

# SCIENTIFIC REPORTS



OPEN

## Distribution of the type III DNA methyltransferases *modA*, *modB* and *modD* among *Neisseria meningitidis* genotypes: implications for gene regulation and virulence

Aimee Tan<sup>1,\*</sup>, Dorothea M. C. Hill<sup>2,\*</sup>, Odile B. Harrison<sup>2</sup>, Yogitha N. Srikhanta<sup>1,†</sup>, Michael P. Jennings<sup>1</sup>, Martin C. J. Maiden<sup>2</sup> & Kate L. Seib<sup>1</sup>

*Neisseria meningitidis* is a human-specific bacterium that varies in invasive potential. All meningococci are carried in the nasopharynx, and most genotypes are very infrequently associated with invasive meningococcal disease; however, those belonging to the 'hyperinvasive lineages' are more frequently associated with sepsis or meningitis. Genome content is highly conserved between carriage and disease isolates, and differential gene expression has been proposed as a major determinant of the hyperinvasive phenotype. Three phase variable DNA methyltransferases (ModA, ModB and ModD), which mediate epigenetic regulation of distinct *phase variable regulons* (phasevarions), have been identified in *N. meningitidis*. Each *mod* gene has distinct alleles, defined by their Mod DNA recognition domain, and these target and methylate different DNA sequences, thereby regulating distinct gene sets. Here 211 meningococcal carriage and >1,400 disease isolates were surveyed for the distribution of meningococcal *mod* alleles. While *modA11-12* and *modB1-2* were found in most isolates, rarer alleles (e.g., *modA15*, *modB4*, *modD1-6*) were specific to particular genotypes as defined by clonal complex. This suggests that phase variable Mod proteins may be associated with distinct phenotypes and hence invasive potential of *N. meningitidis* strains.

*Neisseria meningitidis*, the meningococcus, is a human-specific bacterium that exists as a commensal in approximately 10% of the population<sup>1</sup>; however, some meningococci are associated with severe pathology with rapid disease onset of sepsis and/or meningitis<sup>2,3</sup>. Although certain host factors have been identified that contribute to disease susceptibility (including age, medical conditions and genetic factors<sup>2,4,5</sup>), the precise mechanisms that determine invasive potential and that mediate transition of a given meningococcus from carriage to invasive disease remain unclear. Invasive isolates are typically characterized as one of six capsule-based meningococcal serogroups (A, B, C, X, Y and W). These are globally distributed with varying rates of disease incidence<sup>6</sup>. Meningococci are characterized at the genetic level by multi locus sequence typing (MLST), which uses sequences of seven housekeeping genes to determine isolate sequence type (ST)<sup>7</sup>. Groups of closely related STs are termed clonal complexes (cc) and these are good surrogates of bacterial lineage. A subset of clonal complexes, known as hyperinvasive (or hypervirulent) lineages, is responsible for the majority of disease worldwide. These hyperinvasive lineages are significantly overrepresented in collections of invasive isolates (i.e. those from blood or cerebrospinal fluid) relative to collections of asymptomatic carriage isolates (i.e., those from the nasopharynx), and include cc4/5, cc8, cc11, cc32, cc41/44, and cc269<sup>7-10</sup>. On the other hand, meningococci isolated from

<sup>1</sup>Institute for Glycomics, Griffith University, Gold Coast, Queensland 4222, Australia. <sup>2</sup>Department of Zoology, University of Oxford, Oxford, UK. <sup>†</sup>Present address: Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, 3800, Australia. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.L.S. (email: k.seib@griffith.edu.au)

asymptomatic carriers consist of highly diverse genotypes with some clonal complexes, such as cc23, cc35, cc106 and cc116, very rarely if ever associated with invasion<sup>10</sup>. Even amongst the hyperinvasive lineages, different genotypes often vary in pathogenicity; for example within the cc41/44 lineage, ST-41 meningococci are more commonly associated with disease, and ST-44 with carriage<sup>10</sup>.

Attempts to identify the specific factors responsible for differential virulence of *N. meningitidis* strains have been largely unsuccessful. Whilst numerous factors important for virulence have been identified (e.g. capsule, pili, lipooligosaccharide (LOS) and opacity proteins)<sup>11</sup>, there are no virulence factors that clearly distinguish highly pathogenic isolates that cause invasive disease from less pathogenic isolates. Many proteins considered to be key virulence determinants are also present in *N. meningitidis* carriage isolates and commensal *Neisseria* species<sup>12–14</sup>. Indeed, most meningococcal isolates have a similar overall genome composition, consisting of 79% core genes, 21% accessory genes and less than 0.1% strain specific genes<sup>15</sup>, with a core meningococcal genome of approximately 1600 genes<sup>15,16</sup>. Restriction-modification systems are among the few isolate and clonal complex-specific genes identified<sup>15,17</sup>. Other genetic elements associated with invasive isolates include the *hmbR* hemoglobin receptor gene<sup>18</sup> and a prophage<sup>12,19,20</sup>; however, the mechanistic contribution of these factors to virulence is unclear. Consequently, the pathogenic potential of isolates is hypothesized to be a polygenetic phenomenon arising from varied adhesin and metabolic gene content, and expression differences<sup>21,22</sup>.

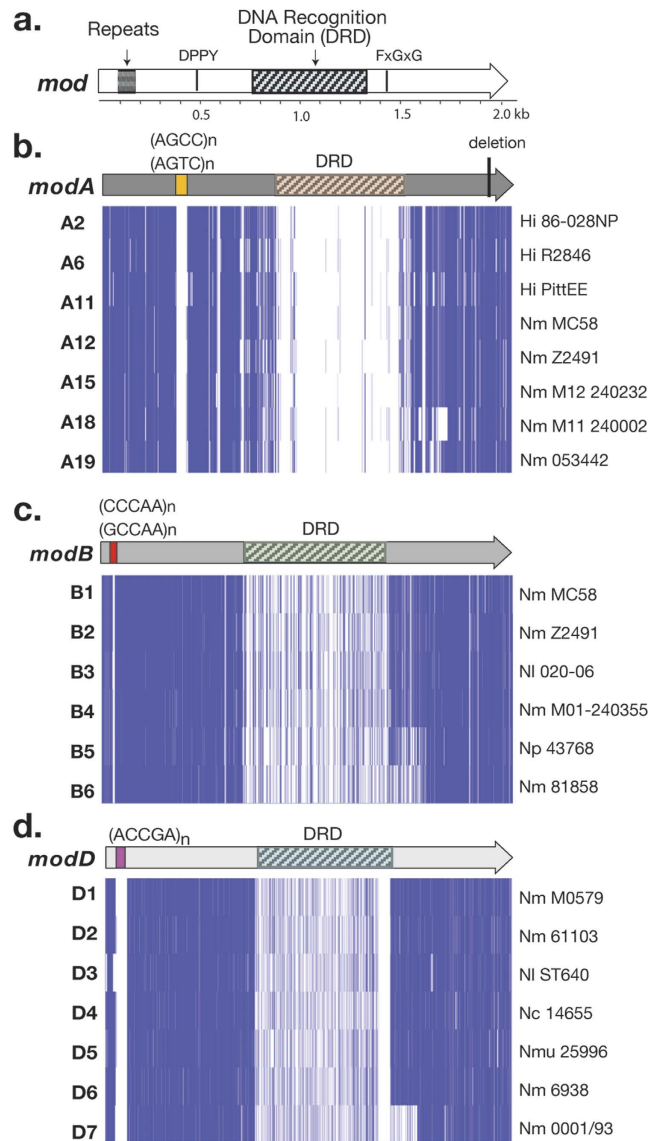
While meningococci are considered to have relatively few transcriptional regulators when compared to other bacterial species<sup>23</sup>, they do contain numerous phase variable genes<sup>24–26</sup> and differential gene expression provides a possible explanation for phenotypic differences. Furthermore, *N. meningitidis* contain a number of phase variable DNA methyltransferases (Mod), associated with type III restriction-modification systems, which mediate epigenetic regulation<sup>27–30</sup>. Random, reversible, hypermutation of repetitive DNA tracts within the open reading frame of *mod* genes lead to frame-shift mutations and the ON/OFF switching of Mod expression (i.e., phase variation). Mod phase variation results in distinct bacterial populations with different patterns of genome methylation, and altered expression of specific sets of genes. These *phase variable* regulons have been defined as phasevarions<sup>29,31</sup>. Mod phasevarions studied to date in pathogenic *Neisseria* have been shown to contain genes encoding outer membrane proteins, stress response proteins and other metabolic components<sup>27,28</sup>. These phasevarions represent an epigenetic mechanism by which meningococcal cells can alter complex phenotypes, which may affect carriage or invasion.

