Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes

Julie C. Dunning Hotopp,¹ Renata Grifantini,² Nikhil Kumar,¹ Yih Ling Tzeng,³ Derrick Fouts,¹ Elisabetta Frigimelica,² Monia Draghi,^{2†} Marzia Monica Giuliani,² Rino Rappuoli,² David S. Stephens,³ Guido Grandi² and Hervé Tettelin¹

Correspondence

Julie C. Dunning Hotopp
jdunning@tigr.org

¹The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

²Novartis Vaccines and Diagnostics Ltd, Via Fiorentina 1, 53100 Siena, Italy

³Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia 30322 and Research Service, VA Medical Center, Decatur, GA 30033, USA

To better understand *Neisseria meningitidis* genomes and virulence, microarray comparative genome hybridization (mCGH) data were collected from one *Neisseria cinerea*, two *Neisseria lactamica*, two *Neisseria gonorrhoeae* and 48 *Neisseria meningitidis* isolates. For *N. meningitidis*, these isolates are from diverse clonal complexes, invasive and carriage strains, and all major serogroups. The microarray platform represented *N. meningitidis* strains MC58, Z2491 and FAM18, and *N. gonorrhoeae* FA1090. By comparing hybridization data to genome sequences, the core *N. meningitidis* genome and insertions/deletions (e.g. capsule locus, type I secretion system) related to pathogenicity were identified, including further characterization of the capsule locus, bioinformatics analysis of a type I secretion system, and identification of some metabolic pathways associated with intracellular survival in pathogens. Hybridization data clustered meningococcal isolates from similar clonal complexes that were distinguished by the differential presence of six distinct islands of horizontal transfer. Several of these islands contained prophage or other mobile elements, including a novel prophage and a transposon carrying portions of a type I secretion system. Acquisition of some genetic islands appears to have occurred in multiple lineages, including transfer between *N. lactamica* and *N. meningitidis*. However, island acquisition occurs infrequently, such that the genomic-level relationship is not obscured within clonal complexes. The *N. meningitidis* genome is characterized by the horizontal acquisition of multiple genetic islands; the study of these islands reveals important sets of genes varying between isolates and likely to be related to pathogenicity.

Received 28 June 2006
Revised 17 August 2006
Accepted 18 August 2006

INTRODUCTION

Neisseriaceae are Gram-negative β -proteobacteria predominantly found on mucosal surfaces in warm-blooded animals

[†]Present address: Departments of Structural Biology and Microbiology and Immunology, Stanford University, School of Medicine, Stanford, CA 94305, USA.

The ArrayExpress accession numbers for the array data related to this paper are A-TIGR-22 and E-TIGR-129.

A supplementary figure and two supplementary tables are available with the online version of this paper.

Abbreviations: IHT, island of horizontal transfer; MLEE, multi-locus enzyme electrophoresis; MLST, multi-locus sequence typing; mCGH, microarray comparative genome hybridization; RMS, restriction/modification system.

(Bovre, 1984; Janda & Knapp, 2003). Most are commensal organisms, including *Neisseria lactamica*, *Neisseria polysaccharea* and *Neisseria cinerea* (Janda & Knapp, 2003). The two best-studied *Neisseria* species are the human pathogens: *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *N. gonorrhoeae* is the causative agent of gonorrhoea; it colonizes the genitourinary tract, invading the epithelium and causing a localized inflammatory process.

N. meningitidis is a causative agent of epidemic bacterial meningitis but is also a commensal isolated from 8–20 % of healthy individuals (Janda & Knapp, 2003). Upon acquisition, *N. meningitidis* may colonize the nasopharynx and can cross the epithelial barrier to enter the bloodstream. In the bloodstream, it can replicate causing septicaemia and/or cross the blood–brain barrier to cause meningitis.

There are five major pathogenic serogroups of *N. meningitidis* (A, B, C, W135 and Y) based on different capsular polysaccharide structures (Janda & Knapp, 2003). However, these pathogenic serogroups arise from a limited number of genetically defined clonal complexes that emerge and spread globally (Stephens, 1999). For example, W135 strains were known to be pathogenic but were not usually responsible for widespread outbreaks (Aguilera *et al.*, 2002). Recently, novel W135 strains in the ST-11/ET-37 clonal complex have emerged and were responsible for worldwide meningitis outbreaks in pilgrims returning from the 2000 and 2001 Hajj pilgrimages and for regional outbreaks in 2002 and 2003 in Burkina Faso (Aguilera *et al.*, 2002).

N. meningitidis strains MC58 and Z2491 have published genome sequences (Parkhill *et al.*, 2000; Tettelin *et al.*, 2000). *N. gonorrhoeae* FA1090 (GenBank AE004969) and *N. meningitidis* FAM18 (<http://www.sanger.ac.uk/Projects/>) genome sequences are also available. Other genome-based techniques such as comparative genome hybridization (mCGH) and subtractive hybridization have addressed issues of (1) cross-species comparisons, (2) particular islands of horizontal transfer, (3) phylogeny, (4) particular regions of biological interest, and/or (5) total genomic content (Bille *et al.*, 2005; Perrin *et al.*, 1999, 2002; Snyder *et al.*, 2004, 2005; Snyder & Saunders, 2006; Stabler & Hinds, 2006; Stabler *et al.*, 2005). We sought to explore the variable gene content of *N. meningitidis* in an effort to understand the prevalence and importance of horizontal gene transfer within this important, naturally competent organism.

N. meningitidis strains can be differentiated by multi-locus enzyme electrophoresis (MLEE) patterns and multi-locus sequence typing (MLST) (Achtman, 1995; Maiden *et al.*, 1998). We used mCGH to examine many diverse meningococcal strains, comparing genomes and assessing large insertion/deletion events. We found that meningococci can be placed into groups based on their mCGH profiles that significantly overlap with clonal complexes. Other significant insights into pathogenicity genes, invasive strains and the emerging W135 epidemic strains are discussed.

METHODS

Bacterial strains. The strains used are listed in Table 1. The spatial and temporal distribution, the extensive phenotypic and genetic typing, and the characterization of most of these *N. meningitidis* strains have been previously reported (Grifantini *et al.*, 2002; Maiden *et al.*, 1998; Parkhill *et al.*, 2000; Pizza *et al.*, 2000; Seiler *et al.*, 1996; Tettelin *et al.*, 2000; Wang *et al.*, 1993). The strains represent the population of *N. meningitidis*, making them ideal candidates for drawing biological conclusions from a whole-genome study.

Microarray construction. Primers were designed from the unique predicted genes annotated in MC58, as previously described (Grifantini *et al.*, 2002). Subsequently, primers were designed to unique regions of Z2491, FAM18 and FA1090. Amplicons of an average size of 495 bp were produced using AmpliTaq (Applied Biosystems), purified using filtration plates (Millipore), and analysed on agarose gels. The purified amplicons were diluted 1:1 with DMSO, spotted in triplicate onto Corning UltraGap slides with a

Lucidea printing robot (GE Healthcare), and irradiated with ultraviolet light. The microarray slide type has been deposited in ArrayExpress (A-TIGR-22).

Hybridizations. Cy3 and Cy5 probes were synthesized from genomic DNAs as previously described (Tettelin *et al.*, 2001). Briefly, amino-allyl-dUTP-labelled probes were synthesized from 4 µg genomic DNA using Klenow Fragment (3'→5' exo-) (New England Biolabs). The reactions were purified using the QIAquick PCR purification kit (Qiagen) with modified phosphate buffers. Cy3 or Cy5 dyes (GE Healthcare) were chemically coupled to the incorporated amino-allyl-dUTP in carbonate buffer. Cy3- and/or Cy5-labelled probes were synthesized at least twice from each genomic preparation. Probe pairs were resuspended in 5% SSC, 50% formamide and 0.1% SDS, and hybridized to slides overnight at 42 °C. The hybridized slides were washed in 2× SSC, 0.1% SDS at 55 °C; 0.1× SSC, 0.1% SDS at room temperature; 0.1× SSC at room temperature; and MilliQ water at room temperature. They were then dried, and scanned using a GenePix4000B scanner (Molecular Devices). The corresponding images were analysed using TIGR Spotfinder (Saeed *et al.*, 2003). The microarray study has been deposited in ArrayExpress (E-TIGR-129).

