Microarray genomotyping of key experimental strains of *Neisseria gonorrhoeae* reveals gene complement diversity and five new neisserial genes associated with Minimal Mobile Elements.

Lori AS Snyder¹, John K Davies² and Nigel J Saunders*¹

Address: ¹Bacterial Pathogenesis and Functional Genomics Group, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom and ²Bacterial Pathogenesis Research Group, Department of Microbiology, Monash University, VIC 3800, Australia

Email: Lori AS Snyder - Lori.Snyder@pathology.oxford.ac.uk; John K Davies - John.Davies@med.monash.edu.au; Nigel J Saunders* - Nigel.Saunders@pathology.oxford.ac.uk

* Corresponding author

Abstract

**Background:** There are four widely used experimental strains of *N. gonorrhoeae*, one of which has been sequenced and used as the basis for the construction of a multi-strain, multi-species pan-neisserial microarray. Although the *N. gonorrhoeae* population structure is thought to be less diverse than *N. meningitidis*, there are some recognized gene-complement differences between strains, including the 59 genes of the Gonococcal Genetic Island. In this study we have investigated the three experimental strains that have not been sequenced to determine the extent and nature of their similarities and differences.

**Results:** Using the Pan-Neisseria microarray, three commonly used gonococcal laboratory experimental strains were investigated (F62, MS11, & FA19). Genes absent from these strains, but present in strain FA1090, were assessed as is possible with typical microarrays. Due to the design of this microarray, additional genes were also identified. Differences were associated with Minimal Mobile Elements (MMEs) or known divergences. Genomotyping indicates the presence of genes previously only described in meningococci and shows the presence of the complete Gonococcal Genetic Island in *N. gonorrhoeae* strain FA19. Five new neisserial genes were identified through microarray genomotyping and subsequent sequencing of two divergent MMEs in *N. gonorrhoeae* strain MS11 and four MMEs in *N. gonorrhoeae* strain FA19. No differences were identified between *N. gonorrhoeae* strains FA1090 and F62, indicating that these strains are very similar.

**Conclusion:** This study shows extensive similarity between the experimental strains, associated with a varying number of strain-specific genes. This provides a framework for those working with these strains to refer to the available gonococcal genome sequence, and is the first detailed comparison of gene complements between gonococcal strains.

---

**Background**

The *Neisseria gonorrhoeae* population structure is not clonal [1-5], its panmictic structure being the result of horizontal genetic exchange [6]. The pathogenic *Neisseria* spp. are naturally transformable [7] and chromosomal changes are largely mediated by homologous
recombination. Natural competence for transformation and homologous recombination allow the generation of mosaic genes, such as the divergent *opa* genes [8,9], antigenic variation through recombination with silent cassettes, such as in the *pilE/pilS* system [10,11], and the horizontal exchange of whole genes or groups of genes in Minimal Mobile Elements (MMEs) [12]. A Minimal Mobile Element (MME) is defined as a region between two conserved genes within which different genes are found in different strains. These elements are described as 'minimal' reflecting the fact that they do not have features suggesting mobilization by associated transposases, nor do they currently appear to have features by which remotely encoded systems would facilitate mobilization. These were initially identified through comparisons of the available *Neisseria* spp. genome sequences, on which basis the MM*EpheS-pheT* was the first to be studied in an extended set of strains [12]. In this region, flanked by the highly conserved *pheS* and *pheT* genes, eight different intergenic regions where found in the *Neisseria* strains that were investigated. Polymorphisms within the flanking gene sequences are indicative of the points of homologous recombination between the imported DNA and the native DNA. Further, assessments of five additional MMEs in the genome sequences show these to be present in different combinations in different strains, supporting the model that these are horizontally, rather than vertically, transferred elements [12].

*N. gonorrhoeae* may show less diversity than *Neisseria meningitidis* because it is more likely to be isolated away from other *Neisseria* spp. in its primary host niche [1]. The exchange of highly similar sequences between strains would prevent sequence analysis revealing the full extent of horizontal exchange. It is quite possible that recombination within genes coding for surface exposed proteins, or those under direct selection, may reveal a higher degree of recombinatorial diversity than is currently established from the predominantly housekeeping gene sequences investigated to date [10,13-15].

