uncharacterized membrane protein	TAGGCT	11, 213, 41/44, 22, 23, 162, 461	SSR present
NEIS0092	Uncharacterized membrane protein	G	11, 22, 461, new ST	SSR present

of the numbers of *opa* loci (Fig. S5) indicates that the modal number of *opa* detected across all ccs was three, with 1–3 PV copies per genome. However, the modal number of *pilC* loci detected was 2–3 per genome. This is contrary to the expected two loci. Resolution of multi-copy PV genes is likely to improve with the availability of higher-quality genome sequences. Values for these loci are, therefore, representative of the percentage of PV versus non-PV for all these loci grouped together. We found that all ccs have at least one PV copy of both the *opa* and *pilC* loci.

Exemplars for the absence of phase variation: short SSRs, interrupted SSRs and problems in genome assembly

Lack of an SSR in a widely distributed PV gene can be either of evolutionary importance or due to errors in assembly of these repetitive sequences. Absences of SSRs in key loci were investigated by aligning gene sequences across multiple isolates with differing propensities for phase variation. Alignment of *fetA* sequences of representative isolates from the cc11 (PV) and cc32 (non-PV) lineages revealed that non-PV tracts had an SSR of length C6, which is below the threshold for a PV gene (Fig. S6a). Contrastingly, alignment of the *kdtA* sequences of a representative non-PV isolate, from cc11, and a PV isolate, from cc32, revealed a polyG tract in the non-PV isolate that was interrupted by a C nucleotide substitution (Fig. S6b). The *kdtA* gene is organized as part of a putative operon with NMB0013 and NMB0012 downstream (Fig. S6c). While translation of *kdtA* with a G10 tract is predicted to yield three distinct protein products, deletion of a single G nucleotide from the SSR results in a longer protein product due to fusion of the first two ORFs (Fig. S6d, e). This gene was shown to be non-essential but required for optimal fitness, as knockout *kdtA* mutants have impaired growth in rich media, with longer logarithmic doubling times and a lower maximum OD₆₀₀ [33].

An unexpected result was a lack of phase variation in some of the *pgl* genes. Aligned sequences of *pglA*, *pglI* and *pglH* were generated for ccs where these genes are either present or absent from the core phasomes. The findings were mixed, with *pglA* losing the propensity for phase variation due to interruption of the SSR (Fig. S7a), while *pglI* was missing from the list of PV genes due to difficulties in genome assembly producing contig breaks within the SSR (Fig. S7b). The *pglH*

gene was absent due to either highly dissimilar repeat regions or contig breaks (Fig. S7c). Resolution of these assembly problems and improved SSR detection could be achieved by the use of long-read sequencing technologies to supplement draft genomes produced by the Illumina platform.

Comparison of core- and accessory-phasome size between carriage and virulent strains

Incidences of IMD are thought to occur soon after strain acquisition, as the carriage state elicits a protective immune response and may also result in adaptation to the nasopharyngeal niche, driving meningococci into a non-invasive state. Differences in the ability to generate genetic and phenotypic diversity, therefore, may play a role in the propensity for strains to cause disease following recent host colonization. Therefore, we examined whether differential core PV gene repertoires could be observed between typically carriage-associated or endemic meningococcal ccs, and hypervirulent lineages. Fig. 4(a) indicates the size of core phasomes for each cc for three different cut-offs for conservation of phase variation: 95, 80 and 50%. With the most stringent cut-offs, we found that the hypervirulent meningococci had smaller core phasomes than carriage-associated lineages ($P < 0.05$). This significant difference is due to the loss of PV genes and/or absence of PV genes of a small gene set in the hypervirulent lineages that disappears at lower cut-offs as these missing genes merge into larger PV-containing gene homology groups. We interpret this modest difference as indicating that the core phasome is required for establishing a carriage state without impacting on the virulence potential of this organism.