Data analysis and bioinformatics. Ratios (Cy5/Cy3) were normalized using iterative log-mode centring, whereby the mode of the log₂(Ratio) histogram with bin size 0.1 was centred at zero (Lindroos *et al.*, 2005; Read *et al.*, 2003). The mean normalized log₂(Ratio) and standard deviation was then calculated from all replicates (at least two slides each with three spotted replicates) with good Spotfinder flags (B or C). The program GACK (Kim *et al.*, 2002), which uses dynamic thresholds and has been used in other *Neisseria* mCGH studies (Snyder *et al.*, 2004; Stabler *et al.*, 2005), was explored as an alternative to normalization. However, between closely related species (e.g. H44/76) or in self/self hybridizations a sharp histogram resulted which led to GACK erroneously picking genes as divergent. Therefore GACK was not further explored.

Prediction of absent/present genes. Since the microarray contains four organisms and MC58 (reference) DNA will not hybridize to Z2491-, FA1090- and FAM18-specific amplicons, nucleotide similarity results between spotted amplicons and genome sequences were used in combination with the observed ratios to predict absence/presence when presenting microarray data in the context of the four genomes (e.g. in Fig. 7). An amplicon from the other three strains was considered present in the MC58 genome if it was >70% identical to MC58 over >90% of the length of the amplicon. When an amplicon was present in MC58, a ratio >3 predicted absence or significant divergence, a ratio <2 predicted presence, and a ratio between 2 and 3 was determined to not be predictive. In correspondence, when an amplicon was absent from MC58, a ratio >0.5 predicted absence, a ratio <0.33 predicted presence, and a ratio between 0.5 and 0.33 was determined to not be predictive. When fluorescence was undetected in either channel, no prediction was made. The reliability of this correspondence analysis is evident on the FA1090 circle in Fig. 7, where the rim is nearly empty compared to the *N. meningitidis* rims. Further support for this analysis is evident in the results and subsequent PCR verification presented in Fig. 3.

Clustering analysis. The mCGH data were clustered using various algorithms in the TIGR MeV software available in the TM4 software package (Saeed *et al.*, 2003). Support trees were constructed using a Euclidean distance and average linkage. Experiment trees were resampled by bootstrapping genes.

K-means clustering was carried out using Euclidean distance and calculated means. Fifteen clusters were selected and the solution converged after 22 iterations. The number of clusters selected was

Table 1. Strain information

NA, not applicable; ND, not described; NG, not a groupable serogroup; CDC, from the CDC collection with no associated publication.

Inner rim*	Strain	Serogroup	Country	Year	Disease	MLST	MLEE	Reference
<i>N. cinerea</i>								
	CIN	NA	ND	ND	NA	ND	ND	
<i>N. lactamica</i> Pink								
1	NL19	NA	ND	ND	NA	ND	ND	
2	NL17	NA	ND	ND	NA	ND	ND	
<i>N. gonorrhoeae</i> Yellow								
3	F62	NA	ND	ND	Gonorrhoea	ND	ND	Pizza <i>et al.</i> (2000)
4	FA1090	NA	ND	ND	Gonorrhoea	ND	ND	Dempsey <i>et al.</i> (1991)
<i>N. meningitidis</i> mCGH-1 Orange								
5	GA1272	C	USA	1991	Invasive	ND	Subgroup IV	Raymond <i>et al.</i> (1997)
6	M4950	NG	USA	1998	Carrier	ND	ET1202	Dolan-Livengood <i>et al.</i> (2003)
7	M6049	W135	USA	1999	Invasive	1062	ST23 complex/ Cluster A3	CDC
8	M5016	NG (Y)	USA	1998	Carrier	ND	ST-23 complex, ET1216	Dolan-Livengood <i>et al.</i> (2003)
9	GA3470	Y	USA	1995	Invasive	ND	ET508	Kahler <i>et al.</i> (2001)
10	GA0929	Y	USA	1991	Invasive	ND	ET501	Swartley <i>et al.</i> (1997)
11	M4060 (68961)	Y	USA	1998	Carrier	ND	ET508	Kahler <i>et al.</i> (2001)
12	GA1002	W135	USA	1991	Invasive	ND	ND	Swartley <i>et al.</i> (1997)
13	M1762	Y	ND	ND	Invasive	ND	ET509	Kahler <i>et al.</i> (2001)
14	GA1655	C	USA	1992	Invasive	ND	Subgroup 6	Raymond <i>et al.</i> (1997)
15	M4041	W135	Canada	1997	Invasive	1482	ST-22	CDC
16	GA18736	W135	USA	2002	Invasive	ND	ND	CDC
17	6083	W135	USA	1980	Invasive	ND	ET916	McAllister & Stephens (1993)
<i>N. meningitidis</i> mCGH-2 Green								
18	Z2491	A	Gambia	1983	Invasive	4	ST-4/subgroup IV	Maiden <i>et al.</i> (1998)
19	F6124	A	Chad	1988	Invasive	5	ST-5/subgroup III	Pizza <i>et al.</i> (2000)
20	F8229 (CDC1750)	A	Kenya	1989	Invasive	ND	Subgroup III	Pinner <i>et al.</i> (1992)
21	Z4096	A	China	1963	Carrier	ND	Subgroup VIII	Wang <i>et al.</i> (1992)
22	Z5035 (5035, Z4081)	A	China	ND	Invasive	ND	Subgroup V	Wang <i>et al.</i> (1992)
23	Z5005 (5005, B40)	A	Morocco	1967	Invasive	ND	Subgroup I	Wang <i>et al.</i> (1992)
24	BZ 133	C	Netherlands	1977	Invasive	1	ST-1/subgroup I/II	Seiler <i>et al.</i> (1996)
<i>N. meningitidis</i> mCGH-3 Blue								
25	M5020	NG	USA	1998	Carrier		ST-198 complex/ classIII	Dolan-Livengood <i>et al.</i> (2003)
26	C11	C	Cuba	1983	Invasive	345	other	Comanducci <i>et al.</i> (2002)
27	BZ 232	B	Netherlands	1964	Invasive	38	NA	Seiler <i>et al.</i> (1996)
28	NG H38	B	Norway	1988	Carrier	36	NA	Seiler <i>et al.</i> (1996)
29	2996	B	UK	1975	Invasive	540	ST-8/Cluster A4	van der Ley & Poolman (1992)
30	961-5945	B	Australia	1996	Invasive	153	ST-8/Cluster A4	Comanducci <i>et al.</i> (2002)
31	FAM18	C	USA	1983	Invasive	ND	ET-37	Comanducci <i>et al.</i> (2002)
32	GA1442	Y	USA	1992	Invasive	ND	ET515 serosub: P1-5,2	Kahler <i>et al.</i> (2001)
33	M7124	W135	Saudi Arabia	2000	Invasive	11	ST-11/ET-37	CDC
34	M7257	W135	USA	2000	Invasive	11	ST-11/ET-37	CDC
35	NG P165	B	Norway	1974	ND	ND	ST-11/ET-37	Seiler <i>et al.</i> (1996)
<i>N. meningitidis</i> mCGH-4 Violet								
36	1000	B	Russia	1988	Invasive	20	ST-18	Seiler <i>et al.</i> (1996)
37	528	B	Russia	1989	Invasive	18	ST-18	Seiler <i>et al.</i> (1996)
38	NG E28	B	Norway	1988	Carrier	26	NA	Seiler <i>et al.</i> (1996)
39	NG E31	B	Norway	1988	Carrier	15	ST-364	Seiler <i>et al.</i> (1996)
40	297-0	B	Chile	1987	Carrier	49	ST-254	Seiler <i>et al.</i> (1996)

Table 1. cont.