Three *mod* genes have been described in *N. meningitidis*: *modA*, *modB*, and *modD*<sup>27,28</sup>. These share a similar overall structure, with N-terminal simple DNA repeats, N- and C-terminal domains that mediate DNA methylation, and a central DNA recognition domain (DRD) that determines the recognition and methylation site of the enzyme (See Fig. 1a). For each of the *modA*, *modB*, and *modD* genes, different alleles have been identified that have conserved N- and C-terminal domains (>90% amino acid identity), but which vary in the DRD sequence (>95% amino acid identity within alleles, and typically <40% identity among alleles). Different *mod* alleles (i.e. DRD variants) methylate different DNA sequences<sup>28,30,32</sup> and regulate different phasevarions<sup>27,28</sup>. The *modA* gene has the highest known allelic diversity, with 20 known alleles (*modA1–20*), many of which are found in *Haemophilus influenzae* (*modA1–10*, *modA14–17*, *modA20*) and/or *N. meningitidis* (*modA4*, *modA11–13*, *modA15*, *modA18–19*)<sup>28,33</sup>. In contrast, only four *modB* alleles have been reported: *modB1*, *modB2*<sup>28</sup>, *modB3*<sup>34</sup> and *modB4*<sup>15</sup>. The *modB* gene has only been found in *Neisseria* species to-date, with *modB1* found in *N. meningitidis* and *N. gonorrhoeae*, *modB2* and *modB4* in *N. meningitidis*<sup>28</sup>, and *modB3* in *Neisseria lactamica*<sup>29</sup>. The *modD* gene also appears to be *Neisseria* specific, and has five known alleles: *modD1* and *modD2* in *N. meningitidis*, *modD3* in *N. lactamica*, *modD4* in *Neisseria cinerea* and *modD5* in *Neisseria mucosa*<sup>27</sup>. To date, the DNA methylation target sequences of *N. meningitidis* *ModA11*, *ModA12* and *ModD1*<sup>30</sup>; *N. gonorrhoeae* *ModA13* and *ModB1*<sup>28,32</sup>; *H. influenzae* *ModA1*, *ModA2*, *ModA4*, *ModA5*, *ModA9* and *ModA10*<sup>35</sup>; and *M. catarrhalis* *ModM2* and *ModM3* have been determined<sup>36</sup>, with a unique site methylated by each allele. This study surveyed the distribution and combination of *mod* genes and their associated alleles in four *N. meningitidis* isolate collections, and identified their association with certain hyperinvasive lineages. These data are consistent with Mod proteins playing a role in the survival of meningococci in the different environments encountered during colonization and invasion in the human host.

## Results

**Distribution of *mod* genes and alleles.** To investigate the distribution of *mod* genes and alleles in *N. meningitidis* (Fig. 1), 1,689 isolates were surveyed, comprising 211 carriage and 1,478 disease isolates from four collections originating in diverse geographic locations (including the USA, UK, Czech Republic, and Australia) and time periods (1993–2013). These analyses determined that a *modA* gene was present in all isolates examined (although two isolates possessed only fragments of *modA*). A *modB* gene was identified in 78% (1,298) of isolates, while a *modD* gene was present in only 25% (423) of isolates (Fig. 2). Overall, 364 isolates contained *modA* only, 900 isolates contained both *modA* and *modB*, and 398 isolates contained *modA*, *modB* and *modD* genes. The number of DNA repeat units among isolates ranged from 2–34 tetranucleotide repeats in *modA*; 2–28 pentanucleotide repeats in *modB*; and 2–15 pentanucleotide repeats in *modD* (Fig. 1). Phase variation of tetranucleotide repeat tracts containing  $\geq 3$  repeat units occurs at a high frequency<sup>37</sup>, and the majority of these *mod* genes are predicted to be phase variable (>98% of fully assembled alleles contain  $\geq 3$  repeat units).

For each of the three *mod* genes, common and rare *mod* alleles (DRD variants) were found (Figs 1 and 2). The majority of *modA* positive isolates contained *modA12* (1,159 isolates, 70%) or *modA11* (456 isolates, 27.5%), with *modA15* comprising 2% of *modA* positive isolates (38 isolates) (Figs 1b and 2b). All other *modA* alleles were found at low frequency in the dataset, two of which, *modA2* and *modA6*, had not been reported in *N. meningitidis* before. In addition, several minor variations in the conserved regions of the *modA11* and *modA12* were seen among the isolates. The most frequent of these was a 15-nucleotide deletion (encoding [S(A/V)KNQ]) in the region encoding the C-terminus of ModA (Fig. 1b), found in 67% of *modA11* alleles and 59% of *modA12* alleles