The accessory-phasome sizes were also examined for each meningococcal cc as a measure of the variability in phasomes. We defined the accessory phasome as gene groupings absent from the core phasome. A count of the accessory phasome was performed independently for each complete cc dataset with a 50% cut-off. No significant differences in the accessory-phasome size were detected between meningococcal ccs suggesting that phase-variation capacity is similar across all lineages of this genus. Due to the differing number of strains in each cc, we analysed a random sample of 50 genomes from each cc. In this case, a large degree of variation was observed in the accessory-phasome size between ccs; however, carriage and disease-associated ccs were not significantly divergent (Fig. 4b). The cc with the largest accessory phasome was cc32, which had an accessory PV gene pool of ~250 deriving from

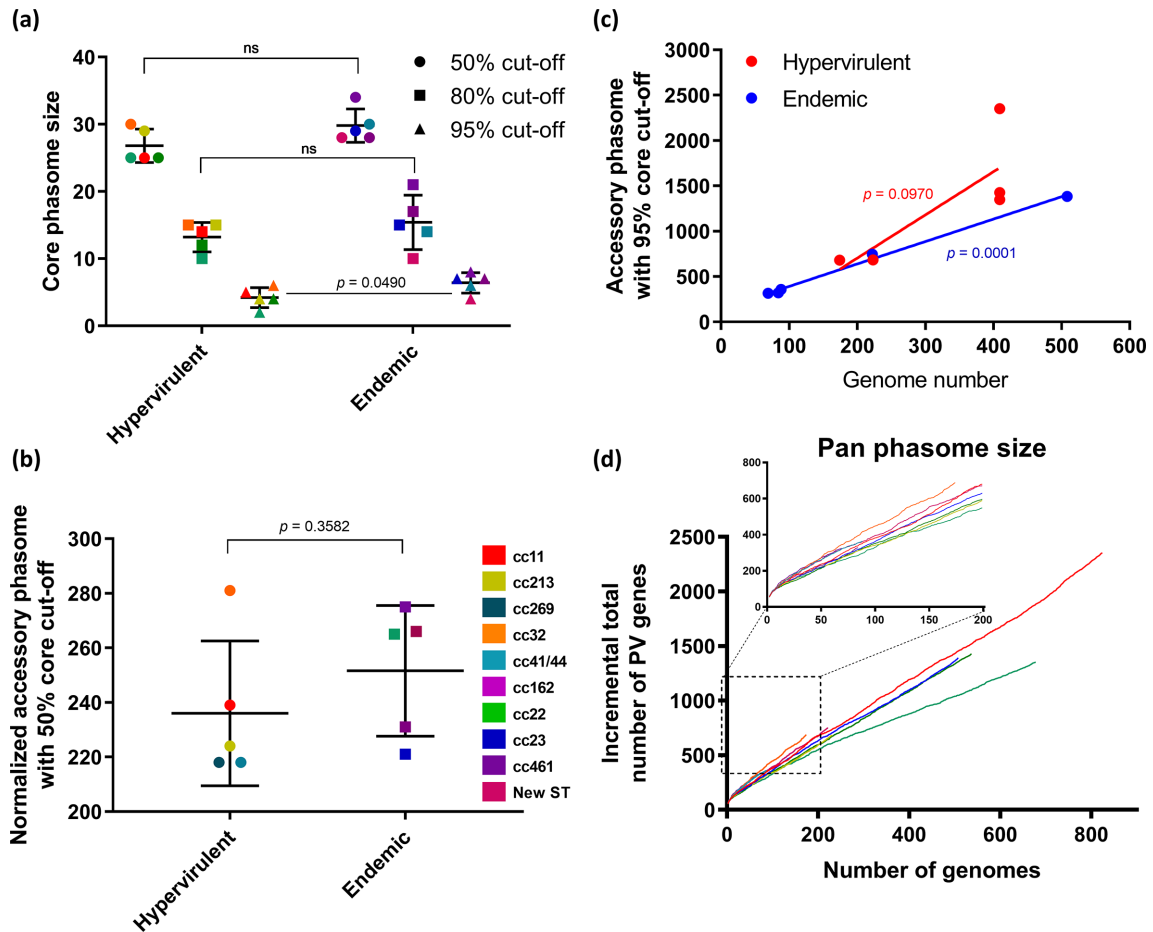


Fig. 4. Core- and accessory-phasome sizes of meningococcal ccs. (a) The size of core phasome was derived for each cc, by setting a cut-off of being present and PV in greater than 50, 80 or 95% of isolates. Statistical significance was determined by *t*-test, ns denotes no significance. (b) Accessory-phasome sizes of each cc derived from a random sample of 50 genomes from each cc. Statistical significance was determined using a *t*-test. (c) Correlation between total number of genome sequences analysed for each cc and the size of the accessory phasome. Statistical significance was determined using Pearson's correlation. (d) Pan-phasome analysis of the number of PV genes detected with an incremental number of genomes analysed for each cc. The inset shows the range where genomes were still being added for all ccs. Data points in (a), (b) and (d) corresponding to different ccs are colour coded as displayed in the key.

162 genomes, and a core-phasome size of only 28 genes. To further assess the accessory phasome, we compared the number of genomes present in a cc with the number of accessory PV loci (Fig. 4c). There was a significant correlation between genome number and accessory-phasome sizes for all combined ccs and for carriage-associated strains, but not for hypervirulent strains. As the number of PV genes has expanded with analysis of additional genomes, we examined whether the meningococcus has an open or closed pan phasome. We observed a trend for continual detection of PV genes (open pan phasome), but with differences between lineages such that cc11 trends towards more PV genes than cc41/44, indicative of a higher degree of genomic plasticity (Fig. 4d).

DISCUSSION

Hypervirulent and carriage-associated meningococcal lineages have differing propensities to cause disease that are likely

to have a genetic component [24]. Our analyses observed a modest difference in the combined frequency of PV genes between typically hypervirulent and carriage-associated ccs, but with no clear pattern of specific genes differentiating these phasomes. This indicates that PV gene content is not a determinant of hypervirulence, despite experimental evidence for many of these genes contributing to experimental invasive disease [46, 47]. Only *nadA* exhibited a bias towards hypervirulent lineages, being present in 3/5 of these lineages but absent from the carriage-associated lineages. Conversely, the majority of PV loci were present in both hypervirulent and carriage-associated ccs, indicating that PV gene content is a major determinant of carriage dynamics not disease propensity. This finding places the onus on analysis of the expression state of PV genes when considering whether these genes contribute to hypervirulence.