Inner rim*	Strain	Serogroup	Country	Year	Disease	MLST	MLEE	Reference
41	BZ 198	B	Netherlands	1986	Invasive	41	ST-44/Lineage 3	Seiler <i>et al.</i> (1996)
42	NZ98/254 (NZ394/98)	B	New Zealand	1998	Invasive	42	ST-44/Lineage 3	Comanducci <i>et al.</i> (2002); Oster <i>et al.</i> (2005)
43	BZ 147	B	Netherlands	1963	Invasive	48	ST-44/Lineage 3	Seiler <i>et al.</i> (1996)
44	NG H15	B	Norway	1988	Carrier	43	ST-44/Lineage 3	Seiler <i>et al.</i> (1996)
45	NG H36	B	Norway	1988	Carrier	47	ST-44/Lineage 3	Seiler <i>et al.</i> (1996)
<i>N. meningitidis</i> mCGH-5 Grey								
46	GA139	C	USA	1989	Invasive	ND	ND	Raymond <i>et al.</i> (1997)
47	SWZ107	B	Switzerland	1986	Invasive	74	ST-35	Seiler <i>et al.</i> (1996)
48	NG 6/88	B	Norway	1988	Invasive	13	ET-5	Seiler <i>et al.</i> (1996)
49	NG F26	B	Norway	1988	Carrier	14	ST-269	Seiler <i>et al.</i> (1996)
50	M2436	Y	USA	1996	Invasive	ND	ET 546	Dempsey <i>et al.</i> (1991)
51	H44/76	B	Norway	1976	Invasive	ND	ST-32/ET-5	Seiler <i>et al.</i> (1996)
52/reference	MC58	B	UK	1983	Invasive	74	ST-32/ET-5	McGuinness <i>et al.</i> (1991)

*Rim designations refer to Fig. 7.

determined by running figures of merit on the dataset 10 times with a maximum number of clusters of 50 and with 100 iterations. Consistently, 20 seemed to be an appropriate number of clusters.

Confirmation of variable regions by amplification. Amplicons were generated using High Fidelity Taq (Invitrogen) according to the manufacturer's suggestions with 1·0 µM of each primer. Reactions were initiated with a 2 min incubation at 95 °C followed by 35 cycles of 95 °C for 30 s, 55 °C for 30 s, 68 °C for 10 min and with a final elongation at 72 °C for 10 min. Primers were designed using the MC58 genome: NMB_0895F, ATTTTAATTACGAAGGCTACGCATT; NMB_0901R, GGGACACCCGCGAAGTTTTGGAAGC; NMB_0910F, CTGTCAAGTTGTCTCGTGCATTGTCA; NMB_0912R, GTTGCGGG-CTGTTGCGTCCGAAACC; NMB_0917F, ATGGATAAGCGCGACC-AGTTCGCCG; NMB_0919F, GATGTGTTTGGCAATCATGGCTTG; NMB_0920R, CACAAGTGATGCGTCCGAGCGTAA.

RESULTS

Microarray specificity

To delineate absent and divergent genes from present genes, two hybridization ratio thresholds were defined. An excessively high ratio threshold fails to identify all variation present; an excessively low ratio threshold fails to identify all conservation present. Numerous ratio thresholds and types of thresholds have been reported previously; they vary depending on the organism and microarray. Two ratio thresholds were used in this study that minimized the noise-to-data ratio.

After comparing ratios to nucleotide percentage identity for sequenced *Neisseria* species, 70 % nucleotide identity was chosen to delineate presence from absence/divergence (Fig. 1, Fig. 2). A ratio greater than 3 predicted absence or divergence (< 70 % sequence identity); a ratio less than 2 predicted conservation (> 70 % sequence identity); and a ratio between 2 and 3 was not predictive. PCR amplification of one variable region from selected strains confirmed the significance of these predictions (Fig. 3).

These thresholds minimized the false positives, since incorrectly calling a locus conserved is preferable to incorrectly calling a locus absent. Indeed, gene absence underlies most of the critical conclusions in this study. Additionally, the false negative results were nonrandom and occurred in amplicons spotted on the microarray that gave lower than average fluorescence intensity. When hybridization of the reference DNA yielded low fluorescence for an amplicon, a large ratio was less likely to be obtained (Fig. 4). Indeed, the highest ratios correlated with more fluorescent amplicons [those with higher log₁₀(Product) values].

Capsule region

The hybridization results at the meningococcal capsule locus for the strains examined were consistent with their known serogroup assignments, further verifying the significance of the mCGH predictions (Fig. 5; Supplementary Fig. S1). For example, all the serogroup A strains had ratios consistent with presence of *sacABCD* and absence of *synABCD* (Fig. S1). Likewise, all the serogroup C strains had a ratio consistent with presence of *synE* (NMC_0050). However, the predicted serogroup C capsule *O*-acetyltransferase gene (*oatC*; NMC_0049) (Claus *et al.*, 2004) was variable across the serogroup C strains (Fig. S1). In addition, strain 528, which has previously been shown to be serogroup B, now contains unique deletions of the capsule region possibly due to routine passaging *in vitro* (Fig. 5).

Three previously characterized non-groupable strains were examined. Strain M5016 is known to be non-groupable due to a point mutation in a Y-like capsule region (Dolan-Livengood *et al.*, 2003); the array results were consistent with the presence of a Y-like capsule, but the point mutation could not be assayed with this method. Strain M4950 is known to have lost the *synABC* genes (Dolan-Livengood *et al.*, 2003), which was also detected by mCGH. Lastly,

Fig. 1. Relationship of mCGH ratios and nucleotide percentage identity ordered along MC58. The nucleotide percentage identity of FAM18, Z2491 and FA1090 amplicons when aligned to MC58 (grey) and the $\log_2(\text{Ratio})$ values (black) are plotted and ordered based on the location along the MC58 chromosome. Delineated on each graph is the range of nucleotide percentage identity considered conserved ($>70\%$ identity) and the ranges of ratios considered conserved (<2) and not predictive ($2-3$). A clear correlation exists between percentage identity and mCGH ratio (see also Fig. 2).

M5020 has been shown to be missing the entire capsule locus between *galE* and *tex*, similar to *N. gonorrhoeae* (Dolan-Livengood *et al.*, 2003). The microarray indicated extensive loss of this region, with the greatest similarity in configuration to *N. lactamica* NL19, which may suggest horizontal transfer of this region from *N. lactamica*.

Other novel variations are seen across the capsule region that will require further characterization: (a) serogroup B gene NMB_0065 displayed a variable distribution in diverse strains across all the serogroups and (b) groups of strains were identified that have absence/divergence of *lipAB* (Fig. S1). The *lipAB* genes are predicted to encode phospholipid lipidation genes that provide the diacylglycerol anchor for the polymer in the outer membrane; but they may encode capsule chaperones, as deletion of *lipAB* leads to intracellular capsule accumulation of lipidated capsule polymers (Tzeng *et al.*, 2005).

Hierarchical clustering and islands of horizontal transfer

A support tree (Graur & Li, 2000) generated using mCGH ratios grouped meningococcal isolates of similar MLST clusters together, but could not differentiate the specific sequence types (e.g. ST-32/ET-5 complex and ST-269) (Fig. 6; Supplementary Table S1). However, various groups of strains were identified that had very similar profiles and were defined as mCGH groups (Table 1). This clustering should not be mistaken as a phylogenetic analysis. These strains are closely related and the large islands of horizontal transfer will lead to stochastic effects in conducting a phylogenetic analysis with these CGH analysis. Such an analysis has been successfully conducted with more divergent *Neisseria* species (Stabler *et al.*, 2005).

Fig. 2. Relationship of mCGH ratios and nucleotide percentage identity. The nucleotide percentage identity of FAM18, Z2491 and FA1090 amplicons when aligned to MC58 was compared to the $\log_2(\text{Ratio})$. Delineated on each graph is the range of nucleotide percentage identity considered conserved ($>70\%$ identity) and the ranges of ratios considered conserved (<2) and not predictive ($2-3$). The majority of the points in this scatter plot reside within the overlap of 70% identity and a ratio <2 , corresponding to the core *Neisseria* genes. The general angle of the cloud of points indicates that ratios increase when the percentage identity decreases, confirming that the microarray results are reliable (see also Fig. 1). These thresholds minimizing the false positives were chosen over thresholds minimizing the false negatives since incorrectly calling a locus conserved seems preferable to incorrectly calling a locus absent. And indeed, the latter are the genes underlying most of the critical conclusions in this study.