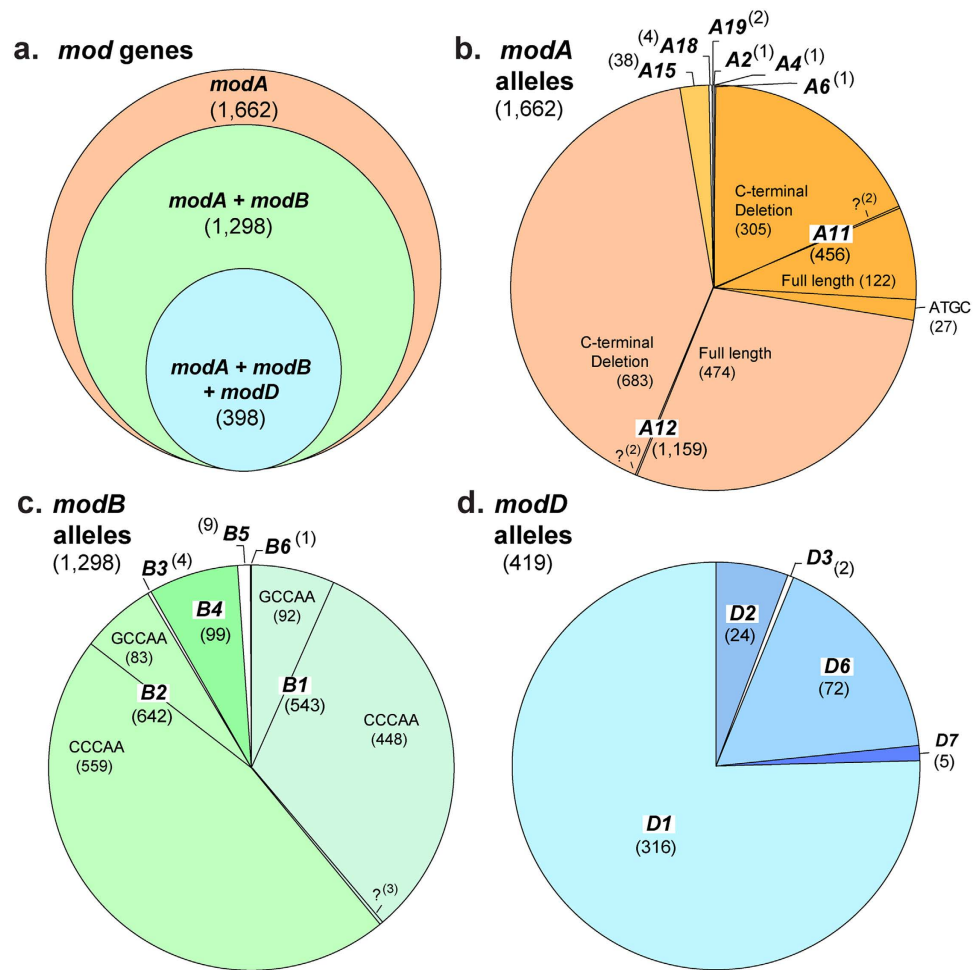


**Figure 1. Key features of *mod* genes and alleles.** (a) Schematic diagram of *mod* genes (represented by an arrow), showing the location of the phase variable DNA repeats, the DPPY and FxGxG motifs, and the central DNA recognition domain (DRD). For (b) *modA*, (c) *modB* and (d) *modD* genes, the location and type of phase variable repeats are shown, along with the location of the DRD and a graphical representation of a clustalW alignment of reference alleles [generated in JalView, with identical residues shown as vertical blue lines]. Strains that define the *mod* alleles are shown on the right (Hi: *Haemophilus influenzae*; Nm: *N. meningitidis*; NI: *N. lactamica*; Np: *N. polysaccharea*; Nc: *N. cinerea*; Nmu: *N. mucosa*). For *modA* (b), the location of the 15-nucleotide deletion is also shown.

(Fig. 2b). This deletion removed 5 amino acids from the full length of the protein. In addition, the N-terminal, phase variable DNA repeat sequences were altered in some isolates. The typical *modA* repeat unit in *N. meningitidis* was 5'-AGCC-3', however *modA11* in 27 isolates (6% of *modA11* isolates; 1.6% of total *modA* positive isolates) had 5'-AGTC-3' repeats (Fig. 2b).

For *modB*, the most common alleles were *modB1* (543 isolates, representing 42% of all *modB* positive isolates) and *modB2* (642 isolates, 49%), with *modB4* present in 7.6% of *modB* positive isolates (Figs 1c and 2c). All other *modB* alleles were found at low frequency, including two new alleles (*modB5* and *modB6*), which have not previously been defined. The *modB5* and *modB6* DRDs shared 11–48% identity at the deduced amino acid level with other *modB* alleles (Fig. 1c). Both were present at low frequency, with 9 *modB5* isolates (0.7%) and 1 *modB6* isolate (0.08%) identified. For *modB1* and *modB2*, the typical repeat unit was 5'-CCCAA-3', with an alternate 5'-GCCAA-3' repeat tract seen in 17% of *modB1* alleles and 13% of all *modB2* (Fig. 2c). *modB3* had both 5'-CCCAA-3' and 5'-TCCAA-3' repeats, and *modB4* and *modB5* typically contained 5'-GCCAA-3' repeats.

The majority of *modD* positive isolates contained the *modD1* allele (316 isolates, 75%), or *modD6* (72 isolates, 17%; Figs 1d and 2d). All other *modD* alleles were found at low frequency, including *modD3* that has



**Figure 2. Distribution of *mod* genes and alleles.** (a) Venn diagram of the distribution of *mod* genes in the dataset. (b–d) Pie charts of *mod* alleles (and allele variations) with the number of isolates containing each allele shown in brackets. (b) *modA* alleles 2, 4, 6, 11, 12, 15, 18, 19 are shown, as well as the A11 5′-AGTC-3′ repeat variant, and the A11 and A12 15bp C-terminal deletion variants (NB. For 2 *modA11* and 2 *modA12* alleles it was unclear from whole genome sequences whether they were full length or had the C-terminal deletion, and these are indicated with a question mark ‘?’). (c) *modB* alleles 1–6 are shown, as well as the B1 and B2 5′-CCCAA-3′ and 5′-GCCAA-3′ repeat variants (NB. The repeat region could not be categorized for 3 *modB1* alleles, and these are indicated with a question mark ‘?’). (d) *modD* alleles 1–7 are shown.

not been reported in *N. meningitidis* before, *modD6* that was previously identified in *N. meningitidis* 6938 but mis-categorized as *modD2*<sup>38,39</sup>, and *modD7* that has not previously been identified. The novel *modD7* allele was present in 5 isolates (1.2% of *modD* positive isolates), and shared 7–15% identity with other *modD* alleles over the length of the DRD (Fig. 1d). In addition, this allele had an extended DRD, with an additional 47 amino acids at the C-terminal end of the DRD, compared to other *modD* alleles (Fig. 1d).

Associations among different *mod* alleles were also identified. For example, of those meningococci containing only *modA* and *modB*, 71% of isolates with *modB1* also contained *modA11*, while 96% of isolates with *modB2*, and 89% of isolates with other *modB* alleles, were associated with *modA12*. Furthermore, 86% of isolates with a *modD* gene also contained *modA12* and *modB2*, which rose to 97% in isolates containing the *modD1* allele.

**Specific *mod* genes are associated with distinct clonal complexes.** Examination of the distribution of *mod* in the dataset demonstrated that most *mod* genes and alleles were associated with specific sequence types and clonal complexes (Table 1). *ModA11*, *modA12*, *modB1* and *modB2* alleles were found in multiple clonal complexes and the majority of isolates from the same clonal complex contained the same allele. For example, *modA11* was present in cc32 (97% of cc32 isolates) and cc116 (85%), whereas *modA12* was present in cc11 (99%) and cc23 (97%) isolates. The C-terminal deletion variations were also clonal complex associated, with the *modA11* C-terminal deletion found in cc35 (82%), cc53 (91%) and cc269 (85%), and the *modA12* C-terminal deletion found in cc41/44 (97%) and cc213 (89%) meningococci. The *modB1* allele was found in cc269 (86%), and cc32 (100%), and the *modB2* allele was associated with cc41/44 (76%) and cc11 (100%) isolates; however, *modB* was absent in all cc23, cc53, cc92, cc106, cc116 and cc461 meningococci examined. *ModB1* and *modB2* sequences with non-typical repeat tracts were frequently found in cc60 (82%) and cc22 (81%), respectively. On the other

