Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp.

Lori A. S. Snyder,¹ Sarah A. Butcher² and Nigel J. Saunders¹

Author for correspondence: Nigel J. Saunders. Tel: +44 1865 275521. Fax: +44 1865 275515.
e-mail: saunders@molbiol.ox.ac.uk

¹ The Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

² Oxford University Bioinformatics Centre, The Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Previously, a complete genome analysis of *Neisseria meningitidis* strain MC58 revealed the largest repertoire of putative phase-variable genes described in any species to date. Initial comparisons with two incomplete *Neisseria* spp. genome sequences available at that time revealed differences in the repeats associated with these genes in the form of polymorphisms, the absence of the potentially unstable elements in some alleles, and in the repertoire of the genes that were present. Analyses of the complete genomes of *N. meningitidis* strain Z2491 and *Neisseria gonorrhoeae* strain FA1090 have been performed and are combined with a comprehensive comparative analysis between the three available complete genome sequences. This has increased the sensitivity of these searches and provided additional contextual information that facilitates the interpretation of the functional consequences of repeat instability. This analysis identified: (i) 68 phase-variable gene candidates in *N. meningitidis* strain Z2491, rather than the 27 previously reported; (ii) 83 candidates in *N. gonorrhoeae* strain FA1090; and (iii) 82 candidates in *N. meningitidis* strain MC58, including an additional 19 identified through cross-comparisons with the other two strains. In addition to the 18 members of the *opa* gene family, a repertoire of 119 putative phase-variable genes is described, indicating a huge potential for diversification mediated by this mechanism of gene switching in these species that is central to their interactions with the host and environmental transitions. Eighty-two of these are either known (14) or strong (68) candidates for phase variation, which together with the *opa* genes make a total of 100 identified genes. The repertoires of the genes identified in this analysis diverge from the different species groupings, indicating horizontal exchange that significantly affects the species and strain complements of these genes.

Keywords: phase variation, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, repeat, genome analysis

INTRODUCTION

The pathogenic *Neisseria* spp. *Neisseria meningitidis* and *Neisseria gonorrhoeae* are causative agents of meningitis and septicaemia, and gonorrhoea respectively. These populations are characterized by genetic diversity at several levels. Horizontal exchange of DNA between strains and neisserial species creates a situation in which sequences present within clones are available

over time to the rest of the population (Bowler *et al.*, 1994; Feil *et al.*, 1995; Saunders *et al.*, 1999). There is extensive allelic diversity in some genes, particularly those under antigenic selection pressures (Malorny *et al.*, 1998). There are genes within a strain that undergo recombination between expressed and silent cassettes to generate diversity within clonal populations (Haas & Meyer, 1987). Finally, there are genes that are switched on and off by phase variation that can provide a large repertoire of phenotypes from within a clonal population, which can provide adaptation to changing environmental conditions (Sparling *et al.*, 1986; Stern

.....
Abbreviation: HPT, homopolymeric tract.

et al., 1986; Meyer & van Putten, 1989; Yang & Gotschlich, 1996; Saunders *et al.*, 2000). In *Neisseria* spp., phase variation is consistently associated with reversible changes within simple DNA repeats, composed of repeated sequence motifs of less than 10 nt. These repeats are located either within ORFs or within promoters. Repeats located within ORFs alter the relative translational reading frame of the sequence 5' and 3' of the repeat as the number of repeats alters (Stern *et al.*, 1986; Sparling *et al.*, 1986; Stibitz *et al.*, 1989). Changes in the length of repeats located within promoters alter the relative position of promoter components and thereby influence transcription (Sarkari *et al.*, 1994; van der Ende *et al.*, 1995).

Phase variation is a strategy used by many bacterial species. The advent of complete bacterial genome sequences has facilitated the investigation and identification of the repertoires of genes that are regulated in this way on a whole-system basis. The first analysis of this type was performed in *Haemophilus influenzae* strain Rd in which 12 genes were identified on the basis of their association with tetrameric repeat sequence motifs (Hood *et al.*, 1996). Search methodology was subsequently improved to include all possible repeat motifs and to present each in its sequence context. This was initially used to analyse the genome sequence of *Helicobacter pylori* strain 26695, identifying 26 candidate phase-variable genes (Saunders *et al.*, 1998), several of which have been subsequently confirmed experimentally (Appelmek *et al.*, 1999, 2000; Peck *et al.*, 1999; Wang *et al.*, 1999; Josenhans *et al.*, 2000; Yamaoka *et al.*, 2000). In addition to increased sensitivity, this approach allowed several genes that had previously been suggested to be phase-variable in this strain, on the basis of repeat searching alone (Tomb *et al.*, 1997), to be excluded on the basis of the sequence context of the repeat. Analysis of the *N. meningitidis* strain MC58 genome sequence revealed 65 putative phase-variable genes of which 44 were either previously recognized (13) or were particularly 'strong' candidates (31) (Saunders *et al.*, 2000). Since the capacity of bacterial populations to generate diversity increases exponentially with the number of genes that can be phase-varied, the number of variable genes found in these whole-genome analyses suggest that clones are capable of switching between vast repertoires of phenotypes. The size of this repertoire and the nature of the genes for which a function can be determined indicate that this type of gene switching is a central aspect of the population's response to changing conditions. As discussed previously (Saunders *et al.*, 2000), the *Neisseria* spp. represent a unique context in which to interpret the results of this type of analysis because of the extensive existing body of knowledge on the length, composition and instability of repeats associated with phase variation. These include relatively short homopolymeric tracts (HPTs), e.g. of (C)₇ in a gene for capsule biosynthesis (Hammerschmidt *et al.*, 1996). These observations were consistent with Markov modelling which identified an over-abundance of repeats

above this length in *N. meningitidis* strain MC58 (Saunders *et al.*, 2000), which is also seen in the other two neisserial sequences (data not presented), suggestive of a mutational bias for instability in these sequences. Although there may be circumstances when additional regions of instability exist (van der Ende *et al.*, 2000), when repeats that are consistently associated with gene switching are identifiable in appropriate sequence contexts this has greater predictive power than in other species. This is enhanced by the opportunity to perform comparative analyses of three genomes.

In contrast to the large repertoire of phase-variable candidates described in *N. meningitidis* strain MC58, the report of the complete sequence of *N. meningitidis* strain Z2491 described only 27 potential phase-variable genes (Parkhill *et al.*, 2000). This suggested that a significant difference exists between the adaptability and behaviour of these two strains from different serogroups. Initial comparisons with the MC58 sequence suggested that strain Z2491 possesses a larger repertoire than initially described (Saunders *et al.*, 2000). However, comparison between different sequences of this type can only be made on the basis of similar analyses using the same methodologies.

This paper describes independent analyses of the complete genome sequences of *N. meningitidis* strain Z2491 and *N. gonorrhoeae* strain FA1090 to determine their repertoires of potentially phase-variable genes. The results of these analyses were then combined with the previous analysis of *N. meningitidis* strain MC58 and a comprehensive comparison was made between the three genome sequences of the candidate genes identified in each individual sequence. This revealed many additional genes in all three genomes and clearly demonstrates the value of performing comparative analysis – the information it provides increasing both the sensitivity and specificity of the search results. It also provides a more comprehensive background for understanding the contribution of phase variation to the behaviour of these species.

METHODS

The complete genome sequences of *N. meningitidis* serogroup A strain Z2491 (Parkhill *et al.*, 2000) and *N. gonorrhoeae* strain FA1090 (publicly available from 1997, downloaded November 2000 from <ftp://ftp.genome.ou.edu/pub/gono/gono-2k.fa>; GenBank no. AE004969) were analysed using previously described whole-genome analysis methodology (Saunders *et al.*, 1998, 2000), using an ACEDB graphical interface (Durbin & Thierry-Mieg, 1991; available from www.sanger.ac.uk). Briefly, whole-genome sequences were analysed using BLASTN and BLASTX (Altschul *et al.*, 1997; available from <http://www.ncbi.nlm.nih.gov>) searches against all sequences in the EMBL, TREMBL and SWISS-PROT databases classified below 'metazoan' and excluding viral sequences. Repeats composed of perfect repeats with motifs of 1–10 bases were identified using ARRAYFINDER (Hancock *et al.*, 1999; software available from J. Hancock on request). As a complementary analysis, tandem repeats composed of repeated motifs of up to 1000 bp were identified

using ETANDEM from EMBOSS (European Molecular Biology Open Software Suite, version 1.9.0). All repeats, were displayed in their sequence contexts with respect to ORFs and termination codons using the tools within ACEDB, and their protein and nucleotide homologies. These complete genome sequence databases were then analysed for simple DNA repeats within their sequence contexts to determine the repertoire of putative phase-variable genes. HPTs of greater than 6 Gs or Cs, and greater than 8 As or Ts, were each investigated and repeats below these thresholds when associated with a frameshift. Other repeats composed of ≥ 4 copies of dinucleotides and ≥ 3 copies of tetramer and longer motifs were also investigated. All repeats were analysed to interpret the significance of the repeat on the basis of sequence context and the potential effect of length variation on the expression of the associated reading frames. To facilitate this analysis the complete genome of *N. gonorrhoeae* strain FA1090 was annotated and assigned sequence feature numbers (XNG#) using ACEDB as described previously (Tettelin *et al.*, 2000) with the addition of BLASTN and BLASTX searches using the two meningococcal annotations (Parkhill *et al.*, 2000; Tettelin *et al.*, 2000). The annotated gonococcal ORFs were extracted from the genome sequence in FASTA format and translated using the TRANSEQ program from EMBOSS. Three new ACEDBs were then generated in which each of the three genomes was compared to the annotations for the other two by BLASTN and BLASTX searches. These were used to investigate the analogous sequence regions containing repeats potentially associated with phase variation in the three genomes. Further analysis of regions was performed using the tools within GCG10 (Genetics Computer Group, Madison, WI, USA) through the Oxford University Bioinformatics Centre and/or Jellyfish (freeware available from <http://www.biowire.com>). Gonococcal reading frames are referred to on the basis of their similarity to meningococcal sequences where possible. The sequences of the remaining in-house-annotated genes and NMA1707a, an ORF not annotated by Parkhill *et al.* (2000), are available from the Saunders group web page (via <http://www.path.ox.ac.uk>). Some of these data are also available as supplementary data on the journal web site (<http://www.mic.sgmjournals.org>).