In addition, using the ratio-thresholds described above, genes were categorized and plotted according to the corresponding amplicon location in each sequenced genome (Fig. 7). The previously described MC58 islands of horizontal gene transfer (IHT) (Tettelin *et al.*, 2000) and putative prophage regions (Masignani *et al.*, 2001; Parkhill *et al.*, 2000; Tettelin *et al.*, 2000) were found to be variable across the *Neisseria* species and meningococcal strains tested (Table 1). Except for the expected changes in the capsule region according to serogroup, most of IHT-A appeared to be present in all of the *N. meningitidis* strains although not in *N. gonorrhoeae* or *N. lactamica*. In addition to these large IHT elements, smaller variable regions of a single or a few genes exist with atypical nucleotide content (Tettelin *et al.*, 2000). These often corresponded to genes that were absent/divergent in a variety of strains in the mCGH results, lending further evidence to these being

smaller potential IHTs (Fig. 7). While variable regions are often associated with atypical nucleotide content, there were several variable regions that had nucleotide content typical of the meningococcus, suggesting transfer between species of similar nucleotide content (e.g. IHT-D in MC58 Fig. 7).

Taken together, six of these islands (MuMenB/PNM2, PNM1, IHT-B, IHT-C, IHT-D, IHT-E) differentiated *N. meningitidis* into the five groups (Fig. 6, Fig. 7, Table 1) also identified by hierarchical clustering. The two *N. gonorrhoeae* strains were missing all of these islands while the two *N. lactamica* strains had IHT-D. The only other set of strains to contain IHT-D were those in the mCGH-5 group (Table 1), which includes MC58, a result that supports the idea that genes can be transferred between pathogenic and commensal *Neisseria* species (Feil *et al.*, 1996; Linz *et al.*, 2000). In

Fig. 3. IHT-E distinguishes mCGH-3 strains. The 44 kB IHT-E region that distinguishes mCGH-3 strains encodes a likely intact prophage (NMC_0836-NMC_0883; boxed in light yellow) and an adjacent transposon containing a type I secretion system (NMC_0884-NMC_0894; boxed in light blue) (A). In Z2491, the prophage and transposon are completely absent. In MC58, an internal portion of the prophage has been deleted and the transposon is absent. mCGH results demonstrate that the phage and transposons are not genetically linked. Homologous genes are linked vertically with pink bars, genes are colour-coded by role categories according to the key, and tRNAs, transcriptional terminators and predicted *att* sites are illustrated. The mCGH results for 16 selected strains are mapped to the FAM18 genome (red, absent/divergent; blue, not predictive; white, present) and compared to a table of confirmatory PCR results (dashes indicate no amplification or non-specific amplification; two values indicate that two equal-intensity bands were obtained) (B). The mCGH results correlate with expected amplification results; the few PCR results highlighted in light green are unexpected. Unexpected results may be either due to spurious amplification or from genomic alteration not assessed by CGH. The expected amplicon sizes are shown along with their location below the mCGH results. PCR1 spans NMC_0835-NMC_0897; an amplicon is only expected when both the prophage and the transposon are absent. PCR2 spans NMC_0835-NMC_0841; an amplicon is expected if the left flank of the prophage is present. PCR3 spans NMC_0850-NMC_0878; an amplicon is expected if a portion of the prophage has been deleted in a manner similar to MC58. PCR4 spans NMC_0882-NMC_0897; an amplicon is expected if the transposon is absent. PCR 5 spans NMC_0896-NMC_0897; an amplicon is expected if the transposase is present.

Fig. 4. Relationship between the log ratio and log product for a typical hybridization. The $\log_2(\text{Ratio})$ is defined as $\log_2[(\text{Reference Fluorescence})/(\text{Query Fluorescence})]$ and the $\log_{10}(\text{Product})$ is defined as $\log_{10}[(\text{Reference Fluorescence}) \times (\text{Query Fluorescence})]$. The $\log_{10}(\text{Product})$ represents the total fluorescence obtained and relates this to the $\log_2(\text{Ratio})$. At the high $\log_{10}(\text{Product})$ values, no high ratios are found because to obtain a high ratio you must have a large amount of intensity in one channel and low intensity in the other. Therefore, the highest ratios are found in the middle of product range. At the lower $\log_{10}(\text{Product})$ values, no high ratios are found suggesting that at low fluorescence values, not as large a change in ratio is possible. At lower fluorescence intensities a large ratio appears more difficult to obtain. As such, the false negative results were nonrandom and tended to occur in amplicons that gave lower than average fluorescence intensity.

addition to IHT-D, mCGH-5 isolates also contained MuMenB/PNM2 and had variable presence of IHT-B/IHT-C. MuMenB and PNM2 are distinct phage regions in Z2491 and MC58 but have overall high nucleotide similarity (Supplementary Table S2); no attempts were made to differentiate them in this analysis. Examination of individual prophage-specific genes may allow for identification of the presence of MuMenB or PNM2, but that is beyond the scope of this paper. However, careful analysis is required, as some prophage genes have been duplicated in the sequenced *N. meningitidis* genomes; alternative duplications are likely in unsequenced genomes.

The mCGH-4 strains had IHT-B/IHT-C as well as MuMenB/PNM2. The mCGH-2 strains, which contain Z2491 and the well-studied clonal serogroup A strains, had both prophage regions (MuMenB/PNM2 and PNM1) but none of the other IHT elements. Strains from the mCGH-1 group were the most similar of all the *N. meningitidis* strains to *N. gonorrhoeae* and *N. lactamica* as they lacked all six islands except for some variation in the presence of MuMenB/PNM2. However, they were distinctly *N. meningitidis* strains as they possess all or portions of IHT-A (capsule region) and were lacking the *N. gonorrhoeae* islands. The mCGH-1 strains are likely to have their own

IHT element(s) that could not be interrogated by mCGH since there are no sequenced meningococci from this group.

Lastly, the mCGH-3 group was differentiated by the absence of all islands except IHT-E. In FAM18 and the other mCGH-3 strains, the 44 kB IHT-E region encoded an intact prophage (NMC_0836-NMC_0883) and an adjacent transposon (NMC_0884-NMC_0894) (Fig. 3). In *N. gonorrhoeae*, *N. lactamica* and *N. meningitidis* mCGH-1 and mCGH-2 strains, the prophage and transposon were completely absent. In mCGH-4 and mCGH-5 strains, the two plasmid/prophage addiction modules were retained (for a review see Engelberg-Kulka & Glaser, 1999), but an internal portion of the prophage, including the head and tail morphogenesis genes, was replaced with an IS30 transposase homologue (NMB_0911). The loss of this portion may have occurred when a second transposase inserted into the prophage, followed by homologous recombination excising the internal region. The right flank of the excision appeared to be maintained in all mCGH-4 and mCGH-5 strains. In strains 1000, 528, NG E28, NG E311 and M2436, additional genes from this region appeared to be absent or divergent (Fig. 3).

The adjacent transposon appeared to be a separate mobile element, as it has variable presence in all mCGH groups. This transposon contained genes that likely encode the MFP (NMC_0887) and ABC (NMC_0888) components of a type I secretion system.

Genes linked to IHTs

K-means clustering (Soukas *et al.*, 2000) was employed to further elucidate the relationship between gene complement and phenotype. Most of the clusters contained *N. lactamica*- or *N. gonorrhoeae*-specific genes or the IHT elements previously discussed. However, two interesting clusters comprised genes clustering with the IHTs. For mCGH-3 strains, 13 smaller islands were unique to these isolates in addition to IHT-E (Table 2). These islands were scattered throughout the FAM18 genome and include restriction/modification systems, a fimbrial protein precursor, a putative TonB-dependent receptor, a transcription regulator, a putative phage protein, a transferrin-binding protein precursor, and numerous uncharacterized hypothetical, conserved hypothetical, and outer-membrane proteins. Seven small islands were unique to mCGH-2 strains besides the previously described IHTs (Table 3) including capsule genes, putative phage genes, a putative DNA-binding protein and numerous hypothetical proteins. The conservation of these smaller islands in these groups of strains suggests that either the strains within these mCGH groups evolved from a single common ancestor, or the islands were acquired at one time and have been moved around the genomes that have been sequenced.

