BMC Microbiology

Research article

Diversity in coding tandem repeats in related *Neisseria* spp.

Philip Jordan†, Lori AS Snyder† and Nigel J Saunders*

Address: Bacterial Pathogenesis and Functional Genomics Group, The Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford. OX1 3RE, UK

Email: Philip Jordan - philip.jordan@pathology.oxford.ac.uk; Lori AS Snyder - lori.snyder@pathology.oxford.ac.uk; Nigel J Saunders* - nigel.saunders@pathology.oxford.ac.uk

* Corresponding author †Equal contributors

Published: 12 November 2003

BMC Microbiology 2003, **3**:23

Received: 06 August 2003
Accepted: 12 November 2003

This article is available from: http://www.biomedcentral.com/1471-2180/3/23

Abstract

Background: Tandem repeats contained within coding regions can mediate phase variation when the repeated units change the reading frame of the coding sequence in a copy number dependent manner. Coding tandem repeats are those which do not alter the reading frame with copy number, and the changes in copy number of these repeats may then potentially alter the function or antigenicity of the protein encoded. Three complete neisserial genomes were analyzed and compared to identify coding tandem repeats where the number of copies of the repeat will have some structural consequence for the protein. This is the first study to address coding tandem repeats that may affect protein structures using comparative genomics, combined with a population survey to investigate which show interstrain variability.

Results: A total of 28 genes were identified. Of these, 22 contain coding tandem repeats that vary in copy number between the three sequenced strains, three strain specific genes were included for investigation on the basis of having >90% identity between repeated units, and three genes with repeated elements of >250 bp were included although no length variations were seen in the genomes. Amplification, and sequencing of repeats showing altered copy number, of these 28 coding tandem repeat containing regions, from a set of largely unrelated strains, revealed further repeat length variation in several cases.

Conclusion: Eighteen genes were identified which have variation in repeat copy number between strains of the same species, twelve of which show greater diversity in repeat copy number than is present in the sequenced genomes. In some cases, this may reflect a mechanism for the generation of antigenic variation, as previously described in other species. However, some of the genes identified encode proteins with cytoplasmic functions, including sugar metabolism, DNA repair, and protein production, in which repeat length variation may have other functions. Coding tandem repeats appear to represent a largely unexplored mechanism of generating diversity in the *Neisseria* spp.

Background

Variable copy number tandem repeats have been observed in a number of prokaryotic genomes [1,2]. These are adjacent sequences that are directly repeated, the repeated units of which may be identical or partially degenerate. Coding tandem repeats are those tandem repeats that are completely contained within a coding sequence and are composed of repeated units in which copy number will

not disrupt the reading frame. Therefore, all coding tandem repeats have repeated units composed of 3 bp or multiples of 3 bp. These are distinct from intergenic repeats and from repeats such as those that mediate phase variation. There are many examples in which variation in copy number within coding tandem repeats has been shown to affect virulence and alter the ability of antibodies to bind to bacterial antigens. In *Streptococcus agalactiae*, there is a reduction in copies of a coding tandem repeat within the α C-protein from the same strain isolated from mother and neonate [3]. The proteins with deleted repeat units are no longer recognised by anti-α C-protein antibodies, and repeat deletion escape mutants can be generated with enhanced pathogenicity in immune mice [4]. These repeats share similarity with other streptococcal sequences in the Rib and Esp proteins, which also vary in the length of coding tandem repeats between strains [5-7]. Tandem repeated structures in the group A streptococcal M proteins, which are extensively studied virulence determinants, vary in length due to intragenic homologous recombination events [8,9]. Size variation in surface proteins Lmp1 and Lmp3 of *Mycoplasma hominis* has been correlated to tandem repeats at the C-terminal end of the proteins and contributes to immune evasion through antigenic variation [10]. In *Mycoplasma hyorhinis*, immune escape variants of the Vlp proteins are generated through intragenic recombination between the C-terminal coding tandem repeat region in homologues *vlpA*, *vlpB*, and *vlpC* [11,12]. Also, there is evidence that repeat epitopes can influence the overall antigenicity of proteins, as well as the availability of epitopes. For example, addition of tandem repeats in the PAc protein of *Streptococcus mutans*, which normally contains three long repeated regions, induces higher antibody production than the native peptide [13].

In the *Neisseria* spp., variable copy number coding tandem repeats have been observed previously only in PilQ [14], and DcaC [15], while different copy numbers of a coding tandem repeat have been reported separately for Lip / H.8 [16,17]. Although the functional consequences of these variations have yet to be determined, this is a potentially important mechanism of adaptation available to these species. A comprehensive analysis to identify genes in which potentially functional variation of this type occurs has not previously been performed in the *Neisseria* or any other bacterial species. In this study, comparisons of the complete genomes of *N. gonorrhoeae* strain FA1090, and *N. meningitidis* strains MC58 and Z2491 were conducted to identify all coding tandem repeats, and to identify which of these varied in copy number between the sequenced strains. Upon its availability, the *N. meningitidis* strain FAM18 genome sequence was added to this analysis. The coding tandem repeats were further investigated in a small diverse collection of strains, to extend the genome-based observations, and to determine which genes are likely to be undergoing functional variation of this type. A range of genes with potentially functionally important diversity in repeat encoded structures was identified.

Results and Discussion

Coding regions identified as containing coding tandem repeats

The three available complete neisserial genome sequences [18-20] were compared to identify genes containing coding tandem repeats associated with variation in the copy number of the repeated units. Each tandem repeat was evaluated to determine whether the entirety of the repeat is located within the predicted coding sequence and that it does not alter the reading frame. Tandem repeats that did not meet these criteria are not coding tandem repeats and as such were not investigated. Twenty-two genes were identified (Table 3), including: *pilQ* [14], and *dcaC* [15], in which diversity in the coding tandem repeats were reported previously, and Lip / H.8 antigen [16,17], in which these two publications report different copy numbers of the coding tandem repeat in the single gene addressed. In addition, 2 genes only present in *N. gonorrhoeae* strain FA1090 (TR23, XNG0938 & TR25, XNG0481) and 1 gene only present in *N. meningitidis* strain MC58 (TR24, NMB1848) were included for further investigation, each having >90% identity between the repeated units (Table 3). Although these could not be assessed for differences in copy number of the coding tandem repeats between the genome sequenced strains, it was felt that due to the high degree of identity between the repeated units they should be further investigated to determine if diversity exists. A further 3 genes (TR26-TR28) were included on the basis of tandem repeats composed of repeated units of greater than 250 bp, although the copy numbers for these did not differ between the sequenced strains (Table 3). Although outside the primary criteria of this study, the unusually long nature of the coding tandem repeated units lead to the inclusion of these three genes for investigation here, to assess if diversity in copy number in such repeats exists. The repeated elements within the coding tandem repeats in the selected candidate genes ranged in size from 6 bp to 273 bp (Table 4).

These 28 genes were assessed using PCR in 11 neisserial strains to identify additional diversity in coding tandem repeat copy numbers. These 11 strains were chosen on the basis of previously observed diversity in repeat copy numbers of *dcaC* [15], and included 6 *N. meningitidis*, 3 *N. lactamica*, and 2 *N. gonorrhoeae* strains (Table 1). *N. meningitidis* strain MC58 was used as a positive and size control in the PCR. The previous *dcaC* study revealed no variability in tandem copy number between the *N. gonorrhoeae* strains studied. For the 2 gonococcus specific genes,

Table 1: *Neisseria* spp. strains used in this study.

Strain set for tandem repeat PCRs	Strain identifier used in Table 4
N. meningitidis serogroup B strain MC58	MC58
N. meningitidis serogroup B strain 44/76	44/76
N. meningitidis serogroup B strain NGE30	NGE30
N. meningitidis serogroup B strain BZ133	BZ133
N. meningitidis serogroup A strain 92001	92001
N. meningitidis serogroup C strain 94/155	94/155
N. meningitidis serogroup W135 strain A22	A22
N. lactamica strain L18	L18
N. lactamica strain L12	L12
N. lactamica strain L22	L22
N. gonorrhoeae strain FA19	FA19
N. gonorrhoeae strain 26034	26034
Strain set for *N. gonorrhoeae*-specific genes	
N. gonorrhoeae strain FA1090	FA1090
N. gonorrhoeae strain MS11	MS11
N. gonorrhoeae strain FA19	FA19
N. gonorrhoeae strain F62	F62
N. gonorrhoeae strain 25534	25534
N. gonorrhoeae strain 28539	28539
N. gonorrhoeae strain 26034	26593
N. gonorrhoeae strain 26241	26241
N. gonorrhoeae strain 27921	27921
N. gonorrhoeae strain 29528	29528
N. gonorrhoeae strain 28516	28516
N. gonorrhoeae strain 27806	27806

11 *N. gonorrhoeae* strains were analyzed, using strain FA1090 as a positive and size control.