RESULTS AND DISCUSSION

Whole-genome analysis of *N. meningitidis* strain Z2491 for potential phase-variable genes

With their description of the complete sequence, Parkhill *et al.* (2000; supplementary information, Table 3) published a table of 27 tandem repeat sequences indicative of phase variation in the serogroup A *N. meningitidis* strain Z2491. The current analysis identified 54 candidate genes (Table 1 & Table 4), including some but not all of those proposed by Parkhill *et al.* (2000) (note that the repeat described in the original sequence paper table NMA0406 is in the gene annotated as NMA0407). Twenty-nine (53.7%) of these genes were not previously identified as candidates in *N. meningitidis* strain MC58 (Saunders *et al.*, 2000).

Five of the initially described genes (NMA0402, NMA0419, NMA0072, an unassigned non-coding region and NMA1996) were excluded from this list because of the nature and/or location of the repeats. The repeat associated with NMA0402 (*truA*) is neither

within the coding region nor does it appear to be associated with a promoter. NMA0419 is associated with a (GC)₆ repeat towards the 3' end of the gene where length variation would not alter the expression of most of the gene and alteration within which would actually increase the length of the final gene product. Since this is an untypical context for a repeat associated with phase variation and also since (GC)_n repeats would have a relatively high T_m and have not yet been demonstrated to be associated with phase variation, this gene was excluded. On similar grounds NMA0072 (*pdxA*) with a (GC)₆ repeat was also eliminated, as was an additional gene containing a (GC)₆ repeat that was not included in the previously published list from strain Z2491 (NMA0263 *lpxC/envA*). Parkhill *et al.* (2000) were correct in ascribing a (G)₁₁ to a non-coding intergenic sequence which was therefore not reported in this study. Finally, the (67 bp)₃ repeat within NMA1996 (*natD*) is located in the 3' end of the gene and is longer than repeats currently associated with phase variation.

Twenty-one of the candidate genes in strain Z2491 (listed in Table 1) were not identified by either of the previous whole-genome analyses of *N. meningitidis* (Parkhill *et al.* 2000; Saunders *et al.* 2000). Seven of these are hypothetical proteins: NMA0640, NMA1313, NMA1373, NMA1436, NMA1457, NMA1562 and NMA2111. NMA1707a was not annotated by Parkhill *et al.* (2000) and is located between NMA1707 and NMA1708. Fourteen of the identified genes have defined functions or have been assigned putative functions on the basis of homology. One of these, NMA1371 (*vacB*) is a homologue of genes associated with virulence through their involvement in host-cell colonization (Tobe *et al.*, 1992; Ruiz-Lozano & Bonfante, 2000) and is the type of virulence-associated gene typically associated with phase variation. A number of the other genes are apparently metabolic function-related genes indicative of a role of phase variation in other forms of niche adaptation. These include: NMA0555 (*ppk*) which encodes polyphosphate kinase and is responsible for the synthesis of polyphosphate (Tinsley & Gotschlich, 1995); NMA0831, a homologue of *potD* from *E. coli*, a polyamine-binding periplasmic protein which is part of an ABC transport system for uptake of polyamine (Antognoni *et al.*, 1999); NMA2028, a homologue of *amiC*, which negatively regulates the amide-inducible aliphatic amidase operon in *Pseudomonas aeruginosa* (Wilson & Drew, 1991; O'Hara *et al.*, 1999); NMA2216, a homologue of *bioH*, which may be involved in biotin biosynthesis (Eisenberg, 1987); and NMA2103, a homologue of *thiL*, which is involved in thiamine biosynthesis (Imamura & Nakayama, 1982). One ORF in *N. meningitidis* strain Z2491 identified in this analysis is not currently annotated as frameshifted by the repeat tract, as suggested by this analysis. Therefore, NMA0286/NMA0285, a putative hydrolase, has two annotation numbers for the 5' and 3' ends of the gene. Also identified were NMA0074 (*gidA*), NMA0579 (a putative prolyl oligopeptidase family protein), NMA0611 (Mrp/NBP35 family pro-

Table 1. Putative phase-variable genes identified by genome analysis of *N. meningitidis* strain Z492 and comparison of the genes in the other sequences

Gene	NMA#	Repeat	FS*	ID P†	ID S‡	NMB#	Repeat	FS*	XNG#	Repeat	FS*
<i>pglA</i>	NMA0048	(G)14	-	yes	yes	NMB0218	(G)11	-	XNG1644	(G)11	-
<i>hypo</i>	NMA0132	(C)9	p	yes							
<i>hypo</i>	NMA0277	(A)11	p		yes	NMB0032	(A)11	p	XNG1934	(A)9	(p)
<i>hydrolase</i>	NMA0286/NMA0285	(C)7	+			NMB0040/NMB0039	(C)6	+	XNG1941	(C)6	+
<i>pilC1</i>	NMA0293	(G)11	+	yes	yes	NMB0049	(G)11	+	XNG1797	(G)11	+
<i>acetyltransferase</i>	NMA0407	(G)12	+	yes§							
<i>hpuA</i>	NMA0475	(G)10	-	yes					XNG2004	(G)9	+
<i>serine protease</i>	NMA0478	(C)10	-	yes	yes	NMB1969	(C)10	-	XNG0986	(G)6	dead
<i>prolyl oligopeptidase family protein</i>	NMA0579	(G)7	-			NMB1877	GGGGCGG	-	XNG0023	GGGGCGG	-
<i>pilC2</i>	NMA0609	(G)13	-	yes	yes	NMB1847	(G)14	+	XNG0049	(G)13	-
<i>wbpC</i>	NMA0619	(G)10	-	yes	yes	NMB1836	(G)13	-	XNG0060	(G)6	-
<i>hypo</i>	NMA0640	(C)7	-						XNG0080	(C)10	-
<i>conserved hypo</i>	NMA0641	CACTCCCT(C)16	+	yes					XNG0081	(C)7	-
<i>phage associated</i>	NMA0782	(C)7(N)10(G)7	-		yes	NMB1634	(C)8(N)9(G)7	dead	XNG1052	(C)8(N)10(G)7	-
<i>conserved hypo</i>	NMA0796	(C)7	-		yes	NMB0593	(C)7	-	XNG0158	(C)5	dead
<i>potD-2</i>	NMA0831	(C)7	-			NMB0623	AACCCCC	-	XNG0188	AACCCCC	-
<i>glycosyltransferase</i>	NMA0832	(CAAACAA)26(AT)3CTAT	-	yes	yes	NMB0624	(CAAACAA)34(AT)5	+	XNG0189	(CAAACAA)4(AT)5	-
<i>hsdSII</i>	NMA1040	(G)8	+	yes	yes	NMB0831	(G)7	+	XNG0388	(G)7	+
<i>phage associated</i>	NMA1167	(C)8(N)10(G)7	+		yes	NMB0970	(C)6(N)10(G)7	+	XNG1076	(C)8(N)10(G)7	+
<i>opc</i>	NMA1251	(C)17	p	yes	yes	NMB1053	(C)12	p			
<i>hsdM</i>	NMA1385	(A)9	dead		yes	NMB1223	(A)9	dead			
<i>glycosyltransferase</i>	NMA1427	(G)9	+	yes		NMB1255	(G)8	dead	XNG1041	¶	
<i>hypo</i>	NMA1436	(C)7 and (AC)4	-						XNG0663	CCCCCTCC and GC(AC)3	+
<i>conserved hypo</i>	NMA1457	(C)6	p			NMB1030	ATCCCT	p	XNG0532	ATCCCT	p
<i>mod Type III</i>	NMA1467	(CCCAA)8	+	yes	yes	NMB1261	(CCCAA)16	+	XNG0520	(CCCAA)12	+
<i>porA</i>	NMA1642	(G)12	p	yes	yes	NMB1429	(G)11	p	XNG0832	(T)11C(G)6	p
<i>dnaX</i>	NMA1656	(C)7	-			NMB1443	ACCCGCC	-	XNG0703	(C)7	-
<i>opa</i>	NMA1676	(CTTCT)7	+	yes							
<i>hypo</i>	NMA1707a	(CAAG)12	-/+ ††		yes	NMB1507	(CAAG)11	-/+ ††	XNG0887	(CAAG)4	-/+ ††
<i>phage associated</i>	NMA1792	(C)10(N)10(G)7	dead	yes	yes	NMB1543	(C)8(N)10(G)7	dead			
<i>opa</i>	NMA1890	(CTTCT)8	+	yes							
<i>hmbR</i>	NMA1925	(G)10	+	yes	yes	NMB1668	(G)9	-	XNG1216	GGCGGCCAA	dead
<i>fixP</i>	NMA1977	(AT)5	-		yes	NMB1723	(AT)5	-	XNG1268	(AT)5	-
<i>phage associated</i>	NMA1999	(C)6(N)10(G)7	+		yes	NMB0486	(C)6(N)10(G)7	+	XNG1175	(C)8(N)10(G)7	+
<i>amiC</i>	NMA2028	(C)7	-			NMB0456	TCCGCC	-	XNG1384	TCCGCC	-
<i>opa</i>	NMA2043	(CTTCT)11	+	yes							
<i>dca</i>	NMA2069	(G)9	+	yes	yes	NMB0415	(G)9	+	XNG1423	TGTGGGGG	-
<i>conserved hypo</i>	NMA2111	(C)7	-			NMB0377	(C)7	-	XNG1460	(C)7	-
<i>vapA</i>	NMA2175	(AAGC)8	- ††	yes	yes	NMB0312	(AAGC)9	+	XNG1563	(AAGC)3	dead
<i>bioH</i>	NMA2216	(G)6	-			NMB0270	(G)6	-	XNG1606	(G)6	-