Whilst case studies of longitudinal carriage are readily available in the literature [16, 48], there is limited information on the lengths of carriage specific to particular lineages. We can speculate that different ccs may be predisposed to different periods and rates of carriage. Phase variation is considered an adaptive mechanism to reduce the fitness costs imposed by transmission bottlenecks, which may facilitate survival in a new host in the face of strong selective pressures by host immunity [49]. In this scenario, the number of PV genes encoded by a cc is a proxy for the ability to adapt to new environments with a consequent influence on lengths and rates of carriage. Interestingly, we observed that cc162 has a significantly larger phasome size than all other ccs leading to the prediction that this cc will have a higher propensity for persistent carriage than other lineages. The influence of PV gene content on carriage dynamics will require further investigation.

Typing schemes based on the genome sequence of bacterial pathogens have allowed for the accurate tracing of population evolution [50]. WGS phylogenetics of meningococci indicate the presence of clades containing ~2 ccs and numerous STs [51]. The clades and ccs have fairly deep evolutionary roots. We show that ccs cluster phylogenetically by their phasome content, indicating that co-evolution of the core genome and the phasome may be a common feature of all meningococcal lineages. However, there are examples of clustering of closely related ccs in the phasome analysis. This observation implies ongoing expansion/retraction of repeat tracts mediated by a diverse set of selection pressures that is independent of the core genome.

The only vaccines licensed for targeting protection against serogroup B meningococci are the protein-based Bexsero and Trumenba [13]. These vaccines are not fully effective against all MenB strains, whilst the emergence of vaccine escape strains and serotype replacement remain important concerns. Evaluation of the conservation of surface-exposed structures and their propensity to antigenically vary is required to develop novel broadly protective vaccines [52]. Our workflow will facilitate characterization of the likelihood of phase or antigenic variation of surface-exposed structures and whether there will be variation between ccs in these processes. For example, LOS is considered a vaccine target for gonococcus [53] and has previously been considered for the meningococcus [54], as in the face of potential vaccine escape, it may be a useful broadly protective antigen in the future [55]. Antibodies targeted against LOS core structures have been shown to be bactericidal for both species [56, 57]. Unexpectedly, our analysis revealed a varied pattern for putative phase variation of KdtA, with potential consequences for phase variation of LOS core structures across all lineages. These differential phase-variation capabilities may be a critical consideration when utilizing a specific LOS epitope as a vaccine antigen.