Primers were designed flanking the tandem repeats such that PCR product size could be used to determine the number of copies of the coding tandem repeated unit. In the case of TR19 (*tonB*), the gene contains 2 tandem repeats, which were addressed separately (TR19a and TR19b). In the case of TR5 (*pilQ*) a compound tandem repeat is present, such that the 5' 24 bp of the 66 bp tandem repeat is then repeated itself as a 24 bp tandem repeat immediately following the 66 bp repeat (Figure 1). Therefore, TR5 was evaluated by sequencing in all strains. Additional sequencing was done for all of the products where the size of the PCR product suggested that the length of the tandemly repeated region might differ from the sequenced strains. In all, over 200 sequencing reactions were conducted to ascertain the sequence of the coding tandem repeat containing region(s) of the 28 coding sequences.

Observed differences between coding repeat lengths

Of the 28 genes containing coding tandem repeats, 6 were found to have differences in the number of coding tandem repeats that appear to divide along species lines in the limited strain collection used (Table 4; TR9, TR14, TR17, TR18, TR19, TR21). There was no length variation in one of the two *N. gonorrhoeae* specific genes (Table 4; TR25), nor in the three additional genes included in the study based on the length of the repeat (>250 bp) (Table 4; TR26, TR27, TR28), suggesting that these long repeats are comparatively stable. Six of the genes displayed no additional length differences to that seen in the sequenced strains (4; TR1, TR3, TR6, TR7, TR12, TR16). Each of these had relatively few copies of the repeat (1 or 2, 2 or 3, or 1 or 3), whereas those which show additional variation to that seen in the genome sequence comparisons tended to have more copies of the repeated unit.

Of the 28 genes selected as potentially containing length-varying coding tandem repeats, 12 were found to have additional differences in copy number between neisserial strains (Table 4; TR2, TR4, TR5, TR8, TR10, TR11, TR13, TR15, TR20, TR22, TR23, TR24). *dcaC* (TR20) was not further investigated, having previously been assessed in these strains [15].

Predicted and known surface proteins with coding repeat copy number variation suggesting antigenic variation

The presence of coding tandem repeats within the genes encoding surface proteins has been recognized in other species as a mechanism of antigenic variation mediated by changes in the number of repeats [3-12]. In these cases, changes in the number of tandem repeat copies alters the

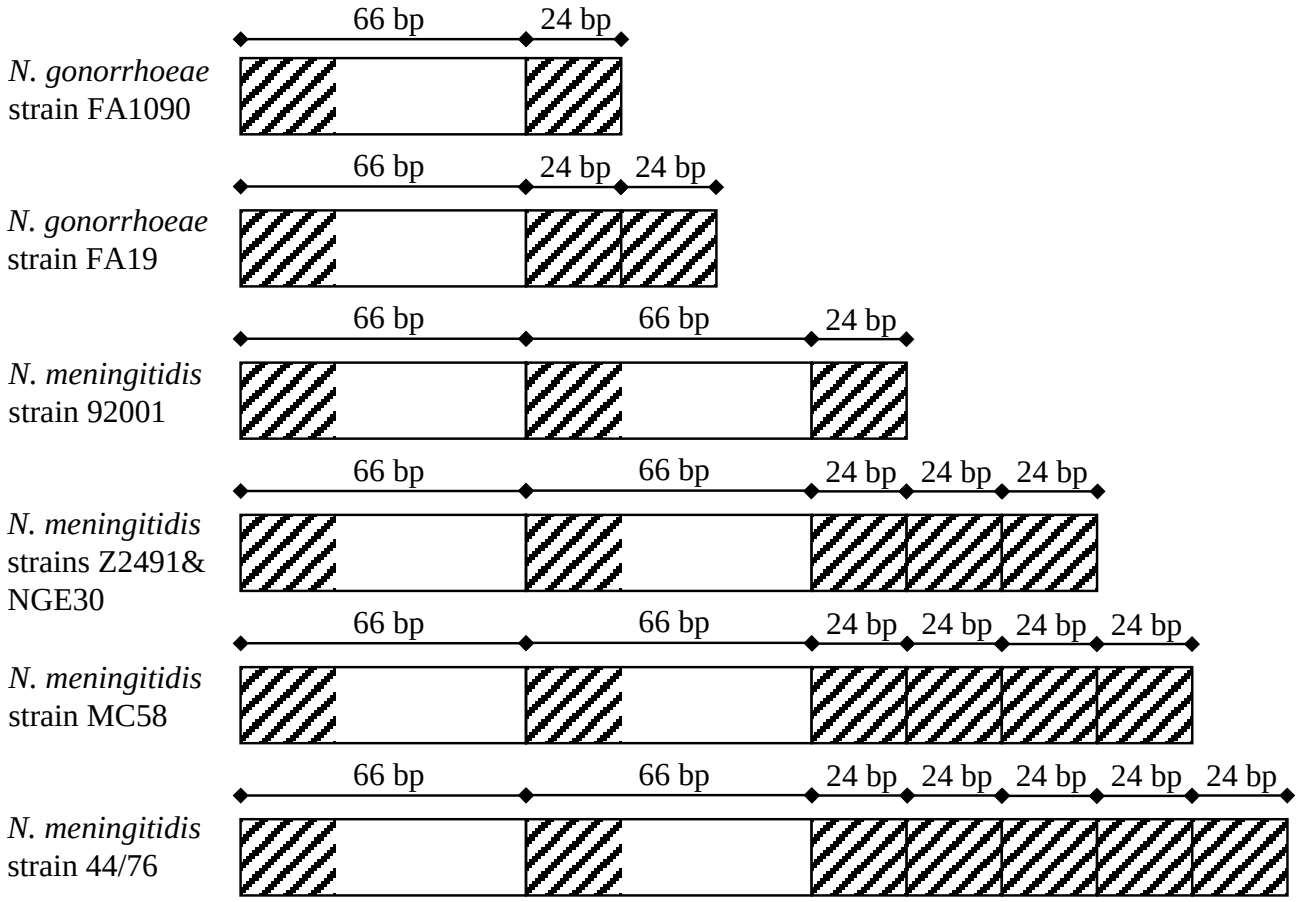

Figure 1
Two consecutive tandem repeat elements exist in pilQ (TR5). The first repeated unit is 66 bp. The first 24 bp of this 66 bp repeat is homologous to the second repeated unit of 24 bp. Both repeats in this compound coding tandem repeat are present in different lengths in the strains.

protein epitopes and presumably offers some benefit to the organism through immune evasion. This process has not been directly demonstrated in the pathogenic *Neisseria* spp., nor has a detailed study of any bacterial genome been conducted in an attempt to identify the repertoire of coding tandem repeats within a strain. This is, therefore, the first report of its kind, and additionally includes data related to genomic comparisons, diverse strain analysis, and sequencing of new copy number difference in the identified coding tandem repeats.

Several of the genes identified in this study are either known to be surface proteins, or are predicted to be surface exposed. Of the 28 genes investigated here, twelve of these are outer membrane proteins, or are predicted to be surface associated (TR1, TR4, TR5, TR7, TR11, TR13, TR14, TR15, TR16, TR20, TR21, TR26). For comparison, analysis of the complete genome of *N. meningitidis* strain MC58 [20] predicted 570 putative surface-exposed proteins out of 2158 annotated features [21]. Six of the 12 genes identified here contained tandem repeat copy numbers that differed from those of the sequence strains (TR4, TR5, TR11, TR13, TR15, TR20). Additionally, two genes are predicted to be cytoplasmic proteins, which are antigenic in other species (TR2 & TR8). This does not necessarily mean these two CDSs encode surface proteins, which is why they are included in a separate section (*Cytoplasmic proteins with variable numbers of coding tandem repeats, which may also be antigenic surface proteins*), but likewise it is possible that these proteins are surface exposed. Overall, half of the genes identified that contain

coding tandem repeats (14 of 28) may be surface exposed proteins.

A potential vaccine candidate

NMB2001 (TR4), a protein with some homology with the p60 invasin from *Listeria monocytogenes* [22], has been identified as a potential vaccine candidate from the study based upon the genome sequencing project [21]. It has been determined to be surface exposed and available for antibody binding. The presence of a tandem repeat was referred to in this paper, but length variation was not described. The 29 amino acid repeat encoded in this protein, which constitutes the majority of the N-terminal portion of the protein, is variable among both the *N. meningitidis* and *N. gonorrhoeae* strains tested (Table 4). It is possible that changes in the coding tandem repeat copy number may alter the antigenicity of the protein, which could complicate its use in any new vaccine.