* FS, frameshift. + and - indicate the presence or absence, respectively, of a frameshift in the ORF; p indicates a promoter-located repeat; (p) indicates sequence changes in the putative promoter region; dead indicates that the gene is not intact in this strain.

† ID P, identified by Parkhill *et al.* (2000).

‡ ID S, identified by Saunders *et al.* (2000).

§ Listed as NMA0406 in Parkhill *et al.* (2000).

|| These ORFs have limited homology to one another.

¶ piv-nm1B has been inserted into this ORF.

** Frameshift is not at the repeats.

†† There are two alternate reading frames, the termination codons for which are separated by 22 bp. NMA1707a and XNG0887 end in one, while NMB1507 ends in the other. On the basis of informatics alone the question of which are frameshifted cannot be resolved.

‡‡ Contains an in-frame termination codon.

tein), NMA0995 (*recB*), NMA1149 (*sucA*), NMA1221 (a phage integrase homologue) and NMA1656 (*dnaX*).

Whole-genome analysis of *N. gonorrhoeae* strain FA1090 for potential phase-variable genes

The genome sequence of *N. gonorrhoeae* strain FA1090 has been publicly available since 1997, and was completed, contiguated and publicly released in this form in September 2000 (GenBank no. AE004969; also <http://www.genome.ou.edu/gono.gb>). Seventy-two putative phase-variable genes, including the 11 *opa* genes, were identified in this gonococcal strain (Table 2 & Table 4). Excluding the *opa* genes, only 20 of these were previously identified between the analyses of Saunders *et al.* (2000) and Parkhill *et al.* (2000) in *N. meningitidis*. In total 39 genes not previously identified as phase-variable were found in this analysis of strain FA1090.

Once the *opa* genes, which are numerous in the gonococci but are also present in the meningococci, *lgtC* (XNG2047) and *lgtD* (XNG2048), which are LPS biosynthetic genes known to be present in other meningococcal strains (Yang & Gotschlich, 1996; GenBank no. U14554; Jennings *et al.*, 1999), and XNG1207a and

XNG1577, which are part of a paralogous family with XNG1788 present in meningococci (NMB0297 and NMA0212), are accounted for, there are currently 11 candidate phase-variable genes that are unique to *N. gonorrhoeae*. Only four of these are not hypothetical proteins: XNG1341, a putative adhesin which is inactivated in strain FA1090 by two frameshifts in addition to that caused by the (CAAG)20 repeat; and XNG0470, XNG1014 and XNG1513 which form a paralogous family of endodeoxyribonucleases (84.3% nucleotide identity). Of these paralogues, XNG0470 has a (G)7 HPT and is frameshifted while XNG1014 and XNG1513 have similarly located (G)6 HPTs and are in-frame. The remaining genes unique to the gonococcal strain FA1090 are XNG0503a, XNG0473, XNG1000, XNG1511, XNG1733, XNG1834 and XNG1856.

Comparisons using three genes that appear to be non-functional in this gonococcal sequence, but which are associated with repeats likely to mediate switching have revealed previously unrecognized potential phase-variable genes in *N. meningitidis*. XNG0400, a putative glycosyltransferase, is inactive in *N. gonorrhoeae* and not a candidate for phase variation on the basis of its

Table 2. Putative phase-variable genes identified by genome analysis of *N. gonorrhoeae* strain FA1090 and comparison of the genes in the other sequences

Gene	XNG#	Repeat	FS*	ID P†	ID S‡	NMB#	Repeat	FS*	NMA#	Repeat	FS*
<i>pilC2</i>	XNG0049	(G)13	-	yes	yes	NMB1847	(G)14	+	NMA0609	(G)13	-
<i>opa</i>	XNG0062	(CTTCT)13	-								
<i>opa</i>	XNG0065	(CTTCT)9	-								
conserved hypo	XNG0075	(C)7	-			NMB1824	CTCCCCC	-	NMA0635	TTCCCCC	-
<i>hypo</i>	XNG0080	(C)10	-						NMA0640	(C)7	-
conserved hypo	XNG0081	(C)7	-	yes					NMA0641	CACTCCCT(C)16	+
glycosyltransferase	XNG0189	(CAAACAC)4(AT)5	-	yes	yes	NMB0624	(CAAACAA)34(AT)5	+	NMA0832	(CAAACAA)26(AT)3CTAT	+
<i>hsdS</i> §	XNG0388	(G)7	+	yes	yes	NMB0831	(G)7	+	NMA1040	(G)8	+
<i>hypo</i>	XNG0412	(G)7	+						NMA1077	(G)6	-
conserved hypo	XNG0419	(C)6	+			NMB0872	(C)6	-	NMA1090	(C)6	-
<i>sodB</i>	XNG0431	(A)7	+			NMB0884	(A)6	-	NMA1104	(A)6	-
<i>azlC</i> related	XNG0439	(G)6	dead			NMB0892	(G)5	-	NMA1111	(G)5	-
<i>hypo</i>	XNG0470	(G)7	+								
<i>hypo</i>	XNG0473	(T)9	+								
<i>hypo</i>	XNG0503a	(C)6A(C)9	+								
<i>mod Type III</i>	XNG0520	(CCCAA)12	+	yes	yes	NMB1261	(CCCAA)16	+	NMA1467	(CCCAA)8	+
<i>lidD</i>	XNG0610	(G)7	-			NMB1377	GGGCAGG	-	NMA1592	GGGCAGG	-
<i>mod</i>	XNG0612	(GCCA)37	-		yes	NMB1375	(GCCA)20	+	NMA1589	(GCCA)3	+
<i>dnaX</i>	XNG0703	(C)7	-			NMB1443	ACCCGCC	-	NMA1656	(C)7	-
<i>porA</i>	XNG0832	(T)11C(G)6	p	yes	yes	NMB1429	(G)11	p	NMA1642	(G)12	p
<i>opa</i>	XNG0877	(CTTCT)16	+								
<i>opa</i>	XNG0952	(CTTCT)20	-								
<i>ppx</i>	XNG0954	(C)7	-			NMB1467	(C)6G	-	NMA1679	(C)6G	-
<i>opa</i>	XNG0987	(CTTCT)2	+								
<i>hypo</i>	XNG1000	(G)7	+								
<i>hypo</i>	XNG1014	(G)6	-								
phage associated	XNG1052	(C)8(N)10(G)7	-		yes	NMB1634	(C)8(N)9(G)7	dead	NMA0782	(C)7(N)10(G)7	-
phage associated	XNG1076	(C)8(N)10(G)7	+		yes	NMB0970	(C)6(N)10(G)7	+	NMA1167	(C)6(N)10(G)7	+
phage associated	XNG1175	(C)8(N)10(G)7	+		yes	NMB0486	(C)6(N)10(G)7	+	NMA1999	(C)6(N)10(G)7	+
<i>opa</i>	XNG1181	(CTTCT)11	+								
<i>hypo</i> ¶	XNG1207a	(G)8	+								
<i>fixP</i>	XNG1268	(AT)5	-		yes	NMB1723	(AT)5	-	NMA1977	(AT)5	-
adhesin	XNG1341	(CAAG)20	dead								
<i>opa</i>	XNG1352	(CTTCT)10	+								
<i>opa</i>	XNG1396	(CTTCT)12	-								
<i>opa</i>	XNG1437	(CTTCT)4	+								
conserved hypo	XNG1460	(C)7	-			NMB0377	(C)7	-	NMA2111	(C)7	-
<i>hypo</i>	XNG1511	(G)8	+								
<i>hypo</i>	XNG1513	(G)6	-								
<i>tspA</i>	XNG1534	(C)7	dead			NMB0341	(C)6	-	NMA2146	(C)6	-
<i>hypo</i> ¶	XNG1577	(G)8	+								
<i>pgiA</i>	XNG1644	(G)11	-	yes	yes	NMB0218	(G)11	-	NMA0048	(G)14	-
<i>hypo</i> **	XNG1733	(G)8	-								
<i>rplK</i>	XNG1740	(C)7	-			NMB0127	CTCCTCC	-	NMA0146	CTCCTCC	-
<i>opa</i>	XNG1746	(CTTCT)13	+								
<i>hypo</i> ¶	XNG1788	(G)8	+			NMB0297	AAGGA	+	NMA0212	AGGGA	+
<i>pilC1</i>	XNG1797	(G)11	+	yes	yes	NMB0049	(G)11	+	NMA0283	(G)11	+
<i>hypo</i>	XNG1834	(C)8	-								
<i>hypo</i>	XNG1856	(TA)5	-								
<i>map</i>	XNG1866	(C)7	+			NMB2093	(C)6	-	NMA0337	(C)6	-
<i>opa</i>	XNG1948	(CTTCT)10	+								
<i>lgtG</i>	XNG1968	(C)11	-		yes	NMB2032	(C)12	+			
<i>frpB</i>	XNG1989	(C)13	p		yes	NMB1988	(C)11	p	NMA0453	ATCCCTCTCTC	p
<i>hpaA</i>	XNG2004	(G)9	+	yes					NMA0475	(G)10	-
<i>lgtA</i>	XNG2045	(G)11	+		yes	NMB1929	(G)14	-	NMA0524	AGGAAT(G)5	-
<i>lgtC</i>	XNG2047	(G)14	+								
<i>lgtD</i>	XNG2048	(G)14	-								