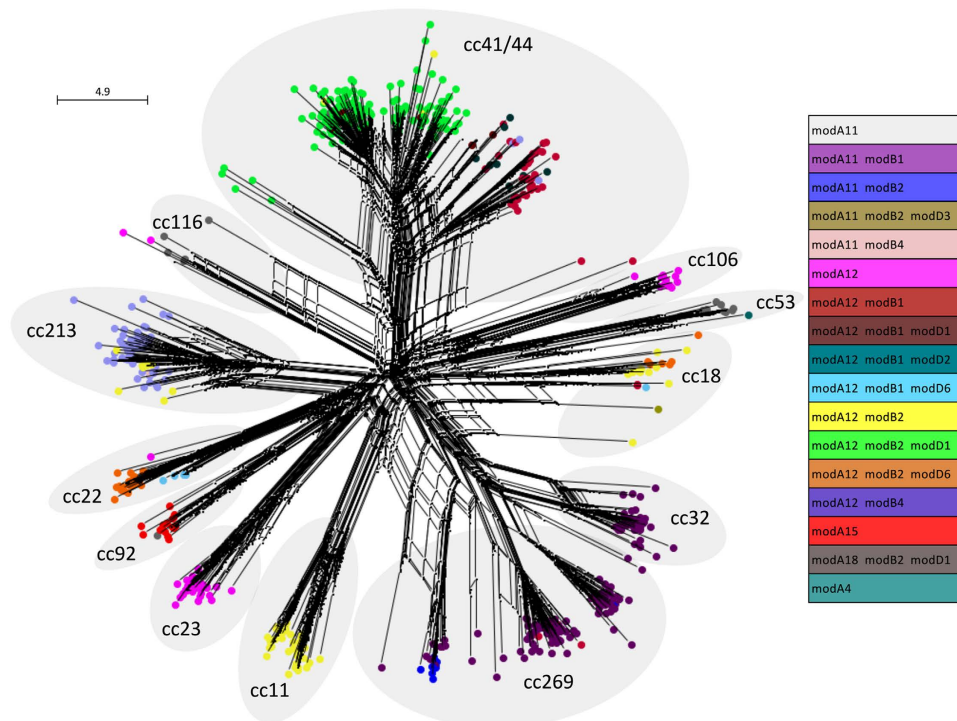
Gene (Locus record number) <sup>a</sup> Allele <sup>b</sup>	Predominant clonal complex associated with allele			
		cc <sup>c</sup>	% (#) <sup>d</sup>	p-value <sup>f</sup>
<i>modA</i> (NEIS1310)				
<i>modA11</i>	Full length	cc32	97 (72/74)	<0.0001
		cc116	85 (11/13)	<0.0001
	C-terminal deletion <sup>e</sup>	cc35	82 (14/17)	<0.0001
		cc53	91 (10/11)	<0.0001
		cc269	85 (231/272)	<0.0001
<i>modA12</i>	Full length	cc162	100 (23/23)	<0.0001
		cc11	99 (175/176)	<0.0001
<i>modA15</i>	C-terminal deletion <sup>e</sup>	cc23	97 (173/179)	<0.0001
		cc41/44	97 (363/368)	<0.0001
		cc213	89 (90/101)	<0.0001
<i>modB</i> (NEIS1194)				
<i>modB1</i>	5'-CCCAA-3' repeats	cc32	100 (74/74)	<0.0001
		cc41/44	16 (60/368)	<0.0001
		cc269	86 (233/272)	<0.0001
	5'-GCCAA-3' repeats <sup>e</sup>	cc60	82 (18/22)	<0.0001
		cc269	8 (22/272)	0.0016
<i>modB2</i>	5'-CCCAA-3' repeats	cc11	100 (176/176)	<0.0001
		cc41/44	76 (281/363)	<0.0001
	5'-GCCAA-3' repeats <sup>e</sup>	cc22	81 (35/43)	<0.0001
<i>modB4</i>		cc174	93 (14/15)	<0.0001
		cc213	90 (91/101)	<0.0001
<i>modB5</i>		cc282	100 (9/9)	<0.0001
No <i>modB</i>		cc23	100 (179/179)	<0.0001
		cc461	100 (31/31)	<0.0001
<i>modD</i> (NEIS2364)				
<i>modD1</i>		cc41/44	80 (308/384)	<0.0001
<i>modD2</i>		cc41/44	2 (9/384)	0.0001
<i>modD6</i>		cc18	30 (7/23)	<0.0001
		cc22	98 (44/45)	<0.0001

**Table 1. Associations between *mod* alleles and clonal complexes.** <sup>a</sup>*mod* genes and corresponding locus record NEIS numbers in the PubMLST database (<http://pubmlst.org/neisseria/>). <sup>b</sup>Alleles were determined based on >95% sequence identity to representative type sequences outlined in the material and methods. <sup>c</sup>Predominant clonal complexes containing the respective *mod* allele, with complexes, determined by the Nm MLST scheme (<http://pubmlst.org/neisseria/>). Note that this analysis was restricted to isolates with full allele, and cc data available. <sup>d</sup>% of isolates from the clonal complex with the allele (number of isolates with the allele/total number of isolates from the clonal complex). <sup>e</sup>Variations to the reference allele. C-terminal deletion: missing 15 nucleotides near the C-terminal of the gene. 5'-AGTC-3' and 5'-GCCAA-3' repeats: contain different repeat tract sequences to the reference allele. Further details on allele variations are given in the text. <sup>f</sup>p-value for the association of the given allele with the clonal complex listed (vs. other clonal complexes), calculated using the Fisher's Exact test (two-tailed).

hand, the less common alleles were associated with a single clonal complex, for example, *modA15* was associated with cc92 (present in 95% of cc92 meningococci vs. 0.1% of other clonal complexes,  $p < 0.0001$ ), *modB4* with cc213 (90% vs. 0.3%,  $p < 0.0001$ ), and *modD1* with cc41/44 (80% vs. 0%,  $p < 0.0001$ ) (Table 1).

Further associations between *mod* alleles and clonal complexes were identified when *mod* allelic combinations were mapped onto a Ribosomal Multilocus Sequence Typing (rMLST) network of the twelve most frequently represented clonal complexes in this dataset (cc106, cc11, cc116, cc18, cc213, cc22, cc23, cc269, cc32, cc41/44, cc53 and cc92) with certain combinations more evident than others (Fig. 3, Supplementary Table 1). For example, *modA11* alone (*i.e.*, without *modB* or *modD*) was present in 91% of cc53 and 85% of cc116 isolates (vs. 1.8% of isolates from other complexes;  $p < 0.0001$ ), *modA12* alone was present in 100% of cc106, 99% of cc23 and 100% of cc461 isolates (vs. 1.7% in other complexes,  $p < 0.0001$ ), while *modA15* alone was present in 95% of cc92 isolates (vs. 0.1% in other complexes;  $p < 0.0001$ ). The *modA11-modB1* combination was found in 89% of cc269 and 99% of cc32 isolates (vs. 1.2% in other complexes,  $p < 0.0001$ ); while the *modA12-modB2* combination was found in 99% of cc11 isolates (vs. 5.2% in other complexes,  $p < 0.0001$ ); however, the *mod*- clonal complex associations were not always completely conserved. For example, cc18 isolates did not share a common allelic combination, and cc41/44 isolates were clustered into two groups, one including ST-41 and the other ST-44 isolates. The ST-41 cluster was associated with *modA12*, *modB2* and *modD1* (78% of isolates, with this combination not seen in other





**Figure 3. Association of *mod* alleles and clonal complexes.** Neighbor-Net graph drawn from Ribosomal Multilocus Sequence Typing (rMLST) allele distances among unique Ribosomal Sequence Types (rSTs;  $n = 639$ ) in clonal complexes comprising  $> 10$  isolates in the dataset. Isolates are colored based on the *mod* allele or allele combination present.

complexes,  $p < 0.0001$ ) whereas the ST-44 cluster was associated with *modA12* and *modB2* but lacked *modD1* (16% of ST-44 isolates vs. 3.8% in other complexes,  $p < 0.0001$ ; Fig. 3). Similarly, most cc22 isolates contained the *modA12-modB2-modD6* combination, but a smaller cluster of isolates had *modA12-modB1-modD6*. The majority of cc269 isolates possessed *modA11-modB1*, however one cluster was associated with *modA11-modB2* (Fig. 3).

**Specific *mod* genes and combinations are associated with invasive or carriage meningococci.** Given that different *mod* alleles regulate distinct sets of genes (phasevarions) that affect the phenotype of the isolate, the distribution of *mod* alleles relative to the isolate's disease outcome (*i.e.*, invasive disease or carriage) was considered. Several associations were observed among *mod* alleles and invasive or carried meningococci, and these associations were particularly strong for atypical alleles and allelic combinations. For example, *modA11* and *modB2* were more commonly associated with invasive, rather than carriage, isolates (respectively, 29% vs. 13% for *modA11* ( $p < 0.0001$ ); 40% vs. 28% for *modB2* ( $p = 0.002$ ); Table 2). The *modA12-modB2-modD1* combination was associated with cc41/44 invasive isolates (20% invasive vs. 0% carriage,  $p < 0.0001$ ), while carriage isolates from this clonal complex possessed *modA12-modB1* but lacked *modD1* (87% carriage vs. 9% invasive with *modA12-modB1*,  $p < 0.0001$ ) (Fig. 3). The *modA12-modB4* combination was associated with cc213 invasive isolates and was only found in two carriage isolates of different sequence types (6.5% of invasive isolates vs. 1% carriage,  $p = 0.0004$ ). Also, the *modA11-modB1* combination was associated with invasive isolates, and present in hyperinvasive lineages cc32 and cc269 (22% of invasive vs. 2.4% carriage isolates,  $p < 0.0001$ ). The atypical *modA15* allele was usually found alone and in cc92 carriage isolates (12% of total carriage isolates vs. 0.2% of invasive isolates,  $p < 0.0001$ ).