Core genome

The core *N. meningitidis* genome consisted of genes present and highly conserved in nucleotide sequence across a wide

Fig. 5. Capsule configuration. The capsule region is shown for each of the four sequenced *Neisseria* species (A). *N. gonorrhoeae* FA1090 does not synthesize capsule and as such does not have the capsule biosynthetic or transport genes. *N. meningitidis* Z2491 is a serogroup A strain and contains *sacABCD* for synthesis capsule composed of (α 1 $\rightarrow$ 6)-linked *N*-acetylmannosamine-1-phosphate. MC58 and FAM18 both synthesize polysialic acid capsules and have *synABC* and serogroup-specific genes. Homologous genes are linked vertically with pink bars, genes are colour-coded by role categories according to the key, and transcriptional terminators are illustrated. Representative mCGH results demonstrate different configurations in the capsule region (B). The results are mapped to the MC58 genome (red, absent/divergent; blue, not predictive; white; present) and unique regions of the other capsule regions. The bottom line indicates colour codes from (A). The *ermC* cassette illustrated in MC58 is the result of insertional mutagenesis of the capsule region.

Fig. 6. Support tree constructed from the mCGH data. The mCGH support tree resulted in grouping isolates with similar MLST profiles (Table 1). Although similar to MLST clusters, this grouping could not differentiate the specific sequence types. Six islands are predominantly responsible for differentiating the meningococcal isolates into five different mCGH groups (see also Fig. 7).

diversity of strains (ratio <3 in all 49 *N. meningitidis* strains) (Table S2). This core gene list included 1706 genes [79 % of MC58 genes similar to the previously reported 78 % (Perrin *et al.*, 2002)] and was consistent with a map of metabolic pathways constructed from the genome sequencing data of MC58 (Fig. 8). Of these core genes, 624 (36.6 %) were hypothetical proteins with no known function.

N. meningitidis- and pathogen-specific genes

By comparing the subsets of genes absent in *N. lactamica* and *N. gonorrhoeae*, potential *N. meningitidis*-specific genes and pathogen-specific genes were identified. This list should not be considered an exhaustive one, considering the small number of non-meningococcal isolates examined. A total of

122 potential *N. meningitidis*-specific genes were identified based on their mCGH ratios, indicating that they were absent/divergent only in *N. lactamica* NL17, *N. lactamica* NL19, *N. gonorrhoeae* FA1090 and *N. gonorrhoeae* F62 (Table S2). More than half of these *N. meningitidis*-specific genes (64 genes) are hypothetical proteins. The remaining functionally annotated *N. meningitidis*-specific genes include genes for capsule biosynthesis, secretion proteins, haemolysins, transcriptional regulators and iron uptake, as well as numerous genes in a previously identified prophage region (Masignani *et al.*, 2001).

Of the *N. meningitidis*-specific genes, 18 were found in all the *N. meningitidis* strains and might be considered core *N. meningitidis*-specific genes (Table 4). Half of these genes were hypothetical proteins. The remaining core *N. meningitidis*-specific genes include three FrpA/C proteins, two putative secretion proteins, a putative transporter, a putative TonB-dependent receptor, a Cu-Zn superoxide dismutase, and a thiol : disulfide interchange protein, DsbA. Although these core *N. meningitidis* genes are important and include some known pathogenicity factors, the selection of *N. meningitidis* strains includes numerous strains not known to cause disease. Therefore, not all *N. meningitidis*-specific pathogenicity factors are likely to be identified.

A total of 60 potential core pathogen-specific genes were absent/divergent in the commensal organisms, *N. lactamica* NL17 and *N. lactamica* NL19 (Table S2) but found in all the pathogenic strains tested. These potential core pathogen-specific genes include known pathogenicity factors and genes identified as differentially regulated upon contact with host cells (Table S2).

Some genes and gene clusters from metabolic pathways were identified as being absent/divergent in commensal *Neisseria* species. These include a gluconate permease and gluconokinase; a D-amino acid dehydrogenase and putative sodium/alanine symporter; a putative 2-methylcitrate pathway; and an arginine decarboxylase, agmatinase and C_4 -dicarboxylate transporter. These include four distinct genomic loci (NMB_2027-NMB_2028, NMB_0176-NMB_0177, NMB_0430-NMB_0431 and NMB_0433, NMB_0468-0470, respectively).

Invasive and carriage isolates

The strains were selected to include invasive and carriage isolates. No genes were identified by either *K*-means clustering or hierarchical clustering that were over-represented in either hypervirulent isolates (from ST-8/Cluster A4, ST-32/ET-5 and ST-44/Lineage 3) or invasive isolates.

DISCUSSION

MLST and mCGH

Meningococcal isolates of similar MLST profiles (Maiden *et al.*, 1998) could be grouped together using hierarchical

Fig. 7. Circular representation of mCGH results. The outer rims present the predicted coding regions on the plus and minus strands (colour-coded by role categories; rim 1 and 2, respectively); predicted genes not represented on the array (grey, rim 3), and plot of the atypical nucleotide content (rim 4). The inner rims present the comparative genome hybridizations results for the strains listed by rim in Table 1. The ratios $[(MC58 \text{ normalized intensity}) / (\text{query normalized intensity})]$ were divided into three categories: ratio < 2 , most likely present in the test strain (not shown); ratio > 3 , mostly likely absent or divergent in the test strain (red); ratio 2–3, not predicted (blue); regions where no prediction can be made (grey). The previously described MC58 islands of horizontal gene transfer (IHT-B, IHT-C) and the putative prophage are found to be variable across the species/strains tested. Taken together, six of these islands (MuMenB/PNM2, PNM1, IHT-B, IHT-C, IHT-D and IHT-E) differentiate the meningococcal isolates into five different groups also identified by hierarchical clustering (see Fig. 6), which are colour coded (Table 1).

Table 2. Genes clustering with IHT-E in mCGH-3 strains

Locus	Common name*
NMC_0209	Conserved hypothetical protein
NMC_0210	Fimbrial protein p9-2 precursor
NMC_0295	Adenine/cytosine DNA methyltransferase
NMC_0296	Conserved hypothetical protein
NMC_0328	Modification methylase
NMC_0387	TonB-dependent receptor, putative
NMC_0678	HpaII very-short-patch-repair endonuclease
NMC_0679	Modification methylase
NMC_0680	Putative type II restriction enzyme
NMC_1117	Conserved hypothetical protein
NMC_1118	Transcription regulator
NMC_1225	AbiD phage protein homologue lin2373, putative
NMC_1519	Hypothetical protein
NMC_1520	Modification methylase
NMC_1691	Transferrin-binding protein 2 precursor
NMC_1730	Protein of unknown function (DUF819) superfamily
NMC_1731	Conserved hypothetical protein
NMC_1810	Hypothetical protein
NMC_2020	Outer-membrane protein class 2

*Common names used are from local/TIGR automated annotation.

clustering, but the specific sequence types (e.g. ST-32/ET-5 complex and ST-269) could not be resolved. The similarity of mCGH and MLST results is likely due to both methods examining sequences throughout the genome, as opposed to

Table 3. Genes clustering with PNM1 in the mCGH-2 strains

Locus	Common name
NMA_0001	Hypothetical protein
NMA_0131	Hypothetical protein
NMA_0132	Hypothetical protein
NMA_0196	Capsule polysaccharide export inner-membrane protein
NMA_0199	Putative UDP-N-acetyl-D-glucosamine 2-epimerase
NMA_0200	Putative capsule biosynthesis protein
NMA_0201	Putative capsule biosynthesis protein
NMA_0202	Putative capsule biosynthesis protein
NMA_0772	Putative DNA-invertase
NMA_0774	Hypothetical protein
NMA_0775	Possible integral membrane protein
NMA_0778	Hypothetical protein
NMA_0781	Hypothetical protein
NMA_0782	Putative phage replication protein
NMA_1210	Hypothetical protein
NMA_1289	Hypothetical protein
NMA_1309	Hypothetical protein
NMA_1312	Putative DNA-binding protein
NMA_1544	Hypothetical protein
NMA_2230	Hypothetical protein

a single sequence. This similarity suggests that although naturally competent, *Neisseria* species inherit/lose foreign genes at a low enough frequency that overall genome content can be predicted based on MLST, and it further validates the use of MLST for meningococcal population studies.