* FS, frameshift. + and - indicate the presence or absence, respectively, of a frameshift in the ORF; p indicates a promoter-located repeat; dead indicates that the gene is sequence-inactivated in this strain.

† ID P, identified by Parkhill *et al.* (2000).

‡ ID S, identified by Saunders *et al.* (2000).

§ These ORFs have limited homology to one another.

|| XNG0470, XNG1014 and XNG1513 are homologues.

¶ XNG1207, XNG1577 and XNG1788 are homologues.

** There is weak homology to a small portion of NMA0586 and NMB1870.

(A)7 repeat alone. The meningococcal homologues are not only intact, but also have polymorphisms in the poly-A tract, which is associated with a frameshift in NMB0846 and is in-frame in NMA1058/NMA1057. The *azlC*-related protein, XNG0439, has a (G)6 in the frameshifted and degenerate gonococcal gene, while in the meningococci (NMB0892 and NMA1111) the (G)5 tract is in-frame. HPTs 5 bp in length are relatively

stable in the LPS genes of *N. meningitidis* (Jennings *et al.*, 1999), which suggests that other strains would have the potential for phase variation of this gene at high frequency only with longer repeats. In *N. gonorrhoeae* strain FA1090, *tspA* (XNG1534) has five frameshifts, one of which is caused by a (C)7 that in the two *N. meningitidis* strains is a (C)6 (NMB0341 and NMA2146). Although the gonococcal gene has multiple

frameshifts, the polymorphism in the HPT suggests the potential for phase variability of the *tspA* gene in *N. meningitidis*. TspA is a T-cell and B-cell stimulating antigen considered to be a possible vaccine candidate in *N. meningitidis* (Kizil *et al.*, 1999, GenBank no. AJ010113).

The putative phase-variable genes from *N. gonorrhoeae* include two genes that have been previously investigated in the *Neisseria* spp. but not previously reported to be phase-variable. *sodB* (XNG0431) is frameshifted in the region of an (A)7 in the FA1090 genome while the meningococcal sequences (NMB0884 and NMA1104) and the GenBank entry for *sodB* from an unspecified *N. gonorrhoeae* strain (GenBank no. AY010758) have (A)6 at the equivalent location. Variation associated with a repeat of this type has only been described in *porA* to date (van der Ende *et al.*, 2000). The other previously identified gene, *lldD* (XNG0610; previously *lldA*; Erwin & Gotschlich, 1996), encodes an L-lactate dehydrogenase. The phase variation of this gene has not previously been recognized.

In *Escherichia coli* and *Thermus thermophilus*, *dnaX* contains a repeat that has been demonstrated to generate ribosomal frameshifting two-thirds of the way into the gene, such that two proteins are translated from one gene, the γ and τ subunits of the DNA polymerase III holoenzyme (Blinkowa & Walker, 1990; Flower & McHenry, 1990; Tsuchihashi & Kornberg, 1990; Yurieva *et al.*, 1997). While the (C)7 identified this as a potential phase-variable gene in *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strain Z2491, this HPT is located one-third of the way through the coding sequence. A second C HPT, a (C)6, is present in two of the *Neisseria* spp. in a location similar to the poly-A tracts of *E. coli* and *T. thermophilus*, but is not present in meningococcal strain MC58. This suggests that these tracts are not involved in the joint generation of equivalent γ and τ subunits of the DNA polymerase III holoenzyme in strain MC58, although their possible instability in the other strains would need further experimental investigation.

It is interesting to note that the analysis of the gonococcal genome indicates that a homologue of *ppx*, the gene encoding exopolyphosphatase, is a phase-variable candidate (XNG0954), while in the meningococcal genomes it is not (NMB1467 and NMA1679). In the meningococcal genomes, however, the *ppk* gene (NMB1900 and NMA0555), encoding polyphosphate kinase, is potentially phase-variable while the gonococcal *ppk* is not (XNG0003). PPX hydrolyses polyphosphate (polyP) to orthophosphate (P_i) and PPK synthesizes polyP from P_i (Akiyama *et al.*, 1993). It appears that each species has the potential to phase-vary one of these two opposing enzymes.

Comparative analysis of the potential phase-variable genes of *N. meningitidis* strain MC58 revisited

Several aspects of the original analysis of *N. meningitidis* strain MC58 can be reappraised. In Table 1 of Saunders

et al. (2000), each repeat identified in *N. meningitidis* strain MC58 was evaluated for its presence in the other two *Neisseria* spp. genome sequences, including the then incomplete sequence of *N. gonorrhoeae* strain FA1090. The other sequences were classified as having a repeat of the same length, a polymorphic repeat (i.e. the repeat is of a different length), no repeat at the equivalent location or no homologue of the gene containing the repeat. These limited designations have been extended in this analysis, in which the NMA ORF designations for *N. meningitidis* strain Z2491 have been included to facilitate cross-comparisons (Table 1, Table 2 & Table 3), as have the actual sequences corresponding to the repeats (Table 3). In addition, two candidates have been removed from the original *N. meningitidis* strain MC58 list (Saunders *et al.* 2000). These are NMB0300 and NMB1140. NMB0300 is a hypothetical protein homologous to NMA2186. Both contain a G6 HPT but only the serogroup A sequence for this gene is intact. In the serogroup B strain MC58, NMB0300 is a dead gene that is frameshifted, but not at the (G)6 repeat. NMB1140, encoding the cell cycle protein *mesJ*, has been removed because it is consistently associated with a repeat tract of (A)8 in all three sequenced strains, it is probably essential and alternative annotation now suggests that this gene is probably not frameshifted.

The more comprehensive comparative analysis reveals that six genes described with either polymorphic repeats or containing no repeat in the other sequences from strain MC58 (Saunders *et al.* 2000) are degenerate in one of the other two sequenced strains. The haemoglobin receptor gene *hmbR* (NMB1668) is degenerate in *N. gonorrhoeae* strain FA1090, as are the serine protease (NMB1969) and the function-unknown genes NMB0312 and NMB0593. The hypothetical protein NMB1786 is non-functional in *N. meningitidis* strain Z2491. Lactoferrin-binding protein gene *lbpB* and the 5' region of the associated *lbpA* (NMB1540), which contains the (G)8 repeat, are not present in *N. gonorrhoeae* strain FA1090. The remaining fragment of *lbpA* contains two frameshifts. The association between 'dead' genes and phase variation may not be co-incidental. While it is possible for a phase-variable gene to be essential for some part of the normal life cycle of the organism, it is self evident that a gene that can be switched off in a programmed way cannot be essential under all conditions. If a mutation were to occur in a gene while it was not adaptive or while switched off when transcription-coupled repair might also be reduced, then these genes might be expected to be more prone to gene loss than others. Even phase-variable genes for which host interactions are well established, such as *opc*, can be lost in this way as in *N. meningitidis* strain S3446 (Saunders, 1999). It may be that phase-variable genes therefore represent a group of genes more prone to degeneration than others and the observation of 'dead' candidate genes may be supportive of their phase variability in other strains.