The compound tandem repeat of pilQ

A compound tandem repeat was identified in *pilQ*, composed of a repeat of 66 bp followed by one of 24 bp, the latter being similar to the 5' 24 bp of the 66 bp repeat (Figure 1). These repeats have been described and studied previously, with a slightly different description of the repeat structure [14]. PilQ forms a dodecameric pore in the neisserial outer membrane, through which the pilus extends from the periplasm to the extracellular space [23]. It is not known how the changes in repeat numbers might affect the protein::protein interactions in the dodecamer, or where these repeats are located in the pore structure. Strain variability in this study is similar to that described previously, with one or two copies of the 66 bp repeat, and one to five copies of the 24 bp repeat, with a notable exception. In *N. meningitidis* strain FAM18, there are no complete copies of either the 66 bp or 24 bp repeats (Table 4). This is due to a deletion in the gene from 50 bp into what would be the first copy of the 66 bp repeat to 361 bp after the end of the tandem repeat containing region. This large deletion in *pilQ* generates a frame-shift mutation in this strain and deletes the site of annealing of the PCR primer TR5R, therefore TR5Rv2 was also designed. A large deletion (303 bp) which comprises a portion of a tandem repeat as well as non-repeated genic sequence is also seen in NMB2050 (TR26), although in this case the deletion does not generate a frame-shift. Deletions associated with the tandem repeats in *pilQ* have not previously been reported. If the compound tandem repeat containing portion of the PilQ protein represents exposed epitopes, then the variation in the tandem repeat structure may be involved in antigenic variation. It has also been suggested that changes in the repeat alter the assembly of pilin in the context of variations in PilE and/or PilC expression [14].

Lipoproteins and putative lipoproteins

Annotated as a hypothetical protein in *N. meningitidis* strain MC58, TR11 (NMB1333), is predicted by PSORT to be an outer membrane or periplasmic protein. The NCBI Conserved Domain Search reveals that this CDS contains a sequence with homology to the Peptidase family M23/M37, which in addition to the eukaryotic proteins of the family, includes bacterial lipoproteins that have no peptidase activity. The most 5' of the two tandem repeat sequences present in the gene, a repeat composed of 9 copies of a 21 bp element, does not display inter-strain differences in length, although in *N. gonorrhoeae* strains FA1090 and FA19 copy 7 has 9 bp deleted. The second repeat within this gene is a 15 bp (5 amino acid) tandem repeat present in two, three, and four copies, differences in lengths being present within both the meningococcal and gonococcal strains. The C-terminus of this protein, 3' of these tandem repeats, contains the region with homology to bacterial lipoproteins.

TR13 (NMB1468), is also predicted to be a lipoprotein. This CDS contains a 21 bp coding tandem repeat that is present in two, three, four, or seven copies in the strains studied (Table 4). This sequence has no significant homology to other sequences in the public databases. It is noteworthy that a number of the proteins encoded by CDSs containing coding tandem repeats are, or are predicted to be, lipoproteins (TR4, TR11, TR13, TR14, TR15, TR21). In addition to the potential to antigenically vary the protein sequence, and therefore the structure of these surface exposed molecules, the change in number of repeated units may also influence the lipid component of the protein, as has been suggested for Lip [24].

The Lip repeat

Lip (TR15, also known as the H.8 antigen) is largely composed of a repeated 5 amino acid motif, and has been sequenced previously by two groups. The two reported sequences differ in the number of tandemly repeated sequences [16,17]. Although variation of the number of these repeats between strains has not previously been addressed at the DNA-level in the literature, Lip is known to vary in gel mobility suggesting significant inter-strain differences in size [25], and in the form of its lipid component [24]. A virulence-associated lipoprotein, this protein was investigated in the 1980's as a vaccine target due to its antigenicity and capacity to generate an antibody response during disseminated gonococcal infections [26]. Changes in the M_r of the protein correlate with serum-resistance and neutrophil enzyme-resistance [25], although these changes were also demonstrated to effect the immunogenicity and/or antigenicity of gonococcal P.1 [27]. Lip can be present as a multimer, but this too is dependent on the M_r of the monomer [25]. Here we demonstrate 7 different length variations in the tandem repeat

Table 2: Primer pairs used in this study.

Tandem repeat	Forward primer (5'-3')	Reverse primer (5'-3')
TR1	GGTCGCTGGATACGCTGC	CGGTAGCCCAAGCCTGCG
TR2	TGCCGGCAGCAAAGTGTCC	GCCCGTTCCAAACGACCG
TR3	AGCGGCAGCGGACTGCC	GTGTGCCTGCCGTGCCG
TR4	CAAACCGGCAGTTTGGGCG	ATGGGAAATGCGGCTGCCG
TR4v2	TGCTCAGCAGCCGCGAGC	GGATGAAGCGGTTGTTGCCG
TR5	CAGCCGTGCGCGTCTGG	TTGGCTGATGTCGGGCTGC
TR5v2		CGGGGCGCGATGTTGACG
TR6	AACGCCATGCCGTCGAGC	GCCCGCGATGATGTGCCG
TR7	CGAAGCGACCAAAGGCATCC	GGGTCATATTCGCCGTGGTC
TR8	CGGTGAAACCGTTGTTGCCG	TCACGGCCGGAACCTTGC
TR9	CCGCCTACATCCTGTTCGG	AAAGCCGGCGATGATGCGG
TR10	GCGCCGAAAGTTTGGGACG	GCTTCGCCTTGTTCCACGC
TR11	CGCTGCCACCAATGATGTCG	TCGTAGCGGCGACTTCCG
TR12	GTAGAAGAAGTCGGCGAGGC	ATCGCCGTATTGCACGCCG
TR13	CAGCCGATTGATGGAACACG	GCCTGAAAATCTTCAGGCGG
TR14	TTGCCGCTCTGTTGGGCG	TGCCCGTATGCGGTCAGG
TR14v2	CTGTTGGGCGGTTGCGCC	CCGCGTTTGACCCTGCTG
TR15	GCGCATTCTAACACAACCGC	AGGAAGGGAATCTGATGCCG
TR16	GCATACCGCCAGAACGGC	ATACGGCCCGGAACGTCG
TR17	AACACGCAAACGCGGACGC	TGCGGCACGGCAGGTGC
TR18	CAGCAGCCAAATGCCCGC	TGCTGCGGCAGGTTTGGC
TR19a	GCGCCCGAACCGCAACC	GTTTTTCCGCCGGTTTCGGG
TR19b	AACGGGGCGCGGAGAAGG	GCCCGGAGAAACCAAAACGC
TR20	Investigated previously [15]	
TR21	GTCGCGTAGAATGCGGCG	GGAGCCTGCTCCACAACG
TR22	GCCGCCGCATTGCTGGC	GCCCTGCTGTTGTGCGGG
TR23	ATCCTGCCGCCGCCTGC	GATGACCGCGGCATCAGC
TR24	GTGCTTTTCGGGCAAGTGCC	CACCAATCCTACACCGTTCCC
TR25	CGCCCGAAGGGTTTACCG	GACACGCCGTCAATGACGC
TR26	GTTTCAGGGCGAGTTTGCCG	CTCGCTGTGCAGCTGCGC
TR27	GGGTGAAGAACGCAAAGCCC	GAGTTCAGTGCTTCGCGGC
TR28	GATTGGATACGCGGCAACCG	ATACGGCGGCAAGCTCCG

that comprises most of the gene, in which only 69 bp (23 amino acids) coding for the gene are outside the tandem repeat. This is the first report in which the DNA repeat from the Lip encoding gene has been sequenced from different strains, demonstrating a high degree of diversity with copy numbers ranging from 10 to 18 copies. No PCR products were generated from the commensal *N. lactamica* strains, which is consistent with restriction of this gene to the pathogenic species [28]. This protein has not been pursued recenty as a vaccine candidate, probably because antibodies directed against it were poorly bactericidal [29,30]. A second gene within the genomes (NMB1533/NMA1733) contains repeated copies of the AAEAP Lip consensus sequence [31]. The seven repeat copies in this 'azurin-like protein' do not vary between the sequenced strains and therefore it is not included under the criteria of this study. It should be noted that this second CDS has been mis-annotated in both published genomes as H.8, while the real Lip/H.8 antigen CDS (NMB1523/NMA1723) is annotated as a hypothetical protein and putative proline-rich repeat protein, respectively [19,20].

The major anaerobically induced outer membrane lipoprotein, AniA [32] (TR16, NMB1623), also contains tandem repeats with homology to the AAEAP conserved repeat in Lip and the 'azurin-like protein'. In this case there are two sets of tandem repeats, each at the extreme N- and C-terminus [33]. AniA appears to be involved in serum resistance, although it is not expressed under aerobic conditions [34]. The crystal structure of AniA would put these tandem repeats on exposed surfaces of the protein [35], although how they are orientated relative to the membrane is not known. Degeneration in the sequence 3' of the gene meant that conserved primers could not be designed flanking the second repeat, therefore it could not be evaluated by PCR in this study. There is some variability in the N-terminal repeat, with some *N. meningitidis* strains having 3 copies and others 2, but these are 12 bp repeats and those identified by genome compar-

ison at the C-terminus are 6 bp repeats, therefore the differences are not in the complete 15 bp tandem repeat, as in Lip, but rather in the smaller subunits which make up the AAEAP repeat.

One of the two genes identified as present only in *N. gonorrhoeae* strain FA1090 (TR25, XNG0481) also has a tandem repeat which is similar to the AAEAP repeat of Lip, the 'azurin-like protein', and AniA. In this case the repeat was identified by ETANDEM as being 30 bp. It is present in 3 copies, or 6 copies of the 15 bp repeat, in all of the gonococcal strains evaluated. This CDS is present in a gonococcal specific island composed of 58 genes. At one end are genes whose homology indicates a prophage, including a putative phage integrase, transcriptional regulator, phage repressor, and DNA helicase. At the other end of this region are *pemK* and *pemI*, which were identified on plasmid R100 and are involved in its maintainance [36]. Therefore this region has features of both an integrated bacteriophage and an integrated plasmid, with a CDS containing a tandem repeat similar to that seen in other neisserial genes in the middle.