## Discussion

The phase variable type III DNA methyltransferases encoded by the *modA*, *modB* and *modD* genes are a global control mechanism by which *N. meningitidis* can randomly alter the expression of distinct sets of genes, known as phasevarions<sup>29</sup>. Several alleles of each of the *mod* genes have been characterized, based on differences in the region encoding the DNA recognition domain (DRD), each of which methylates a different sequence and mediates the epigenetic regulation of different sets of genes<sup>28,30</sup>. Many isolates contain multiple *mod* genes (*i.e.*, have multiple phasevarions) which can phase vary independently. This enables the reversible induction of polygenetic phenotypes, which increases bacterial adaptability to distinct ecological environments and may affect invasive capacity. For example, the phase variable ON/OFF switching of meningococcal ModD1 alters resistance to oxidative stress<sup>27</sup>, while phase variation of ModA11 and ModA12 results in altered antibiotic resistance<sup>40</sup>. Similarly, previous studies of the gonococcal ModA13 phasevarion have shown that ON/OFF variants have distinct phenotypes for biofilm formation, resistance to antimicrobials, and survival in primary human cervical epithelial cells<sup>28</sup>.

Allele			Isolates with allele <sup>a</sup>	% Carriage (#) <sup>b</sup>	cc(s) <sup>c</sup>	% Invasive (#) <sup>b</sup>	cc(s) <sup>c</sup>	p-value <sup>d</sup>
<i>modA11</i>			454	13 (28)	18, 32, 53, 116, 269	29 (426)	23, 32, 41/44, 92, 213, 269	<0.0001
<i>modA12</i>			1,135	70 (148)	11, 18, 22, 41/44, 106, 116	67 (987)	11, 18, 22, 23, 41/44, 213, 269	0.3478
<i>modA15</i>			35	12 (25)	92	1 (10)		<0.0001
<i>modB1</i>			465	23 (49)	18, 32, 41/44, 269	28 (416)	18, 22, 32, 41/44, 269	0.1389
<i>modB2</i>			634	28 (60)	11, 18, 22	40 (574)	11, 18, 22, 41/44, 213, 269	0.0018
<i>modB4</i>			97	1 (2)	41/44	6.5 (95)	41/44, 213	0.0004
<i>modD1</i>			315	0 (0)		21 (315)	41/44	<0.0001
<i>modD6</i>			71	8 (17)	18, 22	3 (54)	18, 22, 41/44	0.0014
Allele combination								
modA	modB	modD	Isolates with combo <sup>a</sup>	% Carriage (#) <sup>b</sup>	cc(s) <sup>c</sup>	% Invasive (#) <sup>b</sup>	cc(s) <sup>c</sup>	p-value <sup>d</sup>
<i>modA11</i>	–	–	109	9.5 (20)	53, 116	6 (89)	23, 32, 92, 269	0.1743
<i>modA12</i>	–	–	282	19 (40)	22, 106, 116	17 (242)	22, 23, 106, 116, 213,	0.4326
<i>modA15</i>	–	–	28	12 (25)	92	0.2 (3)		<0.0001
<i>modA11</i>	<i>modB1</i>	–	317	2.4 (4)	32, 269	22 (313)	32, 269	<0.0001
<i>modA12</i>	<i>modB1</i>	–	103	15 (31)	41/44	5 (72)	18, 41/44, 269	<0.0001
<i>modA12</i>	<i>modB2</i>	–	257	22 (47)	11, 18	15 (210)	11, 18, 41/44, 213	0.057
<i>modA12</i>	<i>modB4</i>	–	94	1 (2)	41/44	6.5 (92)	41/44, 213	0.0004
<i>modA12</i>	<i>modB2</i>	<i>modD1</i>	286	0 (0)		20 (286)	41/44	<0.0001
<i>modA12</i>	<i>modB2</i>	<i>modD6</i>	45	3 (7)	22	3 (38)	18, 22, 41/44	0.6514

**Table 2. Distribution of individual alleles and allele combinations in carriage and invasive meningococci.**

<sup>a</sup>Total number in the dataset with each of the most frequent *modA*, *modB* and *modD* alleles, or allele combinations. (–) indicates the *mod* gene is absent. <sup>b</sup>% (total number) of carriage or invasive isolates with the *mod* allele or allele combination, calculated as a proportion of total carriage or invasive isolates, respectively. For these analyses, alleles/isolates were included if full allele data was available, and if the allele was considered to encode a potentially functional methyltransferases (*i.e.*, contained no point mutations or insertions/deletions (other than insertions/deletions in the phase variable repeat region); *e.g.*, an allele profile of *modA11* - *modB1::IS1301* was considered to only contain a putatively functional *modA11*). Total number of isolates included in analyses: 211 carriage and 1478 invasive isolates for allele distribution; 211 carriage and 1448 invasive isolates for allele combinations (30 invasive isolates removed due to incomplete allele combination data). <sup>c</sup>Predominant clonal complexes containing the allele or allele combination. Clonal complexes are only included if they contain >10 isolates in the database. Sequence types not assigned to a clonal complex are excluded. <sup>d</sup>p-value for the association of the given allele or allele combination with either carriage or invasive isolates, calculated using the Fisher's Exact test (two-tailed).

Our analysis of the distribution of *mod* genes and alleles in meningococci isolated from carriage and invasive disease revealed high levels of diversity in type III DNA methyltransferases and their respective DRD regions, and identified associations between *mod* alleles and clonal complexes. These clonal complexes included both hyperinvasive lineages, responsible for the majority of disease worldwide (*e.g.*, cc11, cc32, cc41/44, and cc269), as well as those that are rarely associated with invasive disease (*e.g.*, cc35, cc92, cc106 and cc116)<sup>7–10,41</sup>. While some alleles, such as *modA12*, were found in both hyperinvasive (cc11 and cc41/44) and non-invasive (cc106) lineages, some alleles were more commonly associated with one or the other. For example, in terms of hyperinvasive lineages, *modA11* was associated with cc32 isolates, *modB1* with cc269 and cc32 isolates, and *modD1* with cc41/44 isolates. This is consistent with previous studies conducted in smaller datasets<sup>27,28</sup>. The *modA11-modB1* combination was associated with hyperinvasive lineages cc32 and cc269; and the *modA12-modB2-modD1* combination with the hyperinvasive ST-41 cluster. On the other hand, the *modA15* allele was predominantly found in isolates of the cc92 lineage, which is a non-invasive lineage. Furthermore, the distribution of *mod* alleles relative to the disease outcome or phenotype of each isolate (*i.e.*, invasive or carriage) showed associations of atypical alleles and allelic combinations with invasive or non-invasive meningococci. For example, *modA11-modB1*, *modA12-modB2-modD1* and *modA12-modB4* combinations were associated with invasive isolates, while *modA15* or the *modA12-modB1* combination were associated with isolates from carriage collections. While the *mod* alleles could be considered a marker of clonal complex rather than infection outcome, it is important to note that there is not always a strict correlation between clonal complex and infection outcome. This is highlighted within the cc41/44 lineage that contains two central STs, where ST-41 meningococci are commonly associated with invasive disease, while ST-44 are typically associated with carriage<sup>10</sup>. The distribution of *modD1* in this clonal complex is specifically associated with invasive disease; *modD1* is present in 84% of cc41/44 invasive isolates (105/105 ST-41 and 203/259 other STs) compared to 0% of cc41/44 carriage isolates (0/8 ST-44 and 0/22 other STs). Future

experimental work will clarify whether the observed associations directly correspond to a difference in the phenotype and virulence of strains. An increased focus should also be placed on collecting and sequencing carriage isolates, as to date these are underrepresented in the available meningococcal isolate panels.