As an example, W135 strains have two distinctly different placements on the hierarchical support tree. Two W135 strains isolated from Hajj pilgrims were located with another ST-11/ET-37 isolate (NG P165) and with isolates in the ST-8/Cluster A4. Two W135 strains isolated from sporadic cases in North America prior to the W135 outbreaks of the Hajj pilgrimage were found clustered in mCGH-1 next to the serogroup A strains of mCGH-2. The close relationship of the mCGH-3 W135 Hajj strains to the other ST-11/ET-37 strains is suggestive that the ancestor to these strains was an ST-11/ET-37 isolate that acquired the W135 capsule locus. This may not have been a recent event, but the organism may have had a recent selective advantage and emerged possibly from the usage of the serogroup A and C meningococcal polysaccharide vaccine in Hajj pilgrims prior to 2002. A similar vaccine-induced pressure has been observed in *Streptococcus pneumoniae*, with an increase in serotype 19A-associated disease after use of a vaccine that does not include this serotype (Kyaw *et al.*, 2006).

Pathogenesis and horizontal gene transfer

The *N. meningitidis* genome is characterized by the horizontal acquisition of multiple genetic islands that contain pathogenicity factors as well as numerous hypothetical proteins. These genetic islands appear to be acquired both from within *N. meningitidis* strains as well as from *N. gonorrhoeae* and/or *N. lactamica*. Examples of the latter may include (a) replacement of the *N. meningitidis* capsule locus in strain M5020 with that of *N. gonorrhoeae* or *N. lactamica* and (b) transfer of IHT-D between *N. lactamica* and mCGH-5 *N. meningitidis* strains. MuMenB, PNM2, PNM1 and IHT-E seem to all be intact prophages. IHT-D and IHT-C encode proteins with homology to prophage elements. Likewise, IHT-B encodes proteins with homology to plasmid replication genes. This suggests that these may be less characterized mobile elements. Mobile elements are often associated with moving pathogenicity islands between strains and this continues to be the case with *N. meningitidis*.

For example, a genetic island containing a chromosomally integrated bacteriophage was found previously to be associated with 100 % of hypervirulent lineages and absent from 90 % of noninvasive isolates (Bille *et al.*, 2005). In the current study, this island was present in 60 % of the virulent strains and 42 % of the carriage strains examined, suggesting that this phage may be a marker for certain hypervirulent clonal groups (e.g. ST-11), but is not a characteristic of all invasive meningococcal strains.

Islands may also be transferred to/from other respiratory colonizers. The transposon adjacent to IHT-E, which varies

Fig. 8. *Neisseria* metabolism. The proposed metabolic pathways for *N. meningitidis* MC58 based on the genome annotation are mapped (adapted from Nelson *et al.*, 2001; Tettelin *et al.*, 2000). Compounds are colour-coded by role: glycolysis/gluconeogenesis/TCA cycle intermediates (purple), amino acids (green), vitamins (orange), and other (red). Transporters are grouped by substrate specificity: inorganic cations (green), inorganic anions (magenta), carbohydrates and carboxylates (yellow), amino acids/peptides/amines/nucleotides/nucleosides (red) and drug/polysaccharide efflux or unknown (black). Outer-membrane porins are represented in blue. Question marks associated with transporters indicate a putative gene, uncertainty in substrate specificity, or uncertainty in direction of transport. Dashed lines are only used for aesthetic reasons and do not differ from solid lines. Pathways marked with a large pink 'X' are those predicted to be found only in pathogenic strains and missing in *N. lactamica*. Both gluconate and methylcitrate utilization have been shown in other pathogens to be involved in intracellular growth. Gluconate and/or methylcitrate utilization may be important for growth/survival in epithelial cells.

Table 4. Core *N. meningitidis* genes absent/divergent in *N. lactamica* and *N. gonorrhoeae* but present in all *N. meningitidis* strains

Common name	MC58	Z2491	FAM18
Conserved hypothetical protein	NMB_0226	NMA_0032	NMC_0225
Conserved hypothetical protein	NMB_0227	NMA_0031	NMC_0226
Conserved hypothetical protein	NMB_0228	NMA_0030	NMC_0227
Conserved hypothetical protein, authentic frameshift	NMB_0229	NMA_0029	NMC_0228
Conserved hypothetical protein	NMB_0230	NMA_0028	NMC_0229
TonB-dependent receptor, putative	NMB_0293	NMA_2193	NMC_1888
Thiol : disulfide interchange protein DsbA	NMB_0294	NMA_2191	NMC_1885
Hypothetical protein	NMB_0504	NMA_0692	NMC_0452
Iron-regulated protein FrpA, putative	NMB_0585	NMA_0788	NMC_1805
Cu-Zn superoxide dismutase	NMB_1398	NMA_1617	NMC_1339
FrpA/C-related protein*	NMB_1405	772077..772619	NMC_1345
	NMB_0585	1529861..1530442	NMC_1346
	NMB_1415		NMC_0527
Iron-regulated protein FrpC	NMB_1415	NMA_0788	NMC_1805
Transporter, putative	NMB_1732	NMA_1988	NMC_1652
Secretion protein, putative	NMB_1737	NMA_1994	NMC_1657
Secretion protein, putative	NMB_1738	NMA_1996	NMC_1658
Hypothetical protein	NMB_1739	NMA_1997	NMC_1659
Hypothetical protein	NMB_1740	NMA_1998	NMC_1660
Hypothetical protein	NMB_1782	NMA_0684	NMC_0440

*Numerous regions of each chromosome can hybridize to the amplicon for the FrpA/C-related protein. In Z2491, these regions were not annotated so the genome coordinates are given. In MC58 and FAM18, only the annotated genes are listed.

across all mCGH groups, contained genes that likely encode two of three components of a type I secretion system (the MFP and ABC components). These proteins are most similar (41 % and 55 % amino acid identity) to uncharacterized proteins encoded on a *Moraxella catarrhalis* plasmid. *M. catarrhalis* can cause otitis media, sinusitis, bronchitis and pneumonia, and was once thought to be closely related to *Neisseria* species (Janda & Knapp, 2003). This highest similarity to a γ -proteobacterial protein suggests possible ancient horizontal transfer between these respiratory colonizers. Type I secretion systems are often involved in pathogenicity as they secrete toxins, degradative enzymes and antibiotics. In *Moraxella* species and *N. meningitidis*, other type I secretion systems have been shown to be involved in secretion of RTX toxin-like exoproteins (Angelos *et al.*, 2003; Wooldridge *et al.*, 2005). In *N. meningitidis*, the function of these RTX toxin-like exoproteins has remained elusive, but in *Moraxella bovis* and other γ -proteobacteria, they appear to be functional haemolysins (Angelos *et al.*, 2003; Wooldridge *et al.*, 2005). The MFP and ABC components of the type I secretion system are often genetically linked to the exported substrate/toxin. A third large open reading frame in the transposon (NMC_0891 and a conserved hypothetical protein with significant matches only to *N. meningitidis* strains) should be examined as a potential substrate for this type I secretion system. The third component necessary for a functional secretion system (TolC) has been identified in *N. meningitidis* strains

(Wooldridge *et al.*, 2005) and is typically not genetically linked to the MFP and ABC components.

While many of the islands are in defined subsets of closely related strains, others, like IHT-B and IHT-C, vary in their presence within diverse mCGH groups (Table 1). Despite this variable presence across strains of the different mCGH groups, strains containing IHT-B always contained IHT-C. IHT-B and IHT-C contain genes for numerous hypothetical proteins and large (>5 kb) genes for haemagglutinin/haemolysin-related proteins. Upon comparing the genomes of FA1090 and Z2491, IHT-B and IHT-C are the sites for a synteny break with MC58. This raises the possibility that they were once a single inheritable island that was broken by a genome rearrangement in MC58. The variability of the IHT-B/IHT-C islands across the mCGH groups suggests that these islands have been gained and lost on multiple occasions, possibly owing to the recombinant nature and transformability of *Neisseria* species (Feil *et al.*, 2001). However, these events occur infrequently, such that the genomic-level relationship between strains, apparent from the presence of more characteristic islands, appears to remain intact.

Restriction/modification systems

One class of genes that is found to vary extensively across the different clusters is composed of restriction/modification systems (RMSs). It has been proposed that in *N.*

meningitidis, RMSs could be used to restrict DNA uptake to only organisms within its clonal complex or sequence type (Claus *et al.*, 2000). This conclusion is supported in this study by the clustering of genes by sequence types with RMSs. For example, three intact genes of an RMS are present in the *N. gonorrhoeae* capsule locus, while the endonuclease is absent from all *N. meningitidis* strains examined (Fig. S1). The modification enzymes are truncated and frame-shifted in the sequenced *N. meningitidis* strains, making their functional presence difficult to assess with mCGH. But loss of this system in all the *N. meningitidis* strains may prevent transfer of the *N. meningitidis* capsule genes into *N. gonorrhoeae* strains.