Of the genes evaluated, TR24 (NMB1848), was included in the study due to the high identity (97%) between repeat copies, although the gene itself was only found in *N. meningitidis* strain MC58. TR24 has more repeat copies than the two gonococcal genes also added to the study for this reason, the meningococcal gene having 15 copies of an 18 bp tandem repeat, rather than 3 copies in the two gonococcal genes (an 18 bp element in TR23, and a 30 bp element in TR25). Four length variants of this gene were identified, including differences between the closely related *N. meningitidis* strains MC58 and 44/76 (Table 4). This tandem repeated unit makes up most of the coding sequence of the gene, there being only 72 bp (24 amino acids) of coding sequence that is not within the tandem repeat. The composition of the majority of this gene by a varying number of coding tandem repeats is reminiscent of Lip, but the location and function of this gene product is not known.

The conserved dcaC *repeat*

DcaC (TR20) is predicted to be an outer membrane protein of unknown function containing a 36 amino acid variable copy number tandem repeat, and has been described previously [15]. Although the gene as a whole has no homology to others in the public databases, homologues of the *dcaC* repeat are present in several hypothetical proteins. In *Magnetococcus* sp. MC-1 CDS Mmc10969 there are 14 copies of the repeat (NZ AAAN01000134); nine and ten in *E. coli* strain CFT073 CDSs c1269 and c5321, respectively [37]; four in *Chlorobium tepidum* strain TLS CDS CT0958 [38]; six in *Vogesella indigofera* strain ATCC19706 ORF1 (AF088857.1); five and three in *Haemophilus somnus* strain 129PT CDSs Hsom0164 (NZ AABG01000001) and Hsom1526 (NZ AABG01000013), respectively; and three in *Pasteurella multocida* strain PM70 CDS PM1611 [39]. Such conservation of repeat homology without overall homology of the proteins has not previously been reported. It appears, therefore, that the presence of a protein containing this repeat, and the variability in the copy number of this repeat, is conserved. In the *Neisseria* spp., the number of tandem repeats within the gene clearly increases the number of distinct hydrophobic regions within the protein (Figure 2).

The species-specific rmpM *repeat*

The shortest tandem repeat included in this study is 2 amino acids (6 bp) and is contained within RmpM (NMB0382, TR21). In this case, the presence of 6 copies of the repeat or 2 copies of the repeat appears to be linked with species, meningococci having the former and gonococci the latter. The presence of no PCR products in the *N. lactamica* strains is consistent with other work on this gene in the commensal *Neisseria* spp. [40], which suggests that a homologue is only present in some strains.

A potential adhesin

NMB0586 (TR7) is a putative adhesin that contains a 12 bp tandem repeat, which is actually a 2 amino acid repeat, the translation of the 12 bp repeat being HDHD. Although only 2 to 3 copies of this repeat were identified in this study, the product of this gene is predicted to be expressed on the outer membrane or periplasmic space. Most of the length of the predicted protein sequence shares homology to ABC transport periplasmic components/surface adhesins. The crystal structure for TroA, a periplasmic zinc-binding protein from *Treponema pallidum*, has been solved [41,42] and the placement of the TR7 tandem repeat in the structure suggests that it may alter any substrate binding capacity this neisserial protein may have.

Predicted and known cytoplasmic proteins in which altered coding tandem repeat copy number may alter function

The only one of the 28 genes found to contain no copies of the repeat in one of the strains was TR10, *mfd* (NMB1281), encoding transcription-repair coupling factor. While there are one to three copies of the 207 bp (69 amino acid) repeat in the pathogenic *Neisseria* spp., *N. lactamica* strain L12 has none of the 207 bp that make up the repeat, generating a far shorter protein. Unlike *pilQ* in *N. meningitidis* strain FAM18, only the repeats are absent in the *N. lactamica* strain L12 *mfd* gene, the remainder of the gene remaining intact. Helicase domains, which would be expected in this type of protein, can be found 3' of the repeated structure, and are unaffected by the number of tandem repeats present in the gene. Phase var-

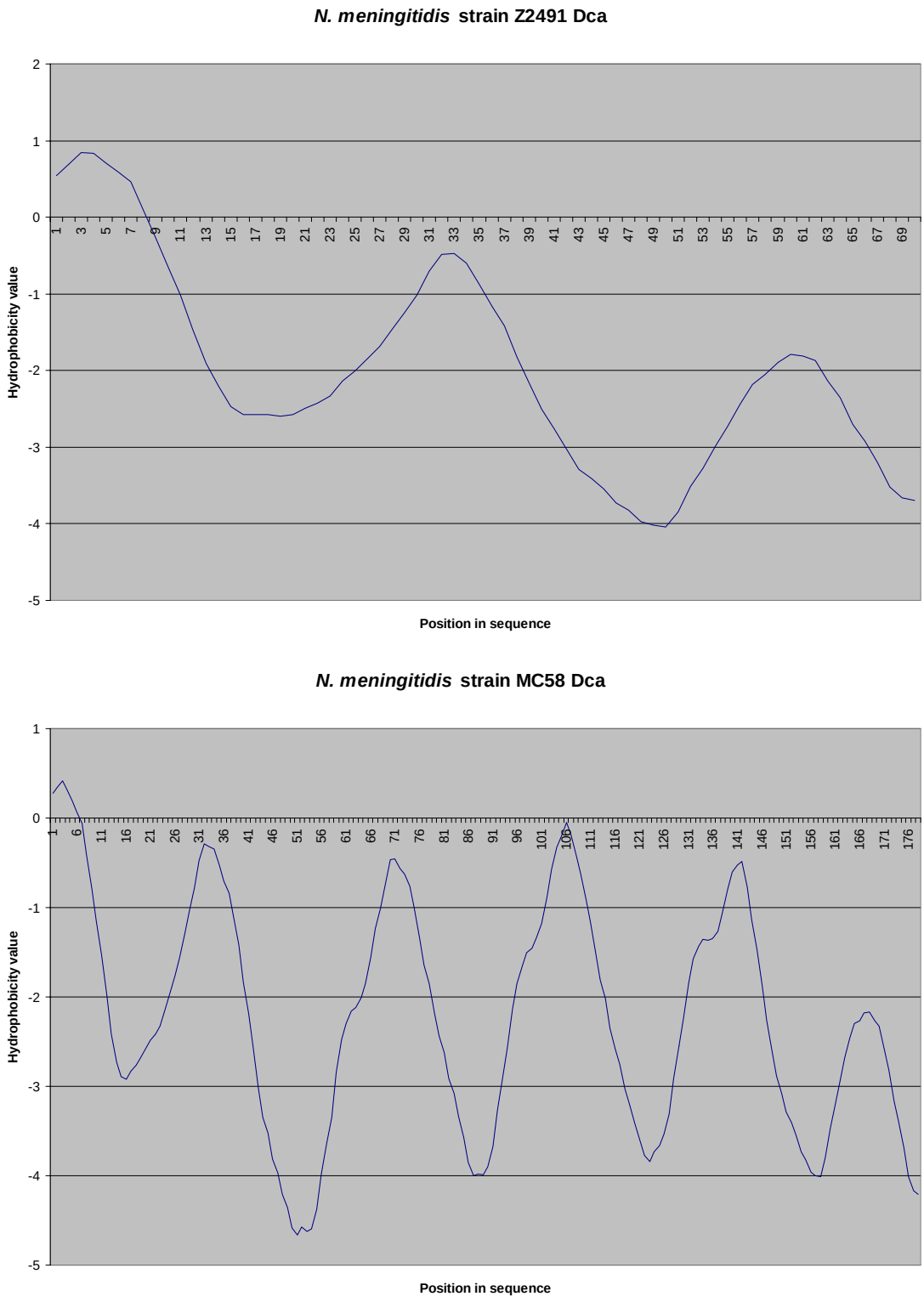

Figure 2
Hydrophobicity profiles of DcaC. The number of coding tandem repeats present in DcaC influences the hydrophobicity profile. In *N. meningitidis* strain Z2491 there is one copy of the 36 amino acid repeat, while in *N. meningitidis* strain MC58 there are four copies. Generated using TopPredII, where the cutoff for certain transmembrane segments is 1, therefore no transmembrane domains are predicted.

Table 3: Genes containing coding tandem repeats with differing copy numbers.