Several questions remain regarding the evolution of *mod* diversity, the frequency of *mod* allele mobilization, and whether the repertoire is being neutrally inherited, or selectively maintained. These data suggest that the more commonly found alleles (e.g., *modA11*, *modA12*, *modB1* and *modB2*), along with their allelic combinations, may have been acquired early in the evolution of *N. meningitidis*, and have been successively inherited by vertical transmission from progenitor cells that have differentiated into multiple clonal complexes. The rarer alleles, and differences in *mod* allele combinations, may have arisen from more recent horizontal gene transfer (HGT) and recombination events common to *Neisseria*<sup>13,38,39,41–43</sup>. For example, the discovery of single isolates containing the *modA2*, *modA4*, *modA6*, and two isolates with *modD3* alleles is consistent with HGT in *N. meningitidis*. Transfer and replacement of alleles is possibly enhanced by the structure of the *mod* genes, as the conserved flanking regions may facilitate homologous recombination and replacement of the central variable DRD. This has previously been reported for the *modA* gene, with sequence analysis suggesting some DRDs originated from other bacterial species<sup>33</sup>. Similarly, the distribution of several alleles (e.g., *modA2*, *modA6*, *modD3* identified in *N. meningitidis* for the first time) suggests horizontal mobilization of DNA from outside the species<sup>44–46</sup>, as these alleles have been previously identified in *H. influenzae*, *N. polysaccharaea* and *N. lactamica*<sup>27,29,33</sup>. In order to investigate the temporal and geographical nature of the associations seen, more diverse and long-term isolate collections are needed. However, it can be noted that the some associations, such as that seen between cc11 isolates and *modA12* and *modB1*, are consistent in all the collections over the time period, while others are more specific to certain locations but may reflect the invasive or carriage focus of the collections, for example, the cc92 carriage isolates associated with the *modA15* allele are largely isolates from 1993 from the Czech carriage study: whereas *modD1* is associated with invasive isolates from the UK and Australian collections.

The evolution and biological significance of the 15-nucleotide C-terminal deletion and the DNA repeat tract variants are unknown. The C-terminal deletion does not affect the well-described conserved motifs of type III Mods, such as the catalytic region (DPPY) and the S-adenosyl-L-methionine methyl donor-binding pocket (FXGXG) (See Fig. 1a)<sup>29,47,48</sup>, and *ModA12* has been shown by methylome analysis to be functional in strains with (M0579) or without (B6116/77) this deletion<sup>30</sup>. It is noteworthy, however, that the typical 5'-AGCC-3' *modA* repeat tract is recognized by the *ModD* methyltransferase DNA recognition domain. Previous studies propose that methylation within gene coding regions may alter transcription<sup>45</sup>: this may suggest that variations arise as a mechanism to circumvent methylation of the repeat tract.

Previous studies posit that meningococcal restriction-modification systems, such as the type III systems that the *mod* genes are part of, are clade associated, and maintain barriers to gene transfer between groups<sup>15</sup>; however, other studies suggest that genome recombination is more frequent than expected<sup>38</sup>, and that restriction-modification systems do not necessarily prevent genetic transfer, either because they are being mobilized themselves, or due to transient or inefficient function<sup>39</sup>. This latter point may be particularly true for the *mod* alleles, given that *Mod* is phase variable and the restriction enzyme is dependent on the presence of *Mod* in type III restriction-modification systems, and that inactivating mutations can be found in the cognate restriction enzymes in some systems<sup>28,49</sup>. If this is indeed the case, then the selective maintenance of the *mod* allele repertoire may be attributable to the epigenetic regulatory functions of *Mod*, as has previously been suggested to explain the dominance of the *modA12* allele in *N. meningitidis*<sup>33</sup>. To support this hypothesis, full characterization of strains, and the phasevarions regulated by *Mod* proteins is needed. These studies will clarify the significance of *mod* allele and clonal complex associations, and how ON/OFF switching affects the phenotype of *N. meningitidis*, especially in light of the fact that isolates can contain up to three distinct *mod* alleles and phasevarions. Given that each *mod* allele can phase vary independently, each isolate can give rise to eight possible distinct sub-populations. These variants would appear genetically identical, differing only in the loss or gain of a few repeat units in the *mod* gene, but would have profound differences in gene expression and potentially in their invasive phenotype. It is also important to note that the potential fluidity of allelic exchange of *mod* alleles means that with a single gene recombination event, *N. meningitidis* strains may acquire a new *mod* allele and the ability to regulate genes in a different phasevarion, and consequently display a different phenotypic profile.

In terms of *mod* allele associations with invasive meningococci, it is important to note that disease does not increase the fitness of meningococci as there is no corresponding benefit to transmission<sup>50</sup>. However, if certain *Mod* proteins do increase the invasiveness of certain clonal complexes, the fact that *Mod* expression is phase variable would allow the *Mod* regulatory system to be maintained. For example, even under conditions where *Mod* ON is deleterious or increases the probability of invasion, cells with *Mod* in phase OFF may persist, in which case, selective pressure against *mod* is removed without affecting the future function of the *Mod* regulatory system or bacterial fitness. Accordingly, these alleles provide a mechanism that enables cells to variably express complex phenotypes, which may facilitate a transition from carriage to invasion and *vice versa* depending on environmental changes. If so, this may help elucidate how transmissibility and virulence are linked in meningococcal lineages<sup>43</sup>. Hence, the study of these regulators, and how they change over time, may provide critical insights into how and why *N. meningitidis* cells generate distinct phenotypes without overt changes in gene content, and how this may mediate transition from carriage to invasive disease. While it is tempting to speculate on the ON vs. OFF status of the *Mods* from the available genome sequences, it is important to note that the samples used for sequencing were not collected for this purpose and are therefore not an accurate reflection of the natural ON/OFF status and ratio of the *in vivo* bacterial population. Therefore, unlike past epidemiological studies that typically isolate and characterize single meningococcal colonies from patient samples, future studies will require direct and unbiased sequencing, or the isolation of representative populations, of bacteria from blood, cerebrospinal fluid and the nasopharynx. The characterization of the presence and expression state of individual *mod* alleles in these samples will help determine the significance of *mod* allele distribution relative to meningococcal carriage and disease.



## Methods

***N. meningitidis* isolate collections.** Four *N. meningitidis* isolate collections were analyzed in this study: i) 20 previously characterized whole-genome sequences (WGS) from 18 disease and two carriage isolates<sup>12,15,25,51–53</sup>; ii) WGS from 54 disease and 209 carriage isolates from the Czech Republic<sup>41,43</sup>; iii) WGS data from 1380 disease isolates in the Meningitis Research Foundation (MRF) Meningococcus Genome Library (MGL) ([www.meningitis.org/current-projects/genome](http://www.meningitis.org/current-projects/genome))<sup>62</sup>; and, iv) 50 disease isolates from Australia<sup>54</sup>.