Comparisons with other genome-based studies

Some of the *N. meningitidis*-specific genes, core genes and pathogen-specific genes have been described previously (Bille *et al.*, 2005; Perrin *et al.*, 1999, 2002; Snyder *et al.*, 2004, 2005; Snyder & Saunders, 2006; Stabler & Hinds, 2006; Stabler *et al.*, 2005). However, the alternative methods, different microarray platforms, the breadth of strains examined, and dissimilar analysis methods lead to some differences (Table S2). The interpretation of mCGH hybridizations is highly dependent on platform and analysis methods. Despite these differences, this study together with the series of CGH studies with *Neisseria* spp. recently presented in the literature with their unique emphases on (1) cross-species comparisons, (2) particular islands of horizontal transfer, (3) phylogeny, (4) particular regions of biological interest, and/or (5) total genomic content each identify important gene sets that may be important for virulence. Because of the diverse emphases between the studies, we have limited a comparison to a tabular format that enables researchers to identify common and differing results of interest in a relatively straightforward manner. This includes select genomic results from genome sequencing (Parkhill *et al.*, 2000; Tettelin *et al.*, 2000), mCGH (Bille *et al.*, 2005; Perrin *et al.*, 1999, 2002; Snyder *et al.*, 2004, 2005; Snyder & Saunders, 2006; Stabler & Hinds, 2006; Stabler *et al.*, 2005), expression microarray (Dietrich *et al.*, 2003; Grifantini *et al.*, 2002) and signature-tagged mutagenesis results (Sun *et al.*, 2000) (Table S2).

Absence/divergence of metabolic genes

Some genes and gene clusters from metabolic pathways were identified as being absent/divergent in commensal *Neisseria* species. Several of these constitute the absence/divergence of an entire pathway in the commensal organisms (Fig. 8). The D-amino acid dehydrogenase and the putative sodium/alanine symporter were found to be differentially regulated upon contact with host cells (Dietrich *et al.*, 2003; Grifantini *et al.*, 2002). Both gluconate and methylcitrate utilization have been shown in other pathogens to be involved in intracellular growth within macrophage cells (Brämer & Steinbüchel, 2001; Eriksson *et al.*, 2003; Stone *et al.*, 1999). *N. meningitidis* and *N. gonorrhoeae* are both internalized by epithelial cells, where they survive and grow in the process of epithelial cell traversal (Merz

& So, 2000; Nassif *et al.*, 1999). Gluconate and/or methylcitrate utilization may thus be important for growth/survival in epithelial cells and may be novel antimicrobial targets.

Concluding remarks

mCGH experiments allow for rapid parallel comparison of the gene content of a wide variety of strains. This facilitates understanding of genomic differences between diverse species and strains, and provides insights into their pathogenic potential. In the case of the *Neisseriaceae*, organisms as distantly related as *N. lactamica* and *N. gonorrhoeae* could be examined reliably owing to the representation of four genomes on the microarray platform. This allowed the detection of a set of genes transferred between *N. meningitidis* and *N. lactamica*, as well as the identification of insertion/deletion events specific to groups of strains, identification of the core *N. meningitidis* genome, and insights into horizontal gene transfer in *N. meningitidis*. Acquisition of some genetic islands appears to have occurred in multiple lineages, and their study revealed important sets of genes related to pathogenicity. Acquisition of islands does not occur so frequently as to obscure the genomic-level relationship within the *Neisseria* population. Genomic changes parallel the MLST profiles of the strains examined, suggesting that subsequent sequencing of genomes based on their MLST profiles would provide additional clues regarding pathogenicity and may identify unique vaccine or antimicrobial targets.

ACKNOWLEDGEMENTS

We would like to thank Scott Peterson for assistance with mCGH methods; Stephen Bentley for access to the FAM18 genome and annotation; Leonard Mayer and the CDC Meningitis and Special Pathogens Laboratory for some strains; and Karen Nelson and Ian Paulsen for their assistance in re-examining *Neisseria* metabolism. This study was supported by Chiron Corporation; D. S. S. and Y. T. were supported by NIH grants (R01 AI-33517 and R01 AI-40247 to D. S. S.) and the Georgia Emerging Infections Program.

REFERENCES

- Achtman, M. (1995). Global epidemiology of meningococcal disease. In *Meningococcal Disease*, pp. 159–175. Edited by K. Cartwright. Wiley.
- Aguilera, J.-F., Perrocheau, A., Meffre, C., Hahné, S. & Group, T. W. W. (2002). Outbreak of the serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. *Emerg Infect Dis* 8, 761–767.
- Angelos, J. A., Hess, J. F. & George, L. W. (2003). An RTX operon in hemolytic *Moraxella bovis* is absent from nonhemolytic strains. *Vet Microbiol* 92, 363–377.
- Bille, E., Zahar, J. R., Perrin, A. & 7 other authors (2005). A chromosomally integrated bacteriophage in invasive meningococci. *J Exp Med* 201, 1905–1913.
- Bovre, K. (1984). Family VIII. *Neisseriaceae*. In *Bergey's Manual of Systematic Bacteriology*, pp. 288–296. Edited by N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins.