Tandem repeat	NMA#*	NMB# †	XNG# ‡	Gene annotation §	Repeats noted previously?
TR1	NMA0227	NMB2141	XNG1829	hypothetical protein	no
TR2	NMA0257	NMB0010	XNG1803	phosphoglycerate kinase (*pgk*)	no
TR3	NMA0338	NMB2092	XNG1869	hypothetical protein	no
TR4	NMA0440	NMB2001	XNG1095	conserved hypothetical protein	no
TR5	NMA0650	NMB1812	XNG0088	PilQ protein (*pilQ*) ||	yes
TR6	NMA0702	NMB0525	XNG0120	aluminium resistance protein, putative	no
TR7	NMA0789	NMB0586	XNG0151	adhesin, putative	no
TR8	NMA1150	NMB0956	XNG0849	SucB protein (*sucB*)	no
TR9	NMA1461	NMB1027	XNG0526	DnaJ protein, truncation (*dnaJ*)	no
TR10	NMA1491	NMB1281	XNG0596	transcription-repair coupling factor (*mfd*)	no
TR11	NMA1547	NMB1333	XNG0546	conserved hypothetical protein	no
TR12	NMA1612	NMB1395	XNG0677	alcohol dehydrogenase, zinc-containing	no
TR13	NMA1680 ¶	NMB1468	XNG0955	hypothetical protein	no
TR14	NMA1692	NMB1483	XNG0967	lipoprotein NlpD, putative	no
TR15	NMA1723	NMB1523	XNG0903	Lip (H.8 antigen) protein **	yes
TR16	NMA1887	NMB1623	XNG1179	major anaerobically induced OMP (*aniA*) ††	yes
TR17	NMA1897	NMB1643	XNG1189	translation initiation factor IF-2 (*infB*)	no
TR18	NMA1977	NMB1723	XNG1268	cytochrome c oxidase, subunit III (*fixP*)	no
TR19	NMA1985	NMB1730	XNG1276	TonB protein (*tonB*) ‡‡	no
TR20	NMA2065	NMB0419	XNG1419	conserved hypothetical protein (*dcaC*) §§	yes
TR21	NMA2105	NMB0382	XNG1455	outer membrane protein class 4 (*rmpM*) || ||	no
TR22	NMA2206	NMB0281	XNG1595	peptidyl-prolyl cis-trans isomerase	no
TR23	¶¶	¶¶	XNG0938	hypothetical protein	no
TR24	¶¶	NMB1848	¶¶	hypothetical protein	no
TR25	¶¶	¶¶	XNG0481	hypothetical protein	no
TR26	NMA0386	NMB2050	XNG1916	conserved hypothetical protein	no
TR27	NMA1515	NMB1301	XNG0578	30S ribosomal protein S1 (*rpsA*)	no
TR28	NMA2213	NMB0274	XNG1602	DNA helicase RecQ (*recQ*) ***	no

* [19] † [20] ‡ Locus numbers from our own annotation of *N. gonorrhoeae* strain FA1090 as used in [43]. § From [20] unless otherwise noted. || Gene annotation from [74] ¶ NMA1680 is annotated on the reverse complement strand compared to NMB1468 and XNG0955. ** Gene annotations from [75] and [30] †† OMP: outer membrane protein. Gene annotation from [33] ‡‡ Gene annotation from [76] §§ Gene annotation from [15] || || Gene annotation from [77] ¶¶ Corresponding gene is not present in this strain. *** Gene annotation from [78]

iation [43] and recombination [44,45] are two characteristic features of the pathogenic *Neisseria* species. Mfd in other species has been linked with both DNA repair and recombination [46]. A knock-out mutant has been investigated to determine whether it influences phase variation rates in *Haemophilus influenzae*, which found no difference between wild-type and mutant [47]. However, the presence of diversity between neisserial species and strains in the length of a relatively long (69 amino acid) repeat within this protein may significantly affect its activity or interactions.

The greatest variation in copy number is seen within the gene with one of the shortest tandem repeated units. TR22 (NMB0281) contains a 9 bp coding tandem repeat at the 5' end, present in 2, 5, 6, 7, 9, 11, 16, 19, and 26 copies (Table 4). The C-terminus of the protein has homology to a rotamase domain. These enzymes increase the rate of protein folding by catalyzing the interconversion of *cis*-proline and *trans*-proline. It is possible that the copy number of the coding tandem repeat influences the rate or substrate preference of this enzymatic reaction. Tandem repeats in glucansucrases have been previously identified near the active site of these enzymes in *Leuconostoc* and *Streptococcus* species where they may contribute to their function through substrate binding [48].

Within a gonococcus-specific region is a CDS (XNG0938; TR23) that has variable numbers of a 18 bp (6 amino acid) coding tandem repeat. The region that contains TR23 also contains 18 other genes not present in the meningococcal genomes including a divergently transcribed CDS with homology to a phage repressor protein. TR23 itself contains a region that is similar to the integrase core domain found in viral integrases and PSORT predicts this to be a cytoplasmic protein. The features of the genes in this region therefore suggest that this region is derived from a prophage.

Cytoplasmic proteins with variable numbers of coding tandem repeats, which may also be antigenic surface proteins

Two, three, and four copies of a 33 bp tandem repeat are found in pgk (TR2, NMB0010). This gene encodes phosphoglycerate kinase (EC 2.7.2.3), a cytoplasmic enzyme involved in the pathway converting glucose to pyruvate [49]. This protein is conserved between prokaryotes and eukaryotes, and the crystal structures of both pig muscle [50] and Thermotoga maritima phosphoglycerate kinase have been determined [51]. The repeated region in the Neisseria spp. phosphoglycerate kinase maps onto an exposed surface portion of the protein. It was recently reported that in group B streptococcus phosphoglycerate kinase is a surface protein and antibodies directed against it provide protection against infection [52]. It is unclear at this time whether the neisserial protein is cytoplasmic, surface associated, or both, although it should be noted that group B streptococci and serogroup B N. meningitidis share capsule characteristics, and that the sugars for these may be a substrate for surface exposed phosphoglycerate kinase, in addition to its cytoplasmic role. Strains which varied in the copy number of tandem repeats were serogroup B N. meningitidis strains NGE30 (2 copies), BZ133 (2 copies), MC58 (3 copies) and 44/76 (4 copies). In contrast, neither the other serogroups of N. meningitidis, nor the N. gonorrhoeae strains, displayed any variability in tandem repeat copy number (Table 4).

A second protein that functions in the cytoplasm has been identified in this study, SucB (TR8, NMB0956). This is the dihydrolipoamide succinyltransferase (E2o) of the 2-oxogluterate dehydrogenase complex, a component of the TCA cycle. Although the sequence of this gene in *E. coli* contains no repeats, the corresponding acetyltransferase component of the pyruvate dehydrogenase complex (E2p) does [53]. In *Brucella melitensis* and *Coxiella burnetii*, the product of *sucB* is immunogenic; antibodies to SucB being present in the serum of infected sheep and Q fever patients, respectively [54,55]. While there is no evidence that SucB is a surface exposed protein in these species, is does raise the possibility that variation of the 30 bp repeat in the neisserial gene may alter the antigenicity of this protein. Alternatively, the changes in the protein due to the differing tandem repeat copy numbers may offer certain neisserial strains adaptive advantages through altered enzymic activity.

The range of mechanisms generating diversity in Neisseria

Neisserial species have a number of different mechanisms by which they generate diversity. At the level of genic composition they are naturally transformable using a species-specific uptake sequence [56], have the capacity to generate mosaic genes [57,58], have a relatively highly panmictic population structure [59,60], and have genetic loci preferentially associated with strain-divergent genes within Minimal Mobile Elements [61]. At the level of phase variation they have many known and candidate switching genes [43,62], and also have systems utilizing recombination to diversify specific genes such as pilE [63]. Each of these influences the dynamic way in which different strains interact with their hosts, and the flexibility with which a colonizing population can diversify and adapt to the differing and changing environments within a single host. Flexibility due to variation in the number of the coding tandem repeats in the genes highlighted in this paper, as reflected by differences in repeat copy number between strains and species, probably represent an additional mechanism by which these host-restricted pathogens can optimize their niche adaptation to their human hosts. Coding tandem repeats within the Neisseria spp. are likely to add an additional level of diversity generation within the already highly adaptable, dynamic, and variable neisserial population.

Conclusions

Through alteration of the copy number of coding tandem repeats, the *Neisseria* spp. may have an additional mechanism of generation of diversity that has not previously been explored in detail. While the alteration of the copy number of coding tandem repeats has been recognized previously in three genes (*pilQ*, *dcaC*, *lip*), the functional consequence of these changes has not been addressed. This is the first report to identify all the sequenced neisserial genes that have coding tandem repeats and determine if these are present in variable copy number. From this assessment, it becomes apparent that this is potentially a mechanism for antigenic variation of surface proteins and / or for functional variation of cytoplasmic proteins.

Methods

Whole genome analysis to identify coding tandem repeats

The previously described whole genome analysis methodology [62,64] was applied, using the ACEDB graphical interface [65]. The complete genome sequences of N. meningitidis serogroup B strain MC58 [20], N. meningitidis serogroup A strain Z2491 [19], and N. gonorrhoeae strain FA1090 [18](publicly available from 1997, downloaded November 2000 from ftp://ftp.genome.ou.edu/pub/gono/gono-2k.fa), were assessed. Direct tandem repeats were identified using ETANDEM for repeat components of up to 100 bp, due to the fact that it consumes computational cycles in a logarithmically expanding fashion with sequence length. EQUICKTANDEM does not have such heavy computational demands, and is used only for the identification of repeats between 100 and 1000 bp. Both programs are from the EMBOSS package [66], and were used with standard parameter settings. Near the completion of this project the complete sequence of N. meningitidis serogroup C strain FAM18

Table 4: Copy number of coding tandem repeats.