Some of the genes are described differently in the different annotations. The 'adhesion and penetration

Table 3. Modified putative phase-variable genes in *N. meningitidis* strain MC58 and comparison of the genes in the other sequences

Gene	NMB#	Repeat	FS*	NMA#	Repeat	FS*	XNG#	Repeat	FS*
hypo	NMB0032	(A)11	p	NMA0277	(A)11	p	XNG1934	(A)9	(p)
<i>pilC1</i>	NMB0049	(G)11	+	NMA0293	(G)11	+	XNG1797	(G)11	+
hypo	NMB0065	(T)10	+	-----	-----	-----	-----	-----	-----
<i>siaD</i>	NMB0067	(C)7	-	-----	-----	-----	-----	-----	-----
<i>mtfB</i>	NMB0098	(G)7	+	-----	-----	-----	-----	-----	-----
<i>pglA</i>	NMB0218	(G)11	-	NMA0048	(G)14	-	XNG1644	(G)11	-
<i>vapA</i>	NMB0312	(AAGC)9	+ †	NMA2175	(AAGC)8	- †	XNG1563	(AAGC)3	dead
hypo	NMB0368	(A)11	-	NMA2117	(A)8	-	XNG1281	(A)7	-
<i>dca</i>	NMB0415	(G)9	+	NMA2069	(G)9	+	XNG1423	TGTGGGGG	-
<i>opa</i>	NMB0442	(CTTCT)10	+	-----	-----	-----	-----	-----	-----
phage associated	NMB0486	(C)6(N)10(G)7	+	NMA1999	(C)6(N)10(G)7	+	XNG1175	(C)8(N)10(G)7	+
hypo ‡	NMB0488	(C)7	-	-----	-----	-----	XNG0973	(C)6	-
conserved hypo	NMB0593	(C)7	-	NMA0796	(C)7	-	XNG0158	(C)5	dead
glycosyltransferase	NMB0624	(CAAACAA)34	+	NMA0832	(CAAACAA)26	-	XNG0189	(CAAACAA)4(AT)5	-
<i>hsdS</i> §	NMB0831	(G)7	+	NMA1040	(G)8	+	XNG0388	(G)7	+
<i>opa</i>	NMB0926	(CTTCT)16	+	-----	-----	-----	-----	-----	-----
<i>funZ</i>	NMB0961	(CAAAAT)5	-	-----	-----	-----	-----	-----	-----
phage associated	NMB0970	(C)6(N)10(G)7	+	NMA1167	(C)6(N)10(G)7	+	XNG1076	(C)8(N)10(G)7	+
Type II restriction enzyme	NMB1032	(C)6	-	NMA1454	CCTCCC	-	XNG0643	CCTCCC	-
<i>opc</i>	NMB1053	(C)12	p	NMA1251	(C)17	p	-----	-----	-----
<i>ner</i>	NMB1080	(A)7	+	-----	-----	-----	-----	-----	-----
<i>hsdM</i>	NMB1223	(A)9	dead	NMA1385	(A)9	dead	-----	-----	-----
<i>mod</i> Type III	NMB1261	(CCCAA)16	+	NMA1467	(CCCAA)8	+	XNG0520	(CCCAA)12	+
conserved hypo	NMB1265	(GAAA)3	-	-----	-----	-----	-----	-----	-----
hypo	NMB1275	(AGCA)3	+ or -	NMA1480	(AGCA)3	+ or -	XNG0508	AACCGGCAAACA	dead or -
<i>mod</i>	NMB1375	(GCCA)20	+	NMA1589	(GCCA)3	+	XNG0612	(GCCA)37	-
<i>nifS</i>	NMB1379	(C)8	-	NMA1594	CCACCACC	-	XNG0608	CCACCACC	-
RTX-type toxin	NMB1407	(ATAACAA)4	+	-----	-----	-----	-----	-----	-----
<i>porA</i>	NMB1429	(G)11	p	NMA1642	(G)12	p	XNG0832	(T)11C(G)6	p
<i>opa</i>	NMB1465	(CTTCT)13	+	-----	-----	-----	-----	-----	-----
hypo ‡	NMB1489	(C)7	-	-----	-----	-----	-----	-----	-----
hypo	NMB1507	(CAAG)11	+/-	NMA1707a	(CAAG)12	-/+	XNG0887	(CAAG)4	-/+
<i>virG</i> -related	NMB1525	(AAGC)5	+	NMA1725	(AAGC)4	+	XNG0905	(AAGC)3	+ †
<i>lbpA</i>	NMB1540	(G)8	-	NMA1739	CGAAACGG	-	XNG0245	-----	dead
phage associated	NMB1543	(C)8(N)10(G)7	p	NMA1792	(C)10(N)10(G)7	p	-----	-----	-----
phage associated	NMB1634	(C)8(N)9(G)7	p	NMA0782	(C)7(N)10(G)7	-	XNG1052	(C)6(N)10(G)7	-
<i>opa</i>	NMB1636	(CTTCT)11	+	-----	-----	-----	-----	-----	-----
<i>hmbR</i>	NMB1668	(G)9	-	NMA1925	(G)10	+	XNG1216	GGCGGCCAA	dead
<i>fixP</i>	NMB1723	(AT)5	-	NMA1977	(AT)5	-	XNG1268	(AT)5	-
glutaredoxin 2	NMB1734	(TGCG)3	-	NMA1990	(TGCG)3	-	XNG1278	(TGCG)3	-
phage associated	NMB1741	(C)6(N)10(G)7	+	-----	-----	-----	-----	-----	-----
conserved hypo	NMB1760	(C)6	+	-----	-----	-----	-----	-----	-----
<i>cvaA</i>	NMB1783	(C)5	+	NMA0683	(C)5	+	XNG0115	(C)4	-
<i>wbpC</i>	NMB1836	(G)13	-	NMA0619	(G)10	-	XNG0060	(G)6	-
<i>pilC2</i>	NMB1847	(G)14	+	NMA0609	(G)13	-	XNG0049	(G)13	-
hypo	NMB1893	(TTCC)4	+	NMA0562	(TTCC)4	+	-----	-----	-----
<i>plsX</i>	NMB1913	(TTCC)3	-	NMA0542	(TTCC)3	-	XNG2061	(TTCC)3	-
<i>lgtA</i>	NMB1929	(G)14	-	NMA0524	AGGAAT(G)5	-	XNG2045	(G)11	+
hypo	NMB1931	(G)7	+	NMA0522	GGCGGGGG	-	-----	-----	-----
serine protease	NMB1969	(C)10	-	NMA0478	(C)10	-	XNG0986	(G)6	dead
<i>hap</i>	NMB1985	(GGCA)3	-	NMA0457	CTCAGGACACCA	-	XNG2001	(GGCA)3	-
<i>trpB</i>	NMB1988	(C)11	p	NMA0453	ATCCCCCTCTCTC	p	XNG1989	(C)13	p
<i>yop1/yadA</i> related	NMB1994	(TAAA)9	p	-----	-----	-----	-----	-----	-----
related to adhesion/invasion & IgA protease	NMB1998	(C)9	-	-----	-----	-----	-----	-----	-----
ABC transporter	NMB2008	(C)7	p	-----	-----	-----	-----	-----	-----
<i>lgtG</i>	NMB2032	(C)12	+	-----	-----	-----	XNG1968	(C)11	-
<i>mafA-3</i>	NMB2104	(G)6	+	NMA0325	(G)5	-	XNG1858	(G)5	-

* FS, frameshift. + and - indicate the presence or absence, respectively, of a frameshift in the ORF; p indicates a promoter-located repeat; (p) indicates sequence changes in the putative promoter region; dead indicates that the gene is sequence-inactivated in this strain.

† Contains an in-frame termination codon.

‡ NMB0488 and NMB1489 are homologues.

§ These ORFs have limited homology to one another.

|| There are two alternate reading frames, the termination codons for which are separated by 22 bp. NMA1707a and XNG0887 end in one, while NMB1507 ends in the other. On the basis of informatics alone the question of which are frameshifted cannot be resolved.

protein homologue'/NMB1985 is included in this paper as *hap* and the 'saccharide acetylase'/NMB1836 is listed as *wbpC*. The 'cell adhesion molecule'/NMB2104 has been included as *mafA-3* and does not contain the same (G)6 repeat and frameshift in strain Z2491 as it does in strain MC58. The hypothetical protein NMB1275/NMA1480 has an (AGCA)3 repeat in both meningococcal sequences, while the gonococcal sequence differs at four bases such that it is not repetitive

(AACCGGCAAACA). The two meningococcal genes are annotated differently and comparisons including the gonococcal sequence now suggest that this gene is not frameshifted (in agreement with NMA1480). On the basis of the differences between the sequences and the presence of a tetramer frequently associated with unstable repeats this gene remains a candidate but should no longer be regarded as 'strong' on the basis of the sequence evidence alone.

Table 4. *opa* genes of three pathogenic *Neisseria* spp.

Strain	ORF	Repeat	FS*
<i>N. meningitidis</i> strain MC58			
	NMB0442	(CTTCT)10	+
	NMB1636	(CTTCT)11	+
	NMB1465	(CTTCT)13	+
	NMB0926	(CTTCT)16	+
<i>N. meningitidis</i> strain Z2491			
	NMA1676	(CTTCT)7	+
	NMA1890	(CTTCT)8	+
	NMA2043	(CTTCT)11	+
<i>N. gonorrhoeae</i> strain FA1090			
	XNG0987	(CTTCT)2	+
	XNG1437	(CTTCT)4	+
	XNG0065	(CTTCT)9	-
	XNG1352	(CTTCT)10	+
	XNG1948	(CTTCT)10	+
	XNG1181	(CTTCT)11	+
	XNG1396	(CTTCT)12	-
	XNG0062	(CTTCT)13	-
	XNG1746	(CTTCT)13	+
	XNG0877	(CTTCT)16	+
	XNG0952	(CTTCT)20	-

* FS, frameshift. + and - indicate the presence or absence, respectively, of a frameshift in the ORF*.