- Brämer, C. O. & Steinbüchel, A. (2001). The methylcitric acid pathway in *Ralstonia eutropha*: new genes identified involved in propionate metabolism. *Microbiology* **147**, 2203–2214.
- Claus, H., Friedrich, A., Frosch, M. & Vogel, U. (2000). Differential distribution of novel restriction-modification systems in clonal lineages of *Neisseria meningitidis*. *J Bacteriol* **182**, 1296–1303.
- Claus, H., Borrow, R., Achtman, M., Morelli, G., Kantelberg, C., Longworth, E., Frosch, M. & Vogel, U. (2004). Genetics of capsule O-acetylation in serogroup C, W-135 and Y meningococci. *Mol Microbiol* **51**, 227–239.
- Comanducci, M., Bambini, S., Brunelli, B. & 12 other authors (2002). NadA, a novel vaccine candidate of *Neisseria meningitidis*. *J Exp Med* **195**, 1445–1454.
- Dempsey, J. A., Litaker, W., Madhure, A., Snodgrass, T. L. & Cannon, J. G. (1991). Physical map of the chromosome of *Neisseria gonorrhoeae* FA1090 with locations of genetic markers, including *opa* and *pil* genes. *J Bacteriol* **173**, 5476–5486.
- Dietrich, G., Kurz, S., Hübner, C., Aepinus, C., Theiss, S., Guckenberger, M., Panzner, U., Weber, J. & Frosch, M. (2003). Transcriptome analysis of *Neisseria meningitidis* during infection. *J Bacteriol* **185**, 155–164.
- Dolan-Livengood, J. M., Miller, Y. K., Martin, L. E., Urwin, R. & Stephens, D. S. (2003). Genetic basis for nongroupable *Neisseria meningitidis*. *J Infect Dis* **187**, 1616–1628.
- Engelberg-Kulka, H. & Glaser, G. (1999). Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu Rev Microbiol* **53**, 43–70.
- Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J. C. D. (2003). Unraveling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* **47**, 103–118.
- Feil, E., Zhou, J., Maynard Smith, J. & Spratt, B. G. (1996). A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: evidence for extensive interspecies recombination within *adk*. *J Mol Evol* **43**, 631–640.
- Feil, E. J., Holmes, E. C., Bessen, D. E. & 9 other authors (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A* **98**, 182–187.
- Graur, D. & Li, W.-H. (2000). In *Fundamentals of Molecular Evolution*, 2nd edn, pp. 209–210. Sunderland, MA: Sinauer Associates.
- Grifantini, R., Bartolini, E., Muzzi, A. & 14 other authors (2002). Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. *Nat Biotechnol* **20**, 914–921.
- Janda, W. M. & Knapp, J. S. (2003). *Neisseria* and *Moraxella catarrhalis*. In *Manual of Clinical Microbiology*, pp. 585–608. Edited by P. R. Murray, E. J. Baron, J. H. Tenover & R. M. Tenover. Washington, DC: American Society for Microbiology.
- Kahler, C. M., Blum, E., Miller, Y. K., Ryan, D., Popovic, T. & Stephens, D. S. (2001). *exl*, an exchangeable genetic island in *Neisseria meningitidis*. *Infect Immun* **69**, 1687–1696.
- Kim, C. C., Joyce, E. A., Chan, K. & Falkow, S. (2002). Improved analytical methods for microarray-base genome-composition analysis. *Genome Biol* **3**, research0065.0061–0065.0017.
- Kyaw, M. H., Lynfield, R., Schaffner, W. & 14 other authors (2006). Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant *Streptococcus pneumoniae*. *N Engl J Med* **354**, 1455–1463.
- Lindroos, H. L., Mira, A., Repsilber, D., Vinnere, O., Naslund, K., Dehio, M., Dehio, C. & Andersson, S. G. (2005). Characterization of the genome composition of *Bartonella koehlerae* by microarray comparative genomic hybridization profiling. *J Bacteriol* **187**, 6155–6165.
- Linz, B., Schenker, M., Zhu, P. & Achtman, M. (2000). Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. *Mol Microbiol* **36**, 1049–1058.
- Maiden, M. C. J., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.
- Masignani, V., Giuliani, M. M., Tettelin, H., Comanducci, M., Rappuoli, R. & Scarlato, V. (2001). Mu-like prophage in serogroup B *Neisseria meningitidis* coding for surface-exposed antigens. *Infect Immun* **69**, 2580–2588.
- McAllister, C. F. & Stephens, D. S. (1993). Analysis in *Neisseria meningitidis* and other *Neisseria* species of genes homologous to the FKBP immunophilin family. *Mol Microbiol* **10**, 13–23.
- McGuinness, B. T., Clarke, I. N., Lambden, P. R., Barlow, A. K., Poolman, J. T., Jones, D. M. & Heckels, J. E. (1991). Point mutation in meningococcal Por-A gene associated with increased endemic disease. *Lancet* **337**, 514–517.
- Merz, A. J. & So, M. (2000). Interactions of pathogenic neisseriae with epithelial cell membranes. *Annu Rev Cell Dev Biol* **16**, 423–457.
- Nassif, X., Pujol, C., Morand, P. & Eugène, E. (1999). Interactions of pathogenic *Neisseria* with host cells. Is it possible to assemble the puzzle? *Mol Microbiol* **32**, 1124–1132.
- Nelson, K. E., Paulsen, I. T. & Fraser, C. M. (2001). Microbial genome sequencing: a window into evolution and physiology. *ASM News* **67**, 310–317.
- Oster, P., Lennon, D., O'Hallahan, J., Mulholland, K., Reid, S. & Martin, D. (2005). MeNZB: a safe and highly immunogenic tailor-made vaccine against the New Zealand *Neisseria meningitidis* serogroup B disease epidemic strain. *Vaccine* **23**, 2191–2196.
- 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.
- Perrin, A., Nassif, X. & Tinsley, C. (1999). Identification of regions of the chromosome of *Neisseria meningitidis* and *Neisseria gonorrhoeae* which are specific to the pathogenic *Neisseria* species. *Infect Immun* **67**, 6119–6129.
- Perrin, A., Bonacorsi, S., Carbonnelle, E., Talibi, D., Dessen, P., Nassif, X. & Tinsley, C. (2002). Comparative genomics identifies the genetic islands that distinguish *Neisseria meningitidis*, the agent of cerebrospinal meningitis, from other *Neisseria* species. *Infect Immun* **70**, 7063–7072.
- Pinner, R. W., Onyango, F., Perkins, B. A. & 7 other authors (1992). Epidemic meningococcal disease in Nairobi, Kenya, 1989. The Kenya/Centers for Disease Control (CDC) Meningitis Study Group. *J Infect Dis* **166**, 359–364.
- Pizza, M., Scarlato, V., Masignani, V. & 33 other authors (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820.
- Raymond, N. J., Reeves, M., Ajello, G., Baughman, W., Gheesling, L. L., Carlone, G. M., Wenger, J. D. & Stephens, D. S. (1997). Molecular epidemiology of sporadic (endemic) serogroup C meningococcal disease. *J Infect Dis* **176**, 1277–1284.
- Read, T. D., Peterson, S. N., Tourasse, N. & 49 other authors (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**, 81–86.
- Saeed, A. I., Sharove, V., White, J. & 18 other authors (2003). TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**, 374–378.

- Seiler, A., Reinhardt, R., Sarkari, J., Caugant, D. A. & Achtman, M. (1996). Allelic polymorphism and site-specific recombination in the *opc* locus of *Neisseria meningitidis*. *Mol Microbiol* **19**, 841–856.
- Snyder, L. A. & Saunders, N. J. (2006). The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as virulence genes. *BMC Genomics* **7**, 128.
- Snyder, L. A., Davies, J. K. & Saunders, N. J. (2004). Microarray genotyping of key experimental strains of *Neisseria gonorrhoeae* reveals gene complement diversity and five new neisserial genes associated with Minimal Mobile Elements. *BMC Genomics* **5**, 23.
- Snyder, L. A., Jarvis, S. A. & Saunders, N. J. (2005). Complete and variant forms of the 'gonococcal genetic island' in *Neisseria meningitidis*. *Microbiology* **151**, 4005–4013.
- Soukas, A., Cohen, P., Socci, N. D. & Friedman, J. M. (2000). Leptin-specific patterns of gene expression in white adipose tissue. *Genes Dev* **14**, 963–980.
- Stabler, R. & Hinds, J. (2006). The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as virulence genes: response. *BMC Genomics* **7**, 129.
- Stabler, R. A., Marsden, G. L., Witney, A. A., Li, Y., Bentley, S. D., Tang, C. M. & Hinds, J. (2005). Identification of pathogen-specific genes through microarray analysis of pathogenic and commensal *Neisseria* species. *Microbiology* **151**, 2907–2922.
- Stephens, D. S. (1999). Uncloaking the meningococcus: dynamics of carriage and disease. *Lancet* **353**, 941–942.
- Stone, B. J., Brier, A. & Kwaik, Y. A. (1999). The *Legionella pneumophila prp* locus; required during infection of macrophages and amoebae. *Microb Pathog* **27**, 368–376.
- Sun, Y.-H., Bakshi, S., Chalmers, R. & Tang, C. M. (2000). Functional genomics of *Neisseria meningitidis* pathogenesis. *Nat Med* **6**, 1269–1273.
- Swartley, J. S., Marfin, A. A., Edupuganti, S., Liu, L. J., Cieslak, P., Perkins, B., Wenger, J. D. & Stephens, D. S. (1997). Capsule switching of *Neisseria meningitidis*. *Proc Natl Acad Sci U S A* **94**, 271–276.
- Tettelin, H., Saunders, N. J., Heidelberg, J. & 39 other authors (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815.
- Tettelin, H., Nelson, K. E., Paulsen, I. T. & 36 other authors (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506.
- Tzeng, Y. L., Datta, A. K., Strole, C. A., Lobritz, M. A., Carlson, R. W. & Stephens, D. S. (2005). Translocation and surface expression of lipidated serogroup B capsular polysaccharide in *Neisseria meningitidis*. *Infect Immun* **73**, 1491–1505.
- van der Ley, P. & Poolman, J. T. (1992). Construction of a multivalent meningococcal vaccine strain based on the class-1 outer-membrane protein. *Infect Immun* **60**, 3156–3161.
- Wang, J. F., Caugant, D. A., Li, X., Hu, X., Poolman, J. T., Crowe, B. A. & Achtman, M. (1992). Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China. *Infect Immun* **60**, 5267–5282.
- Wang, J. F., Caugant, D. A., Morelli, G., Koumare, B. & Achtman, M. (1993). Antigenic and epidemiologic properties of the ET-37 complex of *Neisseria meningitidis*. *J Infect Dis* **167**, 1320–1329.
- Wooldridge, K. G., Kizil, M., Wells, D. B. & Ala'aldin, D. A. (2005). Unusual genetic organization of a functional type I protein secretion system in *Neisseria meningitidis*. *Infect Immun* **73**, 5554–5567.