Repeats	Strains	MC58	44/76	NGE30	BZ133	92001	94/155	A22	Z2491	FAM18	L18	L12	L22	FA1090	FA19	26034	Repeat length	Nature *
TR1		2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	30 bp / 10 aa	no new
TR2		3	4	2	2	2	2	2	2	2	2	2	np	2	2	2	33 bp / 11 aa	more
TR3		1	2	2	2	2	2	2	2	2	2	np	np	2	2	2	18 bp / 6 aa	no new
TR4		3	2	3	3	3	3	3	4	3	np	1	1	2	2	1	87 bp / 29 aa	more
TR5 †		2 / 4	2 / 5	2 / 3	2 / 4	2 / 1	2 / 3	2 / 5	2 / 3	0 / 0‡	np	np	np	1 / 1	1 / 2	1 / 2	†	more
TR6		2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	15 bp / 5 aa	no new
TR7		2	2	2	2	3	2	2	3	2	2	3	2	3	3	2	12 bp / 4 aa	no new
TR8		2	2	2	2	2	3	2	3	4	2	2	2	2	2	2	30 bp / 10 aa	more
TR9		2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	30 bp / 10 aa	species
TR10		3	3	2	3	3	3	3	2	3	2	0	2	1	1	1	207 bp / 69 aa	more
TR11		2	2	3	3	4	3	3	3	3	3	3	3	3	4	4	15 bp / 5 aa	more
TR12		1	1	1	3	3	1	1	3	1	1	1	1	1	1	1	21 bp / 7 aa	no new
TR13		2	2	7	2	2	3	2	2	4	np	np	np	3	3	3	21 bp / 7 aa	more
TR14		3	3	3	3	3	3	3	3	3	np	np	np	2	2	2	240 bp / 80 aa	species
TR15		14	14	15	15	10	14	15	14	13	np	np	np	12	18	16	15 bp / 5 aa	more
TR16		3	3	3	2	3	2	np	2	2	3	3	3	3	3	3	12 bp / 4 aa	no new
TR17		2	np	2	2	2	np	2	2	2	np	np	np	1	np	1	57 bp / 19 aa	species
TR18		3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	12 bp / 4 aa	species§
TR19a		3	3	3	3	3	3	3	3	3	np	np	np	2	2	2	18 bp / 6 aa	species
TR19b		1	1	1	1	1	1	1	1	1	4	4	4	4	4	4	9 bp / 3 aa	species
TR20 ||		4	4	8	1	2	2	1	1	1	5	8	11	2	2	nd	108 bp / 36 aa	more
TR21		6	6	6	6	6	6	6	6	6	np	np	np	2	2	2	6 bp / 2 aa	species
TR22		7	11	19	5	6	9	16	9	26	np	np	np	2	2	2	9 bp / 3 aa	more
TR24		15	14	3	np	np	5	14	---	---	np	np	np	---	np	np	18 bp / 6 aa	more
TR26		2	2	2	2	2	1¶	2	2	2	np	np	np	2	2	2	273 bp / 91 aa	none
TR27		2	2	2	2	2	2	2	2	2	np	np	np	2	2	2	261 bp / 87 aa	none
TR28		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	258 bp / 86 aa	none

Repeats	Strains	FA1090	MS11	FA19	F62	25534	28539	26034	26241	27921	29528	28516	27806					
TR23		3	3	3	3	5	3	3	3	3	3	3	3				18 bp / 6 aa	more
TR25		3	3	3	3	3	3	3	3	3	3	3	3				30 bp / 10 aa	none

np – no product nd – not done * Nature of the length variations of the coding tandem repeat: species – observed lengths are associated with species; more – lengths in addition to what is seen in the three sequence strains used in genome comparison analysis; no new – no new lengths observed in addition to those seen in the sequenced strains used in genome comparison analysis; none – no length differences observed. † Compound tandem repeat, the first is 66 bp / 11 aa and the second is 24 bp / 8 aa ‡ PilQ is frame-shifted in the genome sequence due to a large deletion starting 50 bp into the 66 bp component of the compound tandem repeat element and ending 361 bp after the region that contains the repeats in other strains. § Length differences were only seen in the non-pathogen, *N. lactamica*. || From [15]. ¶ The CDS in this strain contains a 303 bp deletion, which begins 63 bp before the first copy of the tandem repeat in the other strains and ends 32 bp before the second copy of the tandem repeat in the other strains. Thus this strain has one full copy of the tandem repeat and 32 bp remaining of a second copy.

became available from The Wellcome Trust Sanger Institute ftp://ftp.sanger.ac.uk/pub/pathogens/nm/, and the coding tandem repeat copy numbers of the 28 identified genes identified from the initial 3-way genome sequence comparison were similarly assessed.

Bacterial strains and growth conditions

The neisserial strains used are shown in Table 1. These strains were chosen based on the results obtained previously concerning copy number differences in the coding tandem repeat in dcaC [15]. In addition to the information presented in that publication, further information on most of these strains can be obtained from the Neisseria Multi Locus Sequence Typing website http://neisseria.mlst.net developed by Dr Man-Suen Chan and sited at the University of Oxford. Strains were propagated on GC agar (Difco Laboratories) containing the Kellogg supplement and ferric nitrate [67] at 37°C under 5% (v/v) CO_2.

PCR amplification and sequencing

Chromosomal DNA extractions were performed using the method of McAllister and Stephens [68] or Ausubel et al. [69]. PCR from chromosomal DNA was performed using Invitrogen Taq DNA Polymerase or Bioline Bio-X-Act polymerase according to the manufacturers' instructions using the primer pairs shown in Table 2. When necessary, secondary primers were designed to obtain PCR products and sequences, denoted v2 on Table 2. PCR products were resolved on the appropriate concentration of either SeaKem® LE agarose gels (Flowgen) or MetaPhor® agarose gels (Flowgen) containing 0.5 μg/ml Ethidium Bromide (Sigma). PCR product size was determined using Quantity One® Quantitation Software (BIORAD). Automated sequencing used ABI Prism® BigDye™ Terminator Cycle Sequencing version 2.0 or version 3.0 (Applied Biosystems), and was resolved on an ABI Prism® 3100 DNA Sequencer (Applied Biosystems).

Nucleotide sequence analysis

The Basic Local Alignment Search Tool (BLAST) [70] was used to search publicly available microbial genome sequences, GenBank, or EMBL. The complete genome sequence of *N. gonorrhoeae* strain FA1090 was obtained from the *N. gonorrhoeae* Genome Sequencing Project at the University of Oklahoma http://www.genome.ou.edu/gono.html, which was independently annotated as described previously [43]. XNG numbers refer to this annotation, and where no *N. meningitidis* homologue is present to identify these CDSs (TR23 & TR25), the sequences of the *N. gonorrhoeae* CDSs referred to are provided as additional material [see 1]. The *N. meningitidis* serogroup C strain FAM18 genome sequence was obtained from the Wellcome Trust Sanger Institute ftp://ftp.sanger.ac.uk/pub/pathogens/nm/. GenBank and EMBL were accessed through the National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/ and the Oxford University Bioinformatics Centre, respectively. Protein domain determinations were addressed through the NCBI Conserved Domain Search, and crystal structures, where available, were visualized using Cn3D v4.0 (NCBI). The Wisconsin Package from GCG (Accelrys) was used for nucleotide and amino acid sequence analysis and alignments. Staden was used for ABI sequence trace assembly and analysis. Predictions of signal sequences and protein localization were generated using PSORT, which currently claims 83% prediction accuracy [71]. Transmembrane domains and hydrophobicity profiles were predicted using TopPredII [72,73] as implemented by Deveaud and Schuerer (Pasteur Institute; http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html).

Authors' contributions

PJ carried out the majority of the PCR and sequencing. LS conducted the whole-genome analysis, designed the primers, analyzed and aligned the sequences, did some of the PCR and sequencing, and drafted the manuscript. NS was the supervisor of the work, participating in its design, coordination, and data interpretation, as well as manuscript editing. All authors read and approved the final manuscript.

Additional material

Additional File 1

Sequences from the complete genome of N. gonorrhoeae *strain FA1090 that are referred to in the manuscript, but for which there are no annotated meningococcal homologues to which readers can refer. It is necessary to identify of these putative coding sequences in this way due to the lack of publication and public annotation of the* N. gonorrhoeae *strain FA1090 genome sequence.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2180-3-23-S1.doc]

Acknowledgements

The authors would like to thank Dr. Sarah Butcher for continued bioinformatics support and Julian Robinson from the Sir William Dunn School of Pathology sequencing service. LASS was supported by the E. P. Abraham Trust. NJS was supported by a Wellcome Trust Advanced Research Fellowship. This publication made use of the Neisseria Multi Locus Sequence Typing website http://neisseria.mlst.net developed by Dr Man-Suen Chan and sited at the University of Oxford. The development of this site is funded by the Wellcome Trust. The meningococcal genome sequence data were produced by the *N. meningitidis* serogroup C strain FAM18 Sequencing Group at the Wellcome Trust Sanger Institute, and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/nm/. The *N. gonorrhoeae* genome sequence was obtained from the University of Oklahoma, the Gonococcal Genome Sequencing Project which was supported by USPHS/NIH grant #AI-38399, GenBank accession number AE004969 [18].