Comparative analysis of the repeat-associated genes in the three genome sequences

In *N. gonorrhoeae*, 11 of the 72 candidates are the *opa* genes, of which the *N. meningitidis* strains MC58 and Z2491 have four and three, respectively (Table 4). Excluding the *opa* genes on the basis of the current analysis, *N. meningitidis* strain Z2491 has 51, *N. gonorrhoeae* strain FA1090 has 61 and *N. meningitidis* strain MC58 has 59 potential phase-variable genes. The previous large difference between the number of phase-variable genes between the serogroup A and B genome sequences is therefore more a reflection of the different analyses than the different strains. In addition to the *opa* genes (Table 4), cross comparisons reveal there to be a total repertoire of 119 distinct potentially phase-variable genes in the three *Neisseria* spp. strains investigated (Table 5).

In *N. meningitidis* strain Z2491, an additional seven candidates were found by comparison with *N. gonorrhoeae* strain FA1090 (NMA0337, NMA0954, NMA1077, NMA1090, NMA1589, NMA1679 and NMA2146) and a further seven from comparison with *N. meningitidis* strain MC58 (NMA0542, NMA0562, NMA0683, NMA1480, NMA1725, NMA1990 and NMA2146). This brings the total of candidate phase-variable genes to 68 in *N. meningitidis* strain Z2491, including 22 of the 27 previously reported (Parkhill *et al.*, 2000).

Comparative analysis identified 11 additional candidates in *N. gonorrhoeae* strain FA1090. Five of these are from comparison with *N. meningitidis* strain Z2491 (homologues of NMA0286/NMA0285, NMA0478,

Table 5. Repertoire of phase-variable genes in the pathogenic *Neisseria* spp.

Gene	Candidacy*	NMB#†	NMA#†	XNG#†
Surface associated proteins				
fbp	K	NMB1988	(NMA0453)	XNG1989
fbpR	K	NMB1988	NMA1925	(XNG1216)
hpaA	K	-----	NMA0475	XNG2004
hpaA	K	NMB1540	(NMA1739)	(XNG0245)
opc	K	NMB1053	NMA1251	-----
picC1	K	NMB0049	NMA0253	XNG1787
picC2	K	NMB1847	NMA0609	XNG0049
porA	K	NMB1429	NMA1642	XNG0832
adhesin	S	-----	-----	XNG1341
mafA-3	S	NMB2104	(NMA0325)	(XNG1858)
related to adhesion/invasion & IgA protease	S	NMB1998	-----	-----
rspA	S	NMB0341	NMA2146	XNG1534
hap	M	NMB1985	(NMA0457)	XNG2001
osp 1/yadA related	M	NMB1994	-----	-----
Surface sugar biosynthesis				
igfA	K	NMB1929	(NMA0524)	XNG2045
igfC	K	-----	-----	XNG2047
igfD	K	-----	-----	XNG2048
igfG	K	NMB2032	-----	XNG1988
pgiA	K	NMB0218	NMA0048	XNG1644
slsD	K	NMB0067	-----	-----
glycosyltransferase	S	NMB0624	NMA0832	XNG0189
glycosyltransferase	S	NMB0846	(NMA1058/NMA1057)	XNG0400
glycosyltransferase	S	NMB1255	NMA1427	(XNG1041)
wbpC	S	NMB1836	NMA0619	XNG0080
Toxin and secreted enzyme related				
RTX-type toxin	S	NMB1407	-----	-----
serine protease	S	NMB1969	NMA0478	XNG0986
Restriction modification proteins				
hdsS	S	NMB0831	NMA1040	XNG0388
mod	S	NMB1375	NMA1589	XNG0612
mod Type III	S	NMB1261	NMA1467	XNG0520
hdsM	M	NMB1223	NMA1385	-----
Type II restriction enzyme	M	NMB1032	(NMA1454)	(XNG0643)
Others				
acetyltransferase	S	-----	NMA0407	-----
amiC	S	(NMB0456)	NMA2028	(XNG1384)
azlC related	S	(NMB0892)	NMA1111	XNG0439
cvaA	S	NMB1763	NMA0593	(XNG0115)
dca	S	NMB0415	NMA2069	(XNG1423)
dnaX	S	(NMB1443)	NMA1656	XNG0703
flp	S	NMB1723	NMA1977	XNG1288
funZ	S	NMB0061	-----	-----
hydrolase	S	NMB0040/NMB0039	NMA0286/NMA0285	XNG1941
ldd	S	(NMB1377)	(NMA1592)	XNG0610
map	S	NMB2093	NMA0337	XNG1868
mfb	S	NMB0086	-----	-----
nar	S	NMB1080	-----	-----
ritS	S	NMB1379	(NMA1594)	(XNG0608)
phase associated	S	NMB0486	NMA1999	XNG1175
phase associated	S	NMB0970	NMA1167	XNG1076
phase associated	S	NMB1543	NMA1792	-----
phase associated	S	NMB1634	NMA0782	XNG1052
phase associated	S	NMB1741	-----	-----
poilD-2	S	(NMB0823)	NMA0831	(XNG0188)
ppx	S	NMB1467	NMA1679	XNG0654
prolyl oligopeptidase family protein	S	(NMB1877)	NMA0579	(XNG0023)
rpK	S	(NMB0127)	(NMA0148)	XNG1740
sodB	S	(NMB0884)	(NMA1104)	XNG0431
vpsA	S	NMB0312	NMA0175	XNG1583
virG-related	S	NMB1525	NMA1725	XNG0905
ABC transporter	M	NMB2008	-----	-----
bioH	M	NMB0270	NMA2216	XNG1806
glutaredoxin 2	M	NMB1734	NMA1990	XNG1278
plxX	M	NMB1913	NMA0542	XNG2061
Proteins of unknown function and hypothetical proteins				
function unknown	S	-----	NMA0641	XNG0091
function unknown	S	NMB0593	NMA0796	(XNG0158)
function unknown	S	NMB1760	-----	-----
function unknown	S	(NMB1824)	(NMA0635)	XNG0075
function unknown	M	NMB0377	NMA2111	XNG1460
function unknown	M	NMB0872	NMA1090	XNG0419
function unknown	M	(NMB1030)	NMA1457	(XNG0532)
function unknown	M	NMB1265	-----	-----
hypothetical	S	-----	-----	XNG0470
hypothetical	S	-----	-----	XNG0473
hypothetical	S	-----	-----	XNG0503a
hypothetical	S	-----	-----	XNG1000
hypothetical	S	-----	-----	XNG1014
hypothetical	S	-----	-----	XNG1207a
hypothetical	S	-----	-----	XNG1511
hypothetical	S	-----	-----	XNG1513
hypothetical	S	-----	-----	XNG1577
hypothetical	S	-----	-----	XNG1733
hypothetical	S	-----	-----	XNG1834
hypothetical	S	-----	NMA0132	-----
hypothetical	S	-----	NMA0640	XNG0080
hypothetical	S	-----	NMA1077	XNG0412
hypothetical	S	-----	NMA1436	(XNG0663)
hypothetical	S	NMB0032	NMA0277	(XNG1934)
hypothetical	S	NMB0065	-----	-----
hypothetical	S	(NMB0297)	(NMA0212)	XNG1788
hypothetical	S	NMB0368	(NMA2117)	(XNG1281)
hypothetical	S	NMB0488	-----	XNG0973
hypothetical	S	NMB1489	-----	-----
hypothetical	S	NMB1507	NMA1707a	XNG0887
hypothetical	S	NMB1863	NMA0562	-----
hypothetical	S	NMB1931	(NMA0522)	-----
hypothetical	M	-----	-----	XNG1856
hypothetical	M	NMB1275	NMA1480	(XNG0508)

* The predicted likelihood of phase variability. K indicates a previously recognized phase-variable gene; S indicates a strong candidate for phase variation; M indicates a moderate candidate.

† ORF designation numbers listed in parentheses indicate equivalent genes that are present in the strain but are not considered candidates for phase variation.

NMA0619, NMA2175 and NMA2216) and six candidates from analysis of strain MC58 (homologues of NMB0488, NMB1507, NMB1525, NMB1734, NMB1913 and NMB1985). There are, therefore, 83 potential phase-variable genes in *N. gonorrhoeae* strain FA1090.

Comparison with the findings from the other two *Neisseria* spp. genomes reveals 19 more phase-variable

gene candidates in *N. meningitidis* strain MC58, making 82 in total in this strain. This 23% increase over the single genome analysis of Saunders *et al.* (2000) emphasizes the way in which comparative genomic analysis can expand the available information from what is accessible from a single genome sequence. These additional *N. meningitidis* strain MC58 candidates are NMB0040/NMB0039, NMB0270, NMB0341, NMB0377, NMB0385, NMB0741, NMB0785, NMB0846, NMB0872, NMB0955, NMB1001, NMB1202, NMB1255, NMB1350, NMB1467, NMB1846, NMB1900, NMB2093 and NMB2145.