Within-host selection – such as exposure to antibodies – occurs with expression of surface-exposed structures. Modelling approaches have indicated that in the absence of selection, populations would drift towards loss of SSRs

[58]. Despite this, SSRs are prevalent in a variety of bacterial pathogens [22]. In addition, the repeat number directly correlates with the mutability of the locus, with high numbers predisposing to higher mutation rates [4, 5]. Differences in SSRs – either absolute repeat number or the binary below/above threshold for mutability – are likely reflective of the particular set of selection pressures imposed upon a clonal lineage. Our analysis indicated that while the majority of PV genes are conserved in their presence/absence, there are discrete differences in the presence or absence of repeat tracts at these loci. These data provide a conceptual framework to understand how selection pressures shape the evolution of diverse meningococcal lineages. Furthermore, these data allow analysis of the mechanistic basis for the genesis of PV genes in this species, such as the expansion of a pre-existing SSR or recombination resulting in transfer of an SSR between strains.

Funding information

J.J.W. was funded by a Medical Research Council (MRC) IMPACT studentship. L.R.G. was funded by the MRC through grant number MR/M020193/1.

Acknowledgements

This publication made use of the Meningitis Research Foundation (MRF) Meningococcal Genome Library, developed by Public Health England, the Wellcome Trust Sanger Institute and the University of Oxford, a project funded by the MRF. This study also made use of the *Neisseria* MLST website (<https://pubmlst.org/neisseria/>), hosted by the University of Oxford, and developed with funding from the Wellcome Trust and the European Union. This research used the ALICE High Performance Computing Facility at the University of Leicester.

Author contributions

Conceptualization: L.R.G., C.D.B. Supervision: C.D.B., L.R.G. Data curation: J.C.H., L.R.G. Formal analysis: J.J.W., L.R.G. Writing – original draft: J.J.W. Writing – review and editing: C.D.B., L.R.G.

Conflicts of interest

C.D.B. reports the receipt of grants from GlaxoSmithKline, Sanofi Pasteur MSD and Roche outside of this work. All other authors declare that there are no conflicts of interest.

Data Bibliography

1. Wanford JW, Holmes JC, Bayliss CD, Green LR. Data deposited at Zenodo, DOI:10.5281/zenodo.3545849 (2019).

References

1. Pizza M, Rappuoli R. *Neisseria meningitidis*: pathogenesis and immunity. *Curr Opin Microbiol* 2015;23:68–72.
2. Bayliss CD. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. *FEMS Microbiol Rev* 2009;33:504–520.
3. Metruccio MME, Pigozzi E, Roncarati D, Berlanda Scorza F, Norais N et al. A novel phase variation mechanism in the *Meningococcus* driven by a ligand-responsive repressor and differential spacing of distal promoter elements. *PLoS Pathog*;5:e1000710.
4. De Bolle X, Bayliss CD, Field D, van de Ven T, Saunders NJ et al. The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* 2000;35:211–222.
5. Bayliss CD, Bidmos FA, Anjum A, Manchev VT, Richards RL et al. Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. *Nucleic Acids Res* 2012;40:5876–5889.