References

1. Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, Ramisse V, Sylvestre P, Benson G, Ramisse F, Vergnaud G: **A tandem repeats database for bacterial genomes: application to the genotyping of Yersinia pestis and Bacillus anthracis.** *BMC Microbiol* 2001, **1:**2.
2. Achaz G, Rocha EP, Netter P, Coissac E: **Origin and fate of repeats in bacteria.** *Nucleic Acids Res* 2002, **30:**2987-2994.
3. Madoff LC, Michel JL, Gong EW, Kling DE, Kasper DL: **Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein.** *Proc Natl Acad Sci U S A* 1996, **93:**4131-4136.
4. Gravekamp C, Rosner B, Madoff LC: **Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice.** *Infect Immun* 1998, **66:**4347-4354.
5. Michel JL, Madoff LC, Olson K, Kling DE, Kasper DL, Ausubel FM: **Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci.** *Proc Natl Acad Sci U S A* 1992, **89:**10060-10064.
6. Wastfelt M, Stalhammar_Carlemalm M, Delisse AM, Cabezon T, Lindahl G: **Identification of a family of streptococcal surface proteins with extremely repetitive structure.** *J Biol Chem* 1996, **271:**18892-18897.
7. Shankar V, Baghdayan AS, Huycke MM, Lindahl G, Gilmore MS: **Infection-derived Enterococcus faecalis strains are enriched in esp, a gene encoding a novel surface protein.** *Infect Immun* 1999, **67:**193-200.
8. Hollingshead SK, Fischetti VA, Scott JR: **Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats.** *Mol Gen Genet* 1987, **207:**196-203.
9. Hollingshead SK, Fischetti VA, Scott JR: **Complete nucleotide sequence of type 6 M protein of the group A Streptococcus. Repetitive structure and membrane anchor.** *J Biol Chem* 1986, **261:**1677-1686.
10. Ladefoged SA: **Molecular dissection of Mycoplasma hominis.** *APMIS Suppl* 2000, **97:**1-45.
11. Yogev D, Watson_McKown R, Rosengarten R, Im J, Wise KS: **Increased structural and combinatorial diversity in an extended family of genes encoding Vlp surface proteins of Mycoplasma hyorhinis.** *J Bacteriol* 1995, **177:**5636-5643.
12. Citti C, Kim MF, Wise KS: **Elongated versions of Vlp surface lipoproteins protect Mycoplasma hyorhinis escape variants from growth-inhibiting host antibodies.** *Infect Immun* 1997, **65:**1773-1785.
13. Kato H, Takeuchi H, Oishi Y, Senpuku H, Shimura N, Hanada N, Nisizawa T: **The immunogenicity of various peptide antigens inducing cross-reacting antibodies to a cell surface protein antigen of Streptococcus mutans.** *Oral Microbiol Immunol* 1999, **14:**213-219.
14. Tønjum T, Caugant DA, Dunham SA, Koomey M: **Structure and function of repetitive sequence elements associated with a highly polymorphic domain of the Neisseria meningitidis PilQ protein.** *Mol Microbiol* 1998, **29:**111-124.
15. Snyder LAS, Shafer WM, Saunders NJ: **Divergence and transcriptional analysis of the division and cell wall (dcw) gene cluster in Neisseria spp.** *Molecular Microbiology* 2003, **47:**431-441.
16. Baehr W, Gotschlich EC, Hitchcock PJ: **The virulence-associated gonococcal H.8 gene encodes 14 tandemly repeated pentapeptides.** *Mol Microbiol* 1989, **3:**49-55.
17. Woods JP, Spinola SM, Strobel SM, Cannon JG: **Conserved lipoprotein H.8 of pathogenic Neisseria consists entirely of pentapeptide repeats.** *Mol Microbiol* 1989, **3:**43-48.
18. Lewis LA, Gillaspy AF, McLaughlin RE, Gipson M, Ducey T, Ownbey T, Hartman K, Nydick C, Carson M, Vaughn J, Thomson C, Song L, Lin S, Yuan X, Najar F, Zhan M, Ren Q, Zhu H, Qi S, Kenton SM, Lai H, White JD, Clifton S, Roe BA, Dyer DW: **The Gonococcal genome sequencing project.** *unpublished* .
19. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S, Spratt BG, Barrell BG: **Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.** *Nature* 2000, **404:**502-506.
20. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC: **Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.** *Science* 2000, **287:**1809-1815.
21. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R: **Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing.** *Science* 2000, **287:**1816-1820.
22. Hess J, Dreher A, Gentschev I, Goebel W, Ladel C, Miko D, Kaufmann SH: **Protein p60 participates in intestinal host invasion by Listeria monocytogenes.** *Zentralbl Bakteriol* 1996, **284:**263-272.
23. Collins RF, Davidsen L, Derrick JP, Ford RC, Tonjum T: **Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure.** *J Bacteriol* 2001, **183:**3825-3832.
24. Bhattacharjee AK, Moran EE, Ray JS, Zollinger WD: **Purification and characterization of H.8 antigen from group B Neisseria meningitidis.** *Infect Immun* 1988, **56:**773-778.
25. Hitchcock PJ, Hayes SF, Mayer LW, Shafer WM, Tessier SL: **Analyses of gonococcal H8 antigen. Surface location, inter- and intrastrain electrophoretic heterogeneity, and unusual two-dimensional electrophoretic characteristics.** *J Exp Med* 1985, **162:**2017-2034.
26. Black JR, Black WJ, Cannon JG: **Neisserial antigen H.8 is immunogenic in patients with disseminated gonococcal and meningococcal infections.** *J Infect Dis* 1985, **151:**650-657.
27. Judd RC, Shafer WM: **Topographical alterations in proteins I of Neisseria gonorrhoeae correlated with lipooligosaccharide variation.** *Mol Microbiol* 1989, **3:**637-643.
28. Aho EL, Murphy GL, Cannon JG: **Distribution of specific DNA sequences among pathogenic and commensal Neisseria species.** *Infect Immun* 1987, **55:**1009-1013.
29. Tinsley CR, Virji M, Heckels JE: **Antibodies recognizing a variety of different structural motifs on meningococcal Lip antigen fail to demonstrate bactericidal activity.** *J Gen Microbiol* 1992, **138 (Pt 11):**2321-2328.
30. Bhattacharjee AK, Moran EE, Zollinger WD: **Antibodies to meningococcal H.8 (Lip) antigen fail to show bactericidal activity.** *Can J Microbiol* 1990, **36:**117-122.
31. Kawula TH, Spinola SM, Klapper DG, Cannon JG: **Localization of a conserved epitope and an azurin-like domain in the H.8 protein of pathogenic Neisseria.** *Mol Microbiol* 1987, **1:**179-185.
32. Hoehn GT, Clark VL: **The major anaerobically induced outer membrane protein of Neisseria gonorrhoeae, Pan 1, is a lipoprotein.** *Infect Immun* 1992, **60:**4704-4708.
33. Hoehn GT, Clark VL: **Isolation and nucleotide sequence of the gene (aniA) encoding the major anaerobically induced outer membrane protein of Neisseria gonorrhoeae.** *Infect Immun* 1992, **60:**4695-4703.
34. Cardinale JA, Clark VL: **Expression of AniA, the major anaerobically induced outer membrane protein of Neisseria gonorrhoeae, provides protection against killing by normal human sera.** *Infect Immun* 2000, **68:**4368-4369.
35. Boulanger MJ, Murphy ME: **Crystal structure of the soluble domain of the major anaerobically induced outer membrane protein (AniA) from pathogenic Neisseria: a new class of copper-containing nitrite reductases.** *J Mol Biol* 2002, **315:**1111-1127.
36. Tsuchimoto S, Ohtsubo H, Ohtsubo E: **Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100.** *J Bacteriol* 1988, **170:**1461-1466.
37. Welch RA, Burland V, Plunkett GDIII, Redford P, Roesch P, Rasko DA, Buckles EL, Liou S-R, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HLT, Donnenberg MS, Blattner FR: **Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia**