By comparing the results from the three analyses of the *Neisseria* spp. genome sequences, likely candidacy as phase-variable genes can be strengthened by evidence of interstrain and interspecies polymorphisms in the repeat tract. The repeat polymorphism data from this study alone presents compelling evidence of phase variation for nine genes (in addition to the *opa* genes) already known to be phase-variable: *hpuA* (Chen *et al.*, 1998), *pglA* (Jennings *et al.*, 1998), *opc* (Sarkari *et al.*, 1994), *porA* (van der Ende *et al.*, 1995), *bmbR* (Richardson & Stojiljkovic, 1999), *pilC2* (Jonsson *et al.*, 1991), *lgtA* (Yang & Gotschlich, 1996), *frpB* (renamed *fetA* by Carson *et al.*, 2000) and *lgtG* (Banerjee *et al.*, 1998). Twenty-nine ORFs not previously known to be phase-variable also have polymorphisms in the length of the repeat between the strains that make them clear candidates for translational phase variation and two candidates for transcriptional phase variation (Table 5). Although not evident from the sequenced *N. meningitidis* strains, interstrain variations in the HPT of NMB0415, *dca*, has also been demonstrated previously (Snyder *et al.*, 2001).

As previously discussed the candidature of different genes varies with the length and composition of the repeat and the availability of polymorphic information (Saunders *et al.*, 2000). With respect to the HPTs there is evidence for instability in C or G repeats of 7 bp in length (Hammerschmidt *et al.*, 1996), whereas similar repeats of 5 bp are relatively stable (Jennings *et al.*, 1999). A threshold of instability at 7 bp in length is also supported by Markov chain analysis (Saunders *et al.*, 2000). On these grounds (C or G)₆ repeats have not been identified as candidates unless they are also associated with a frameshift in which a longer repeat would be associated with a functional gene and/or variation in length or sequence polymorphism in one of the other sequences. (C or G)₇ repeats are accordingly only considered to be moderate candidates unless they are associated with a frameshift or similar variability between the sequences. From this analysis, all candidate dinucleotide repeats with four copies are invariant between the strains. Further, variation of this type and length of repeat has not been demonstrated in this species, although dinucleotide repeats composed of AT/TA, CT/GA or CA/GT pairs are unstable in other Gram-negative species (van Ham *et al.*, 1993; Saunders *et al.*, 1998; Peck *et al.*, 1999; Eckert & Yan, 2000). Given the weaker candidature of genes associated with

Table 6. Reduced likelihood candidates with four copies of dinucleotide repeats

Gene	NMB#*	NMA#*	XNG#*
function unknown	(NMB0053)	(NMA0217)	XNG1792
<i>omp85</i>	NMB0182	NMA0085	XNG1682
<i>gidA</i>	(NMB0193)	NMA0074	XNG1669
<i>thiL</i>	NMB0385	NMA2103	XNG1453
function unknown	NMB0432	(NMA2053)	(XNG1406)
function unknown	NMB0471	NMA2014	XNG1369
Type II restriction enzyme	NMB0726	-----	-----
function unknown	NMB0741	NMA0954	XNG0305
<i>recB</i>	NMB0785	NMA0995	(XNG0351)
<i>sucA</i>	NMB0955	NMA1149	XNG0850
<i>purH</i>	(NMB0983)	(NMA1182)	XNG1353
phage integrase	NMB1001	NMA1221	XNG0695
hypothetical	(NMB1094)	NMA1313	-----
<i>vacB</i>	(NMB1200)	NMA1371	(XNG0754)
hypothetical	NMB1202	NMA1373	-----
transporter	NMB1277	NMA1483	XNG0506
hypothetical	NMB1350	NMA1562	-----
function unknown	(NMB1626)	(NMA1799)	XNG1521
hypothetical†	NMB1786	NMA0679/NMA0678	(XNG0112)
Mrp/NBP35 family	NMB1846	NMA0611	XNG0050
<i>ppk</i>	NMB1900	NMA0555	(XNG0003)
hypothetical	NMB2145	-----	XNG1825
transmembrane transport	-----	(NMA0047)	XNG1643
hypothetical	-----	(NMA2137)	XNG1487

* ORF designation numbers listed in parentheses indicate equivalent genes that are present in the strain but are not considered candidates for phase variation.

† While the (CA)₄ repeat is present in all three neisserial strains, the associated (A) HPT is variable and may be involved in transcriptional phase variation.

these repeats, the 24 candidates associated with these repeats have been placed in a separate table (Table 6). An indication of the likelihood of phase variability of each of the 95 remaining candidates (in addition to the *opa* genes) is listed in Table 5. Of these, 82 (86.3%) are either known phase-variable genes (14) or are particularly strong candidates for phase variation (68). Of these 95 candidates, 42 contain frameshifts (44.2%) and 14 are dead genes (14.7%) in at least one of the strains analysed, while 8 candidates have promoter-associated repeats.

In addition to the sheer number of genes identified, another noteworthy feature is the diversity of functions that are observed in these genes. While a significant number (31 of 61 with known functions or homologies; 50.8%) of genes fall within the functional categories traditionally associated with phase variation and virulence – namely surface proteins, LPS and sugar metabolism genes, toxins and restriction modification genes – a large number of the remainder (30 of 61; 49.2%) appear to have other functions (Table 5). This in part probably reflects the success of this approach in identifying the role of these genes in as yet undefined aspects of virulence. In addition, it also probably reflects the true role of phase variation as a general mechanism of environmental and niche adaptation, of which changes due to colonization niches and immune responses are only an important subset.

Although the difference between the meningococcal strains is not as great as was initially indicated from the original sequence papers, there is still a significant difference between the two sequenced strains. To put

Fig. 1. Venn diagram showing the distribution of different candidate genes in the three genomes.

this in context, there is less difference in the total repertoire of variable genes between the gonococcus (83 candidates) and the meningococcal strain MC58 (82 candidates) than there is between serogroup A (68 candidates) and B meningococci. Of the candidate genes that are not common to all three genomes, two are unique to strain Z2491 (NMA0132 and NMA0407) while the other sequenced strains have 13 (*N. gonorrhoeae* strain FA1090) and 14 (*N. meningitidis* strain MC58) unique candidates. It should be noted that there is no clear distinction between which genes are present in each of the species, for seven candidate genes are common only to the gonococcus and the meningococcal strain Z2491 and three are common only to the gonococcus and meningococcal strain MC58 (Fig. 1). This suggests that there is a substantial degree of horizontal transfer between these species and strains that affects the gene complement as well as the more recognized mosaicism, or that there have been multiple acquisition events from a common source. It is therefore meaningful to analyse and consider the genes of this type in these species as a whole, which further increases the potential for diversification and adaptability in these species. That there are more than 100 such genes, including over 80 known or strong candidates within only three strains, indicates that this type of switching event is central to the behaviour, adaptability and virulence of these organisms in a way that is unprecedented in other species studied to date.

ACKNOWLEDGEMENTS

L.A.S.S. is supported by the E. P. Abraham Trust. N.J.S. is supported by a Wellcome Trust Advanced Research Fellowship. This work is supported by the Wellcome Trust (N.J.S.). The *N. gonorrhoeae* sequence was obtained from the University of Oklahoma, the Gonococcal Genome Sequencing Project which was supported by USPHS/NIH grant no. AI-38399 (L. A. Lewis,¹ A. F. Gillaspay,¹ R. E. McLaughlin,¹ M. Gipson,¹ T. Ducey,¹ T. Ownbey,¹ K. Hartman,¹ C. Nydick,¹ M. Carson,¹ J. Vaughn,¹ C. Thomson,¹ L. Song,² S. Lin,² X. Yuan,² F. Najar,² M. Zhan,² Q. Ren,² H. Zhu,² S. Qi,² S. M. Kenton,² H. Lai,² J. D. White,² S. Clifton,² B. A. Roe² &

D. W. Dyer¹; ¹The University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; ²The University of Oklahoma, Norman, OK, USA).

REFERENCES

- Akiyama, M., Crooke, E. & Kornberg, A. (1993). An exopolyphosphatase of *Escherichia coli*. The enzyme and its *ppx* gene in a polyphosphate operon. *J Biol Chem* **268**, 633–639.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Antognoni, F., Del Duca, S., Kuraishi, A., Kawabe, E., Fukuchi-Shimogori, T., Kashiwagi, K. & Igarashi, K. (1999). Transcriptional inhibition of the operon for the spermidine uptake system by the substrate binding protein PotD. *J Biol Chem* **274**, 1942–1948.
- Appelmek, B. J., Martin, S. L., Monteiro, M. A. & 10 other authors (1999). Phase variation in *Helicobacter pylori* lipopolysaccharide due to changes in the lengths of poly(C) tracts in alpha3-fucosyltransferase genes. *Infect Immun* **67**, 5361–5366.
- Appelmek, B. J., Martino, M. C., Veenhof, E. & 7 other authors (2000). Phase variation in H type I and Lewis a epitopes of *Helicobacter pylori* lipopolysaccharide. *Infect Immun* **68**, 5928–5932.
- Banerjee, A., Wang, R., Uljon, S. N., Rice, P. A., Gotschlich, E. C. & Stein, D. C. (1998). Identification of the gene (*lgtG*) encoding the lipooligosaccharide beta chain synthesizing glucosyl transferase from *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* **95**, 10872–10877.
- Blinkowa, A. L. & Walker, J. R. (1990). Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res* **18**, 1725–1729.
- Bowler, L. D., Zhang, Q. Y., Riou, J. Y. & Spratt, B. G. (1994). Interspecies recombination between the *penA* genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis*: natural events and laboratory simulation. *J Bacteriol* **176**, 333–337.
- Carson, S. D., Stone, B., Beucher, M., Fu, J. & Sparling, P. F. (2000). Phase variation of the gonococcal siderophore receptor FetA. *Mol Microbiol* **36**, 585–593.
- Chen, C. J., Elkins, C. & Sparling, P. F. (1998). Phase variability of hemoglobin utilization in *N. gonorrhoeae*. *Infect Immun* **66**, 987–993.
- Durbin, R. & Thierry-Mieg, J. T. (1991). *A. C. elegans* DataBase. Documentation, code and data available from <http://www.acedb.org>
- Eckert, K. A. & Yan, G. (2000). Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: influence of sequence on expansion mutagenesis. *Nucleic Acids Res* **28**, 2831–2838.
- Eisenberg, M. A. (1987). Biosynthesis of biotin and lipoic acid. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 544–550. Edited by F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.
- van der Ende, A., Hopman, C. T., Zaat, S., Essink, B. B., Berkhout, B. & Dankert, J. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the –10 and –35 regions of the promoter. *J Bacteriol* **177**, 2475–2480.