6. Bayliss CD, Hoe JC, Makepeace K, Martin P, Hood DW *et al.* *Neisseria meningitidis* escape from the bactericidal activity of a monoclonal antibody is mediated by phase variation of IgtG and enhanced by a mutator phenotype. *Infect Immun* 2008;76:5038–5048.
7. Tauseef I, Ali YM, Bayliss CD. Phase variation of PorA, a major outer membrane protein, mediates escape of bactericidal antibodies by *Neisseria meningitidis*. *Infect Immun* 2013;81:1374–1380.
8. Saunders NJ, Jeffries AC, Peden JF, Hood DW, Tettelin H *et al.* Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol Microbiol* 2000;37:207–215.
9. Aidley J, Wanford JJ, Green LR, Sheppard SK, Bayliss CD. Phasomelt: an 'omics' approach to cataloguing the potential breadth of phase variation in the genus *Campylobacter*. *Microb Genomics* 2018;4:mgen.0.000228.
10. Wanford JJ, Green LR, Aidley J, Bayliss CD. Phasome analysis of pathogenic and commensal *Neisseria* species expands the known repertoire of phase variable genes, and highlights common adaptive strategies. *PLoS One* 2018;13:e0196675.
11. Mandal S, Campbell H, Ribeiro S, Gray S, Carr T *et al.* Risk of invasive meningococcal disease in university students in England and optimal strategies for protection using MenACWY vaccine. *Vaccine* 2017;35:5814–5818.
12. Trayner KMA, Cameron JC, Anderson N. Meningococcal ACWY (MenACWY) vaccine uptake, and barriers and motivations towards vaccination, in undergraduate students: a mixed-methods study. *The Lancet* 2017;390:S89.
13. Ladhani SN, Campbell H, Parikh SR, Saliba V, Borrow R *et al.* The introduction of the meningococcal B (MenB) vaccine (Bexsero®) into the national infant immunisation programme – new challenges for public health. *J Infect* 2015;71:611–614.
14. Kretz CB, Retchless AC, Sidikou F, Issaka B, Ousmane S *et al.* Whole-genome characterization of epidemic *Neisseria meningitidis* serogroup C and resurgence of serogroup W, Niger, 2015. *Emerg Infect Dis* 2015;22:1762–1768.
15. Green LR, Lucidarme J, Dave N, Chan H, Clark S *et al.* Phase variation of NadA in invasive *Neisseria* isolates impacts on coverage estimates for 4C-MenB, a MenB vaccine. *J Clin Microbiol* 2018;56:e00204–18.
16. Oldfield NJ, Cayrou C, AlJannat MAK, Al-Rubaiawi AAA, Green LR *et al.* Rise in group W meningococcal carriage in university students, United Kingdom. *Emerg Infect Dis* 2017;23:1009–1011.
17. Harrison OB, Schoen C, Retchless AC, Wang X, Jolley KA *et al.* *Neisseria* genomics: current status and future perspectives. *Pathog Dis* 2017;75:ftx060.
18. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*;3:124.
19. Findlow H, Campbell H, Lucidarme J, Andrews N, Linley E *et al.* Serogroup C *Neisseria meningitidis* disease epidemiology, seroprevalence, vaccine effectiveness and waning immunity, England, 1998/99 to 2015/16. *Euro Surveill* 2019;24:1700818.
20. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
21. Henderson IR, Owen P, Nataro JP. Molecular switches – the ON and OFF of bacterial phase variation. *Mol Microbiol* 1999;33:919–932.
22. Bayliss CD, Palmer ME. Evolution of simple sequence repeat-mediated phase variation in bacterial genomes. *Ann N Y Acad Sci* 2012;1267:39–44.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
24. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol* 2011;7:539.
25. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* 2000;16:944–945.
26. Weidlich L, Baethgen LF, Mayer LW, Moraes C, Klein CC *et al.* High prevalence of *Neisseria meningitidis* hypervirulent lineages and emergence of W135:P1.5.2:ST-11 clone in Southern Brazil. *J Infect* 2008;57:324–331.