**Screening for *mod* genes.** The *mod* genes were identified by bioinformatic analysis of WGS data using the BIGSdb platform hosted on [pubmlst.org/neisseria](http://pubmlst.org/neisseria)<sup>55</sup>. Locus records for *modA* (NEIS1310), *modB* (NEIS1194) and *modD* (NEIS2364) were generated in the PubMLST database (<http://pubmlst.org/neisseria>)<sup>55</sup> using previously identified reference *modA* and *modB* nucleotide sequences from *N. meningitidis* MC58<sup>28,53</sup> and the *modD* sequence from *N. meningitidis* M0579<sup>15,27</sup>. The BIGSdb ‘Scan Tag’ tool<sup>56</sup> was implemented for the identification of *mod* loci within WGS data of each isolate, returning BLAST matches greater than 30% alignment and 50% identity to the sequences stored in locus records. The nucleotide diversity of *mod* loci required that hits be exported to MEGA6<sup>57</sup> for manual alignment with previously identified (see reference alleles below) full length *mod* coding sequences and DNA recognition domain (DRD) based alleles. For example, alignment gaps were inserted in the phase variable repeat region of phase OFF sequences to allow comparison with the full-length translated amino acid sequence. Trimmed sequences were uploaded to the appropriate locus record for storage where they were assigned unique numeric identifiers (allele id numbers), and grouped into variants based on the DRD (corresponding to the *mod* alleles described below). Database alleles were flagged with information such as phase-variation status and where the open reading frame was interrupted due to insertions, deletions or point mutations (other than changes to the number of phase variable repeat units), and database alleles were assigned the flags ‘internal stop codon’ and ‘frameshift’ where necessary. These interrupted alleles were not included in the analysis of invasive versus carriage meningococci. The BIGSdb ‘autotagger’ and ‘autodefiner’ tools<sup>56</sup> were then used to identify *mod* genes within genomic data stored for each isolate, allowing tagging of nucleotide positions and database allele id numbers.

The *mod* loci were frequently present on different contiguous sequences of a genome assembly due to break points within phase variable regions or insertion sequences: this permitted identification of the presence of a *mod* gene and its allele, but not assignment of a database allele id number. In these cases, a database flag was inserted at these genomic positions to indicate a partial assembly, and isolates were considered to contain *mod* genes but were not included in subsequent allele analyses. This process was performed iteratively until no new alleles of the three *mod* genes could be identified in the genomes; at this stage, genomes without tagged *mod* genes were investigated using implementations of the BIGSdb BLAST tool (parameters: word size 11; reward 2; penalty -3; gap open 5; gap extend 2). For each locus variant, six hits were investigated per isolate, regardless of E-value significance, and 1000 bp of flanking sequence were extracted with the hit for investigation in MEGA6; new alleles were uploaded to the database as before, otherwise genomes were tagged as ‘missing’ *mod* loci.

To screen for *mod* genes and alleles in the Australian disease isolates (no WGS available), PCR and sequencing analyses was performed as previously described<sup>27,28</sup>. MLST analysis was performed in accordance with the scheme guidelines (<http://pubmlst.org/neisseria/info/>)<sup>55</sup>.

Sequence alignments were performed using ClustalW<sup>58</sup> and were exported into Jalview<sup>59</sup> to generate alignment Figures. The Neighbor-Net graph was generated in SplitsTree4<sup>60</sup> from allelic distances among the 49 rMLST loci of unique ribosomal Sequence Types (rSTs) ( $n = 639$ )<sup>61</sup> among clonal complexes comprising >10 isolates in the dataset. rSTs (available at <http://pubmlst.org>) were annotated with the combination of *mod* alleles present in isolates, and clusters were labeled according to the clonal complexes of these isolates. 186 isolates were not assigned to a known clonal complex.

Statistical analyses were carried out using 2-tailed Fisher’s Exact tests (Graphpad Software Inc., San Diego, CA, USA), with p-values of  $p < 0.05$  taken to indicate statistical significance.

***mod* gene/allele reference sequences.** The *modA*, *modB*, or *modD* genes were defined by >90% sequence identity along the length of the deduced amino acid sequence, excluding the variable DRD, to the *modA11* and *modB1* genes of *N. meningitidis* MC58<sup>28</sup>, or the *modD1* gene of M0579<sup>27</sup>. The alleles of each *mod* gene were defined by  $\geq 95\%$  amino acid identity across the DRD to the reference sequences listed below.

For *modA*, reference sequences were as previously described<sup>28,33</sup>: *modA2*, *H. influenzae* (*Hi*) strain 723; *modA4*, *Hi* 3579; *modA6*, *Hi* C1626; *modA11*, *N. meningitidis* (*Nm*) MC58; *modA12*, *Nm* Z2491; *modA15*, *Hi* R3570; *modA18*, *Nm* NGE28; *modA19*, *Nm* 053422. For *modA* allele alignments, sequences from the following isolates were used for the full-length gene reference: *modA2*, *Hi* 86-028NP; *modA4*, *Hi* R2846; *modA6*, *Hi* PittEE. For *modA15* and *modA18*, no full-length genes were available in GenBank, and isolates M12 240232 and M11 240002 from the MRF MGL have been used, respectively.

For *modB*, reference sequences were as previously described<sup>29</sup>: *modB1*, *Nm* MC58; *modB2*, *Nm* Z2491; *modB3*, *N. lactamica* (*Nl*) 020-60. For *modB4*, *Nm* M01-240355<sup>15</sup>. For newly described *modB5* and *modB6*, GenBank database matches were identified in *N. polysaccharea* 43768 (NEIPOLOT\_01008) for *modB5*; and *Nm* 81858 (NM81858\_1294) for *modB6*.

For *modD*, reference sequences were as previously described<sup>27</sup>: *modD1*, *Nm* M0579; *modD3*, *Nl* ST640; *modD4*, *N. cinerea* 14655; *modD5*, *N. mucosa* 25996. For *modD2*, *Nm* 61103 (NM61103\_0875). For *modD6*, *Nm* 6938<sup>38</sup>. For *modD7*, the Czech isolate *Nm* 0001/93<sup>43</sup> was used from the MRF MGL as no similar sequences were identified in GenBank.