- van der Ende, A., Hopman, C. T. & Dankert, J. (2000). Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. *Infect Immun* **68**, 6685–6690.
- Erwin, A. L. & Gotschlich, E. C. (1996). Cloning of a *Neisseria meningitidis* gene for L-lactate dehydrogenase (L-LDH): evidence for a second meningococcal L-LDH with different regulation. *J Bacteriol* **178**, 4807–4813.
- Feil, E., Carpenter, G. & Spratt, B. G. (1995). Electrophoretic variation in adenylate kinase of *Neisseria meningitidis* is due to inter- and intraspecies recombination. *Proc Natl Acad Sci USA* **92**, 10535–10539.
- Flower, A. M. & McHenry, C. S. (1990). The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc Natl Acad Sci USA* **87**, 3713–3717.
- Haas, R. & Meyer, T. F. (1987). Molecular principles of antigenic variation in *Neisseria gonorrhoeae*. *Antonie Leeuwenhoek* **53**, 431–434.
- van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993). Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **73**, 1187–1196.
- Hammerschmidt, S., Muller, A., Sillmann, H. & 7 other authors (1996). Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol Microbiol* **20**, 1211–1220.
- Hancock, J. M., Shaw, P. J., Bonneton, F. & Dover, G. A. (1999). High sequence turnover in the regulatory regions of the developmental gene hunchback in insects. *Mol Biol Evol* **16**, 253–265.
- Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C. & Moxon, E. R. (1996). DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc Natl Acad Sci USA* **93**, 11121–11125.
- Imamura, N. & Nakayama, H. (1982). *thiK* and *thiL* loci of *Escherichia coli*. *J Bacteriol* **151**, 708–717.
- Jennings, M. P., Virji, M., Evans, D., Foster, V., Srihanta, Y. N., Steeghs, L., van der Ley, P. & Moxon, E. R. (1998). Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis*. *Mol Microbiol* **29**, 975–984.
- Jennings, M. P., Srihanta, Y. N., Moxon, E. R., Kramer, M., Poolman, J. T., Kuipers, B. & van der Lay, P. (1999). The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. *Microbiology* **145**, 3013–3021.
- Jonsson, A. B., Nyberg, G. & Normark, S. (1991). Phase variation of gonococcal pili by frameshift mutation in *pilC*, a novel gene for pilus assembly. *EMBO J* **10**, 477–488.
- Josenshans, C., Eaton, K. A., Thevenot, T. & Suerbaum, S. (2000). Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fljP*, a gene encoding a basal body protein. *Infect Immun* **68**, 4598–4603.
- Kizil, G., Todd, I., Atta, M., Borriello, S. P., Ait-Tahar, K. & Ala'Aldeen, D. A. (1999). Identification and characterization of TspA, a major CD4(+) T-cell- and B-cell-stimulating *Neisseria*-specific antigen. *Infect Immun* **67**, 3533–3541.
- Malorny, B., Morelli, G., Kusecek, B., Kolberg, J. & Achtman, M. (1998). Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins. *J Bacteriol* **180**, 1323–1330.
- Meyer, T. F. & van Putten, J. P. (1989). Genetic mechanisms and biological implications of phase variation in pathogenic neisseriae. *Clin Microbiol Rev Suppl* **2**, S139–S145.
- O'Hara, B. P., Norman, R. A., Wan, P. T., Roe, S. M., Barrett, T. E., Drew, R. E. & Pearl, L. H. (1999). Crystal structure and induction mechanism of AmiC–AmiR: a ligand-regulated transcription antitermination complex. *EMBO J* **18**, 5175–5186.
- Parkhill, J., Achtman, M., James, K. D. & 25 other authors. (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506.
- Peak, I. R. A., Jennings, M., Hood, D. W. & Moxon, E. R. (1999). Tetranucleotide repeats identify novel virulence determinant homologues in *Neisseria meningitidis*. *Microb Pathog* **26**, 13–23.
- Peck, B., Ortkamp, M., Diehl, K. D., Hundt, E. & Knapp, B. (1999). Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. *Nucleic Acids Res* **27**, 3325–3333.
- Richardson, A. R. & Stojiljkovic, I. (1999). HmbR, a hemoglobin-binding outer membrane protein of *Neisseria meningitidis*, undergoes phase variation. *J Bacteriol* **181**, 2067–2074.
- Ruiz-Lozano, J. M. & Bonfante, P. (2000). A *Burkholderia* strain living inside the arbuscular mycorrhizal fungus *Gigaspora margarita* possesses the *vacB* gene, which is involved in host cell colonization by bacteria. *Microb Ecol* **39**, 137–144.
- Sarkari, J., Pandit, N., Moxon, E. R. & Achtman, M. (1994). Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. *Mol Microbiol* **13**, 207–217.
- Saunders, N. J. (1999). *Bacterial phase variation associated with repetitive DNA*. PhD thesis, The Open University.
- Saunders, N. J., Peden, J. F., Hood, D. W. & Moxon, E. R. (1998). Simple sequence repeats in the *Helicobacter pylori* genome. *Mol Microbiol* **27**, 1091–1098.
- Saunders, N. J., Hood, D. W. & Moxon, E. R. (1999). Bacterial evolution: bacteria play pass the gene. *Curr Biol* **9**, R180–R183.
- Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R. & Moxon, E. R. (2000). Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol Microbiol* **37**, 207–215.
- Snyder, L. A. S., Saunders, N. J. & Shafer, W. M. (2001). A putatively phase variable gene (*dca*) required for natural competence in *Neisseria gonorrhoeae* but not *Neisseria meningitidis* is located within the division cell wall (*dcw*) gene cluster. *J Bacteriol* **183**, 1233–1241.
- Sparling, P. F., Cannon, J. G. & So, M. (1986). Phase and antigenic variation of pili and outer membrane protein II of *Neisseria gonorrhoeae*. *J Infect Dis* **153**, 196–201.
- Stern, A., Brown, M., Nickel, P. & Meyer, T. F. (1986). Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**, 61–71.
- Stibitz, S., Aaronson, W., Monack, D. & Falkow, S. (1989). Phase variation in *Bordetella pertussis* by frameshift mutation in a gene for a novel two-component system. *Nature* **338**, 266–269.
- Tettelin, H., Saunders, N. J., Heidelberg, J. & 39 other authors (2000). Complete genome sequence of *Neisseria meningitidis* serotype B strain MC58. *Science* **287**, 1809–1815.
- Tinsley, C. R. & Gotschlich, E. C. (1995). Cloning and characterization of the meningococcal polyphosphate kinase gene: production of polyphosphate synthesis mutants. *Infect Immun* **63**, 1624–1630.

- Tobe, T., Sasakawa, C., Okada, N., Honma, Y. & Yoshikawa, M. (1992).** *vacB*, a novel chromosomal gene required for expression of virulence genes on the large plasmid of *Shigella flexneri*. *J Bacteriol* **174**, 6359–6367.
- Tomb, J. F., White, O., Kerlavage, A. R. & 39 other authors (1997).** The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.
- Tsuhitashi, Z. & Kornberg, A. (1990).** Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A* **87**, 2516–2520.
- Wang, G., Rasko, D. A., Sherburne, R. & Taylor, D. E. (1999).** Molecular genetic basis for the variable expression of Lewis Y antigen in *Helicobacter pylori*: analysis of the $\alpha(1,2)$ fucosyl-transferase gene. *Mol Microbiol* **31**, 1265–1274.
- Wilson, S. & Drew, R. (1991).** Cloning and DNA sequence of *amiC*, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the *amiC* product. *J Bacteriol* **173**, 4914–4921.
- Yamaoka, Y., Kwon, D. H. & Graham, D. Y. (2000).** A M(r) 34,000 proinflammatory outer membrane protein (*oipA*) of *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **97**, 7533–7538.
- Yang, Q. L. & Gotschlich, E. C. (1996).** Variation of gonococcal lipooligosaccharide structure is due to alterations in poly-G tracts in *lgt* genes encoding glycosyl transferases. *J Exp Med* **183**, 323–327.
- Yurieva, O., Skangalis, M., Kuriyan, J. & O'Donnell, M. (1997).** *Thermus thermophilis dnaX* homolog encoding gamma- and tau-like proteins of the chromosomal replicase. *J Biol Chem* **272**, 27131–27139.
-

Received 17 April 2001; revised 21 May 2001; accepted 5 June 2001.