coli. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99:**17020-17024.
38. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA, Fraser CM: **The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium.** *Proc Natl Acad Sci U S A* 2002, **99:**9509-9514.
39. May BJ, Zhang Q, Li LL, Paustian ML, Whittam TS, Kapur V: **Complete genomic sequence of Pasteurella multocida, Pm70.** *Proc Natl Acad Sci U S A* 2001, **98:**3460-3465.
40. Troncoso G, Sanchez S, Kolberg J, Rosenqvist E, Veiga M, Ferreiros CM, Criado M: **Analysis of the expression of the putatively virulence-associated neisserial protein RmpM (class 4) in commensal Neisseria and Moraxella catarrhalis strains.** *FEMS Microbiol Lett* 2001, **199:**171-176.
41. Lee YH, Deka RK, Norgard MV, Radolf JD, Hasemann CA: **Treponema pallidum TroA is a periplasmic zinc-binding protein with a helical backbone.** *Nat Struct Biol* 1999, **6:**628-633.
42. Lee YH, Dorwart MR, Hazlett KR, Deka RK, Norgard MV, Radolf JD, Hasemann CA: **The crystal structure of Zn(II)-free Treponema pallidum TroA, a periplasmic metal-binding protein, reveals a closed conformation.** *J Bacteriol* 2002, **184:**2300-2304.
43. Snyder LA, Butcher SA, Saunders NJ: **Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic Neisseria spp.** *Microbiology* 2001, **147:**2321-2332.
44. Feil EJ, Spratt BG: **Recombination and the population structures of bacterial pathogens.** *Annu Rev Microbiol* 2001, **55:**561-590.
45. Feil EJ, Maiden MC, Achtman M, Spratt BG: **The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis.** *Mol Biol Evol* 1999, **16:**1496-1502.
46. Ayora S, Rojo F, Ogasawara N, Nakai S, Alonso JC: **The Mfd protein of Bacillus subtilis 168 is involved in both transcription-coupled DNA repair and DNA recombination.** *J Mol Biol* 1996, **256:**301-318.
47. Bayliss CD, van_de_Ven T, Moxon ER: **Mutations in poll but not mutSLH destabilize Haemophilus influenzae tetranucleotide repeats.** *EMBO J* 2002, **21:**1465-1476.
48. Janecek S, Svensson B, Russell RR: **Location of repeat elements in glucansucrases of Leuconostoc and Streptococcus species.** *FEMS Microbiol Lett* 2000, **192:**53-57.
49. Conway T, Ingram LO: **Phosphoglycerate kinase gene from Zymomonas mobilis: cloning, sequencing, and localization within the gap operon.** *J Bacteriol* 1988, **170:**1926-1933.
50. Kovari Z, Flachner B, Naray_Szabo G, Vas M: **Crystallographic and thiol-reactivity studies on the complex of pig muscle phosphoglycerate kinase with ATP analogues: correlation between nucleotide binding mode and helix flexibility.** *Biochemistry* 2002, **41:**8796-8806.
51. Auerbach G, Huber R, Grattinger M, Zaiss K, Schurig H, Jaenicke R, Jacob U: **Closed structure of phosphoglycerate kinase from Thermotoga maritima reveals the catalytic mechanism and determinants of thermal stability.** *Structure* 1997, **5:**1475-1483.
52. Hughes MJ, Moore JC, Lane JD, Wilson R, Pribul PK, Younes ZN, Dobson RJ, Everest P, Reason AJ, Redfern JM, Greer FM, Paxton T, Panico M, Morris HR, Feldman RG, Santangelo JD: **Identification of major outer surface proteins of Streptococcus agalactiae.** *Infect Immun* 2002, **70:**1254-1259.
53. Spencer ME, Darlison MG, Stephens PE, Duckenfield IK, Guest JR: **Nucleotide sequence of the sucB gene encoding the dihydrolipoamide succinyltransferase of Escherichia coli K12 and homology with the corresponding acetyltransferase.** *Eur J Biochem* 1984, **141:**361-374.
54. Zygmunt MS, Diaz MA, Teixeira_Gomes AP, Cloeckaert A: **Cloning, nucleotide sequence, and expression of the Brucella melitensis sucB gene coding for an immunogenic dihydrolipoamide succinyltransferase homologous protein.** *Infect Immun* 2001, **69:**6537-6540.
55. Nguyen SV, To H, Yamaguchi T, Fukushi H, Hirai K: **Characterization of the Coxiella burnetti sucB gene encoding an immunogenic dihydrolipoamide succinyltransferase.** *Microbiol Immunol* 1999, **43:**743-749.
56. Elkins C, Thomas CE, Seifert HS, Sparling PF: **Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence.** *J Bacteriol* 1991, **173:**3911-3913.
57. Fudyk TC, Maclean IW, Simonsen JN, Njagi EN, Kimani J, Brunham RC, Plummer FA: **Genetic diversity and mosaicism at the por locus of Neisseria gonorrhoeae.** *J Bacteriol* 1999, **181:**5591-5599.
58. Zhou J, Bowler LD, Spratt BG: **Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of Neisseria meningitidis and commensal Neisseria species.** *Mol Microbiol* 1997, **23:**799-812.
59. Maiden MC: **Population genetics of a transformable bacterium: the influence of horizontal genetic exchange on the biology of Neisseria meningitidis.** *FEMS Microbiol Lett* 1993, **112:**243-250.
60. Smith JM, Smith NH, O_Rourke M, Spratt BG: **How clonal are bacteria?** *Proc Natl Acad Sci U S A* 1993, **90:**4384-4388.
61. Saunders NJ, Snyder LAS: **The minimal mobile element.** *Microbiology (Reading, England)* 2002, **148:**3756-3760.
62. Saunders NJ, Jeffries AC, Peden JF, Hood DW, Tettelin H, Rappuoli R, Moxon ER: **Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58.** *Mol Microbiol* 2000, **37:**207-215.
63. Haas R, Veit S, Meyer TF: **Silent pilin genes of Neisseria gonorrhoeae MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates.** *Mol Microbiol* 1992, **6:**197-208.
64. Saunders NJ, Peden JF, Hood DW, Moxon ER: **Simple sequence repeats in the Helicobacter pylori genome.** *Mol Microbiol* 1998, **27:**1091-1098.
65. Durbin R, Thierry-Mieg JT: **A C. elegans DataBase. Documentation, code and data available from http://www.acedb.org.** 1991.
66. Lewis LA, Gillaspy AF, McLaughlin RE, Gipson M, Ducey T, Ownbey T, Hartman K, Nydick C, Carson M, Vaughn J, Thomson C, Song L, Lin S, Yuan X, Najar F, Zhan M, Ren Q, Zhu H, Qi S, Kenton SM, Lai H, White JD, Clifton S, Roe BA, Dyer DW: **The Gonococcal genome sequencing project.** *unpublished* .
67. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16:**276-277.
68. Kellogg D.S., Jr., Peacock W.L., Jr., Deacon WE, Brown L, Pirkle CI: **Neisseria gonorrhoeae. I. Virulence genetically linked to clonal variation.** *Journal of Bacteriology* 1963, **85:**1274-1279.
69. McAllister CF, Stephens DS: **Analysis in Neisseria meningitidis and other Neisseria species of genes homologous to the FKBP immunophilin family.** *Mol Microbiol* 1993, **10:**13-23.
70. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: **Preparation of genomic DNA from bacteria.** *Current Protocols in Molecular Biology* 1990, **1:**2.4.1-2.4.2.
71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
72. Nakai K, Kanehisa M: **Expert system for predicting protein localization sites in Gram-negative bacteria.** *Proteins* 1991, **11:**95-110.
73. von_Heijne G: **Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule.** *J Mol Biol* 1992, **225:**487-494.
74. Claros MG, von Heijne G: **TopPred II: an improved software for membrane protein structure predictions.** *Comput Appl Biosci* 1994, **10:**685-686.
75. Lewis LA, Gillaspy AF, McLaughlin RE, Gipson M, Ducey T, Ownbey T, Hartman K, Nydick C, Carson M, Vaughn J, Thomson C, Song L, Lin S, Yuan X, Najar F, Zhan M, Ren Q, Zhu H, Qi S, Kenton SM, Lai H, White JD, Clifton S, Roe BA, Dyer DW: **The Gonococcal genome sequencing project.** *unpublished* .
76. Drake SL, Koomey M: **The product of the pilQ gene is essential for the biogenesis of type IV pili in Neisseria gonorrhoeae.** *Mol Microbiol* 1995, **18:**975-986.
77. Cannon JG, Black WJ, Nachamkin I, Stewart PW: **Monoclonal antibody that recognizes an outer membrane antigen common to the pathogenic Neisseria species but not to most nonpathogenic Neisseria species.** *Infect Immun* 1984, **43:**994-999.

78. Stojiljkovic I, Srinivasan N: **Neisseria meningitidis tonB, exbB, and exbD genes: Ton-dependent utilization of protein-bound iron in Neisseriae.** *J Bacteriol* 1997, **179:**805-812.
79. Klugman KP, Gotschlich EC, Blake MS: **Sequence of the structural gene (rmpM) for the class 4 outer membrane protein of Neisseria meningitidis, homology of the protein to gonococcal protein III and Escherichia coli OmpA, and construction of meningococcal strains that lack class 4 protein.** *Infect Immun* 1989, **57:**2066-2071.
80. Mehr IJ, Seifert HS: **Differential roles of homologous recombination pathways in Neisseria gonorrhoeae pilin antigenic variation, DNA transformation and DNA repair.** *Mol Microbiol* 1998, **30:**697-710.