## References

1. Caugant, D. A. & Maiden, M. C. J. Meningococcal carriage and disease—Population biology and evolution. *Vaccine* **27**, Supplement 2, B64–B70 (2009).
2. Hill, Darryl J., Griffiths, Natalie J., Borodina, E. & Virji, M. Cellular and molecular biology of *Neisseria meningitidis* colonization and invasive disease. *Clin. Sci.* **118**, 547–564 (2010).
3. Rosenstein, N. E., Perkins, B. A., Stephens, D. S., Popovic, T. & Hughes, J. M. Meningococcal Disease. *New Engl. J. Med.* **344**, 1378–1388 (2001).
4. Emonts, M., Hazelzet, J. A., de Groot, R. & Hermans, P. W. M. Host genetic determinants of *Neisseria meningitidis* infections. *Lancet Infect. Dis.* **3**, 565–577 (2003).
5. Dale, A. P. & Read, R. C. Genetic susceptibility to meningococcal infection. *Expert Rev. Anti Infect. Ther.* **11**, 187–199 (2013).
6. Jafri, R. *et al.* Global epidemiology of invasive meningococcal disease. *Popul. Health Metr.* **11**, 17 (2013).
7. Maiden, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**, 3140–3145 (1998).
8. Maiden, M. C. J. Population genomics: diversity and virulence in the *Neisseria*. *Curr. Opin. Microbiol.* **11**, 467–471 (2008).
9. Brehony, C., Jolley, K. A. & Maiden, M. C. J. Multilocus sequence typing for global surveillance of meningococcal disease. *FEMS Microbiol. Rev.* **31**, 15–26 (2007).
10. Yazdankhah, S. P. *et al.* Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J. Clin. Microbiol.* **42**, 5146–5153 (2004).
11. Roupahel, N. G. & Stephens, D. S. *Neisseria meningitidis*: biology, microbiology, and epidemiology. *Methods Mol. Biol.* **799**, 1–20 (2012).
12. Schoen, C. *et al.* Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. USA* **105**, 3473–3478 (2008).
13. Marri, P. R. *et al.* Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS ONE* **5**, e11835 (2010).
14. Snyder, L. & Saunders, N. The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as ‘virulence genes’. *BMC Genomics* **7**, 128 (2006).
15. Budroni, S. *et al.* *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl. Acad. Sci. USA* **108**, 4494–4499 (2011).
16. Bratcher, H., Corton, C., Jolley, K., Parkhill, J. & Maiden, M. A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* **15**, 1138 (2014).
17. Claus, H., Friedrich, A., Frosch, M. & Vogel, U. Differential distribution of novel restriction-modification systems in clonal lineages of *Neisseria meningitidis*. *J. Bacteriol.* **182**, 1296–1303 (2000).
18. Harrison, O. B. *et al.* Epidemiological evidence for the role of the hemoglobin receptor, HmbR, in meningococcal virulence. *J. Infect. Dis.* **200**, 94–98 (2009).
19. Bille, E. *et al.* Association of a bacteriophage with meningococcal disease in young adults. *PLoS ONE* **3**, e3885 (2008).
20. Bille, E. *et al.* A chromosomally integrated bacteriophage in invasive meningococci. *J. Exp. Med.* **201**, 1905–1913 (2005).
21. Joseph, B. *et al.* Comparative genome biology of a serogroup B carriage and disease strain supports a polygenic nature of Meningococcal virulence. *J. Bacteriol.* **192**, 5363–5377 (2010).
22. Schoen, C., Kischkies, L., Elias, J. & Ampattu, B. J. Metabolism and virulence in *Neisseria meningitidis*. *Front. Cell. Infect. Microbiol.* **4**, 114 (2014).
23. Mellin, J. R. & Hill, S. In *Neisseria: Molecular Mechanisms of Pathogenesis* (eds Caroline Attardo Genco & Lee Wetzler) Ch. 1, 3–18 (Caister Academic Press, 2010).
24. Snyder, L. A. S., Butcher, S. A. & Saunders, N. J. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria spp.* *Microbiology* **147**, 2321–2332 (2001).
25. Bentley, S. D. *et al.* Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet.* **3**, e23 (2007).
26. Martin, P. *et al.* Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: evidence for a novel mechanism of phase variation. *Mol. Microbiol.* **50**, 245–257 (2003).
27. Seib, K. L. *et al.* A novel epigenetic regulator associated with the hypervirulent *Neisseria meningitidis* clonal complex 41/44. *FASEB J.* **25**, 3622–3633 (2011).
28. Srikhanta, Y. N. *et al.* Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. *PLoS Path.* **5**, e1000400 (2009).
29. Srikhanta, Y. N., Fox, K. L. & Jennings, M. P. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiol.* **8**, 196–206 (2010).
30. Seib, K. L. *et al.* Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6-adenine DNA methyltransferases of *Neisseria meningitidis*. *Nucleic Acids Res.* **43**, 4150–4162 (2015).
31. Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M. & Jennings, M. P. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc. Natl. Acad. Sci. USA* **102**, 5547–5551 (2005).
32. Adamczyk-Popławska, M., Lower, M. & Piekarczyk, A. Characterization of the NgoAXP: phase-variable type III restriction-modification system in *Neisseria gonorrhoeae*. *FEMS Microbiol. Lett.* **300**, 25–35 (2009).
33. Gawthorne, J. A., Beatson, S. A., Srikhanta, Y. N., Fox, K. L. & Jennings, M. P. Origin of the diversity in DNA recognition domains in phasevarion associated modA genes of pathogenic *Neisseria* and *Haemophilus influenzae*. *PLoS ONE* **7**, e32337 (2012).
34. Fox, K. L., Srikhanta, Y. N. & Jennings, M. P. Phase variable type III restriction-modification systems of host-adapted bacterial pathogens. *Mol. Microbiol.* **65**, 1375–1379 (2007).
35. Attack, J. M. *et al.* A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat. Commun.* **6**, 7828 (2015).
36. Blakeway, L. V. *et al.* ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phasevarions in otitis media. *FASEB J.* **28**, 5197–5207 (2014).
37. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **126**, 847–857 (1978).
38. Hao, W. *et al.* Extensive genomic variation within clonal complexes of *Neisseria meningitidis*. *Genome Biol. Evol.* **3**, 1406–1418 (2011).
39. Kong, Y. *et al.* Homologous recombination drives both sequence diversity and gene content variation in *Neisseria meningitidis*. *Genome Biol. Evol.* **5**, 1611–1627 (2013).
40. Jen, F. E., Seib, K. L. & Jennings, M. P. Phasevarions mediate epigenetic regulation of antimicrobial susceptibility in *Neisseria meningitidis*. *Antimicrob. Agents Chemother.* **58**, 4219–4221 (2014).
41. Jolley, K. A. *et al.* Carried meningococci in the Czech Republic: a diverse recombining population. *J. Clin. Microbiol.* **38**, 4492–4498 (2000).
42. Putonti, C., Nowicki, B., Shaffer, M., Fofanov, Y. & Nowicki, S. Where does *Neisseria* acquire foreign DNA from: an examination of the source of genomic and pathogenic islands and the evolution of the *Neisseria* genus. *BMC Evol. Biol.* **13**, 184 (2013).

43. Jolley, K. A., Wilson, D. J., Kriz, P., Mcvean, G. & Maiden, M. C. J. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol. Biol. Evol.* **22**, 562–569 (2005).
44. Furuta, Y. & Kobayashi, I. Movement of DNA sequence recognition domains between non-orthologous proteins. *Nucleic Acids Res.* **40**, 9218–9232 (2012).
45. Furuta, Y. *et al.* Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* **10**, e1004272 (2014).
46. Kroll, J. S., Wilks, K. E., Farrant, J. L. & Langford, P. R. Natural genetic exchange between *Haemophilus* and *Neisseria*: Intergeneric transfer of chromosomal genes between major human pathogens. *Proc. Natl. Acad. Sci. USA* **95**, 12381–12385 (1998).
47. Ahmad, I. & Rao, D. N. Functional Analysis of Conserved Motifs in EcoP151 DNA Methyltransferase. *J. Mol. Biol.* **259**, 229–240 (1996).
48. Malone, T., Blumenthal, R. M. & Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618–632 (1995).
49. Fox, K. L. *et al.* *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic Acids Res.* **35**, 5242–5252 (2007).
50. Buckee, C. O. *et al.* Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. USA* **105**, 15082–15087 (2008).
51. Parkhill, J. *et al.* Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506 (2000).
52. Peng, J. *et al.* Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone. *Genomics* **91**, 78–87 (2008).
53. Tettelin, H. *et al.* Complete genome sequence of *Neisseria meningitidis* serogroup B Strain MC58. *Science* **287**, 1809–1815 (2000).
54. Berrington, A. W. *et al.* Phase variation in meningococcal lipooligosaccharide biosynthesis genes. *FEMS Immunol. Med. Microbiol.* **34**, 267–275 (2002).
55. Jolley, K. & Maiden, M. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595 (2010).
56. Jolley, K. A. & Maiden, M. C. Automated extraction of typing information for bacterial pathogens from whole genome sequence data: *Neisseria meningitidis* as an exemplar. *Euro Surveill.* **18**, 20379 (2013).
57. Tamura, K. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
58. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
59. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
60. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
61. Jolley, K. A. *et al.* Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid Web-based analysis methods. *J. Clin. Microbiol.* **50**, 3046–3053 (2012).
62. Hill, D. M. *et al.* Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. *Lancet Infect Dis.* **15**, 1420–1428 (2015).

## Acknowledgements

This publication made use of: 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, and funded by the Meningitis Research Foundation; and PubMLST (<http://pubmlst.org/neisseria/info/>), developed by Keith Jolley, sited at the University of Oxford, and funded by the Wellcome Trust and European Union). This work was supported by the Australian National Health and Medical Research Council (NHMRC) [NHMRC Early Career Fellowship to Y.N.S.; Program Grants 565526 and 1071659 to M.P.]; Project Grant 1021631 and Career Development Fellowship 1045235 to K.L.S.], The Meningitis Research Foundation [D.M.C.H.] and the Wellcome Trust [M.C.J.M].

## Author Contributions

K.L.S., M.P.J. and M.C.J.M. designed the study. A.T., D.M.C.H., O.B.H. and Y.S. provided data. A.T., D.M.C.H. and K.L.S. wrote the draft manuscript. All authors analyzed the data, and contributed to the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Tan, A. *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.* **6**, 21015; doi: 10.1038/srep21015 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>