27. Read RC. *Neisseria meningitidis*; clones, carriage, and disease. *Clin Microbiol Infect* 2014;20:391–395.
28. Tibayrenc M, Ayala FJ. How clonal are *Neisseria* species? The epidemic clonality model revisited. *Proc Natl Acad Sci* 2015;112:8909–8913.
29. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;13:36–46.
30. Tan A, Hill DMC, Harrison OB, Srikhanta YN, Jennings MP *et al.* Distribution of the type III DNA methyltransferases modA, modB and modD among *Neisseria meningitidis* genotypes: implications for gene regulation and virulence. *Sci Rep* 2016;6:21015.
31. Power PM, Roddam LF, Rutter K, Fitzpatrick SZ, Srikhanta YN *et al.* Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol Microbiol* 2003;49:833–847.
32. Zhu P, Klutch MJ, Bash MC, Tsang RSW, Ng L-K *et al.* Genetic diversity of three *Igt* loci for biosynthesis of lipooligosaccharide (LOS) in *Neisseria* species. *Microbiology* 2002;148:1833–1844.
33. Tzeng Y-L, Datta A, Kolli VK, Carlson RW, Stephens DS. Endotoxin of *Neisseria meningitidis* composed only of intact lipid A: inactivation of the meningococcal 3-deoxy-D-manno-octulosonic acid transferase. *J Bacteriol* 2002;184:2379–2388.
34. Francis F, Ramirez-Arcos S, Salimnia H, Victor C, Dillon JR. Organization and transcription of the division cell wall (*dcw*) cluster in *Neisseria gonorrhoeae*. *Gene* 2000;251:141–151.
35. Tzeng Y-L, Datta A, Strole C, Kolli VSK, Birck MR *et al.* KpsF is the arabinose-5-phosphate isomerase required for 3-deoxy-D-manno-octulosonic acid biosynthesis and for both lipooligosaccharide assembly and capsular polysaccharide expression in *Neisseria meningitidis*. *J Biol Chem* 2002;277:24103–24113.
36. Perkins-Balding D, Rattliff-Griffin M, Stojiljkovic I. Iron transport systems in *Neisseria meningitidis*. *Microbiol Mol Biol Rev* 2004;68:154–171.
37. Harrison OB, Maiden MCJ, Rokbi B. Distribution of transferrin binding protein B gene (*tbpB*) variants among *Neisseria* species. *BMC Microbiol* 2008;8:66.
38. Pettersson A, Prinz T, Umar A, van der Biezen J, Tommassen J. Molecular characterization of LbpB, the second lactoferrin-binding protein of *Neisseria meningitidis*. *Mol Microbiol* 1998;27:599–610.
39. Rattliff M, Zhu W, Deshmukh R, Wilks A, Stojiljkovic I. Homologues of neisserial heme oxygenase in gram-negative bacteria: degradation of heme by the product of the *pigA* gene of *Pseudomonas aeruginosa*. *J Bacteriol* 2001;183:6394–6403.
40. Russell JE, Jolley KA, Feavers IM, Maiden MCJ, Suker J. PorA variable regions of *Neisseria meningitidis*. *Emerg Infect Dis* 2004;10:674–678.
41. Comanducci M, Bambini S, Caugant DA, Mora M, Brunelli B *et al.* NadA diversity and carriage in *Neisseria meningitidis*. *Infect Immun* 2004;72:4217–4223.
42. Turner DPJ, Marietou AG, Johnston L, Ho KKL, Rogers AJ *et al.* Characterization of MspA, an immunogenic autotransporter protein that mediates adhesion to epithelial and endothelial cells in *Neisseria meningitidis*. *Infect Immun* 2006;74:2957–2964.
43. van Ulsen P, van Alphen L, Ten Hove J, Fransen F, van der Ley P *et al.* A neisserial autotransporter NalP modulating the processing of other autotransporters. *Mol Microbiol* 2003;50:1017–1030.
44. Morand PC, Drab M, Rajalingam K, Nassif X, Meyer TF. *Neisseria meningitidis* differentially controls host cell motility through PilC1 and PilC2 components of type IV pili. *PLoS One* 2009;4:e6834.
45. Aho EL, Dempsey JA, Hobbs MM, Klapper DG, Cannon JG. Characterization of the *opa* (class 5) gene family of *Neisseria meningitidis*. *Mol Microbiol* 1991;5:1429–1437.

46. Mackinnon FG, Borrow R, Gorrington AR, Fox AJ, Jones DM et al. Demonstration of lipooligosaccharide immunotype and capsule as virulence factors for *Neisseria meningitidis* using an infant mouse intranasal infection model. *Microb Pathog* 1993;15:359–366.
47. Echenique-Rivera H, Muzzi A, Del Tordello E, Seib KL, Francois P et al. Transcriptome analysis of *Neisseria meningitidis* in human whole blood and mutagenesis studies identify virulence factors involved in blood survival. *PLoS Pathog* 2011;7:e1002027.
48. van Ravenhorst MB, Bijlsma MW, van Houten MA, Struben VMD, Anderson AS et al. Meningococcal carriage in Dutch adolescents and young adults; a cross-sectional and longitudinal cohort study. *Clin Microbiol Infect* 2017;23:573.e1–573.e7.
49. Moxon R, Kussell E. The impact of bottlenecks on microbial survival, adaptation, and phenotypic switching in host–pathogen interactions. *Evolution* 2017;71:2803–2816.
50. Feil EJ, Enright MC. Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* 2004;7:308–313.
51. Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D et al. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc Natl Acad Sci USA* 2011;108:4494–4499.
52. de Filippis I. Quest for a broad-range vaccine against *Neisseria meningitidis* serogroup B: implications of genetic variations of the surface-exposed proteins. *J Med Microbiol* 2009;58:1127–1132.
53. Gulati S, Shaughnessy J, Ram S, Rice PA. Targeting lipooligosaccharide (LOS) for a gonococcal vaccine. *Front Immunol* 2019;10:00321.
54. Christodoulides M, Heckels J. Novel approaches to *Neisseria meningitidis* vaccine design. *Pathog Dis* 2017;75:ftx033.
55. Zollinger WD, Donets MA, Schmiel DH, Pinto VB, Labrie JE et al. Design and evaluation in mice of a broadly protective meningococcal group B native outer membrane vesicle vaccine. *Vaccine* 2010;28:5057–5067.
56. Chakraborti S, Lewis LA, Cox AD, St Michael F, Li J et al. Phase-variable heptose I glycan extensions modulate efficacy of 2C7 vaccine antibody directed against *Neisseria gonorrhoeae* lipooligosaccharide. *J Immunol* 2016;196:4576–4586.
57. John CM, Phillips NJ, Stein DC, Jarvis GA. Innate immune response to lipooligosaccharide: pivotal regulator of the pathobiology of invasive *Neisseria meningitidis* infections. *Pathog Dis* 2017;75:ftx030.
58. Palmer ME, Lipsitch M, Moxon ER, Bayliss CD. Broad conditions favor the evolution of phase-variable loci. *mBio* 2013;4:e00430-12.
59. Shafer WM, Datta A, Kolli VSK, Rahman MM, Balthazar JT et al. Phase variable changes in genes lgtA and lgtC within the lgtABCDE operon of *Neisseria gonorrhoeae* can modulate gonococcal susceptibility to normal human serum. *J Endotoxin Res* 2002;8:47–58.
60. Børud B, Bårnes GK, Brynildsrud OB, Fritzsønn E, Caugant DA. Genotypic and phenotypic characterization of the O-linked protein glycosylation system reveals high glycan diversity in paired meningococcal carriage isolates. *J Bacteriol* 2018;200:e00794-17.
61. Rytönen A, Albiger B, Hansson-Palo P, Källström H, Olcén P et al. *Neisseria meningitidis* undergoes PilC phase variation and PilE sequence variation during invasive disease. *J Infect Dis* 2004;189:402–409.
62. Oldfield NJ, Matar S, Bidmos FA, Alamro M, Neal KR et al. Prevalence and phase variable expression status of two autotransporters, NalP and MspA, in carriage and disease isolates of *Neisseria meningitidis*. *PLoS One*;8:e69746.
63. van der Ende A, Hopman CT, Dankert J. Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. *Infect Immun* 2000;68:6685–6690.
64. Richardson AR, Stojiljkovic I. HmbR, a hemoglobin-binding outer membrane protein of *Neisseria meningitidis*, undergoes phase variation. *J Bacteriol* 1999;181:2067–2074.
65. Carson SD, Stone B, Beucher M, Fu J, Sparling PF. Phase variation of the gonococcal siderophore receptor FetA. *Mol Microbiol* 2000;36:585–593.
66. Lewis LA, Gipson M, Hartman K, Ownbey T, Vaughn J et al. Phase variation of HpuAB and HmbR, two distinct haemoglobin receptors of *Neisseria meningitidis* DNM2. *Mol Microbiol* 1999;32:977–989.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.