There are four commonly used laboratory strains of *N. gonorrhoeae*. The complete genome sequence is available for *N. gonorrhoeae* strain FA1090 [16], for which there is also a physical map [17,18]. This strain was originally recovered from a patient with disseminated gonococcal infection (DGI) in 1983 [19]. There is also a physical map of the *N. gonorrhoeae* strain MS11 chromosome [20]. It is from strain MS11 that the Gonococcal Genetic Island (GGI) has been sequenced [21]. The GGI is composed of 59 predicted coding regions, including genes believed to be involved in the secretion of DNA. This strain was isolated from a patient with uncomplicated gonorrhoea [22]. *N. gonorrhoeae* strain F62 was also isolated from a patient with uncomplicated gonorrhoeae [23], but it does not contain the GGI [21]. The last of the four commonly used laboratory strains of *N. gonorrhoeae* is strain FA19, which was isolated from a patient with both gonorrhea and DGI. Additionally, it has caused DGI in a laboratory worker, following a mouth-pipetting accident [24]. While other strains are certainly used experimentally, these four represent those most commonly used and for which there is the greatest body of existing literature available.

Since these strains have been widely used in experimental studies it is important to be able to compare them, and to provide a means of relating each to the available genome sequence of *N. gonorrhoeae* strain FA1090. Some differences between these four commonly used laboratory strains, in addition to the presence/absence of the GGI [21], have been noted previously. The expression of ICAM-1 by human mucosal epithelial cells is upregulated in response to strain FA1090, but not to strain MS11 [25]. Truncations of the protegrin PG-1 showed different potencies against strains FA19 and F62 [26]. There are two immunological classes of the major outer membrane protein Protein I (encoded by *porB*), designated PIA and PIB [2]. Strains F62 and FA19 have PIA, while strains FA1090 and MS11 have PIB [2,27,28]. While this is not an exhaustive list, it does indicate that some difference in gene complement between these strains may be expected, and defining the presence and absence of gene-complement differences may provide keys for the community to address their significant functional differences.

The Pan-*Neisseria* microarray was constructed by an international consortium of neisseriologists (Saunders, Davies, et al – in preparation). The probes printed onto this microarray represent the coding sequences of three complete neisserial genomes (*N. gonorrhoeae* strain FA1090 [16], *N. meningitidis* strain MC58 [29], *N. meningitidis* strain Z2491 [30]) as well as the coding regions from the GGI [21]. For those coding regions of sufficient length, each probe is complementary to 150–450 bp of the most conserved regions of the CDS between the different genome sequences, that are also best able to discriminate between paralogous genes. Basing the design upon the most conserved regions between the different strains and species makes the array minimally strain specific facilitating the use the Pan-*Neisseria* microarray in analyzing strains to which it was not designed.

This paper addresses the gene complement differences between the four most commonly used laboratory strains, utilizing the Pan-*Neisseria* microarray, comparative genomic analysis, and sequencing of the newly identified divergent MMEs revealed by the microarray data. This is the first use of a multi-species, multi-genome microarray to conduct gene complement studies. Additionally, this
work supports and expands our observation of Minimal Mobile Elements in the *Neisseria* spp. [12].

**Results**

**Comparison of *N. gonorrhoeae* strains F62, MS11, and FA19 to the sequenced strain FA1090 using the Pan-*Neisseria* microarray**

Chromosomal DNA from *N. gonorrhoeae* strain FA1090 was fluorescently labelled with Cy3-dCTP and mixed with Cy5-dCTP labelled chromosomal DNA from either *N. gonorrhoeae* strain F62, MS11, or FA19, and hybridized to the Pan-*Neisseria* microarray slide. Analysis of labelled DNA hybridization indicated that there are gene complement differences between strains FA1090 and MS11 (Table 1), and strains FA1090 and FA19 (Table 2). No differences in gene complement were noted in microarray data comparison of *N. gonorrhoeae* strains FA1090 and F62. To confirm the original results for strains FA1090 and F62 these strains were re-sourced from an additional laboratory and the experiments repeated. The repeat experiment confirmed the absence of any differences.

**Table 1: Genomotyping results of *N. gonorrhoeae* strain FA1090 versus *N. gonorrhoeae* strain MS11.**

| Genes known to be associated with divergence: | Lack of hybridization may be due to: |
|-----------------------------------------------|-------------------------------------|
| XNG0049 (pilC1)                               | divergence in MS11 pilC sequence (Z50180) |
| XNG0210 (hypothetical)                        | maf-associated divergence            |
| XNG0211 (hypothetical)                        | maf-associated divergence            |
| XNG0212 (hypothetical)                        | maf-associated divergence            |
| XNG1280 (hypothetical)                        | maf-associated divergence            |
| XNG0442 (hypothetical)                        | prophage region-associated divergence|
| XNG1509 (hypothetical)                        | prophage region-associated divergence|

| Genes associated with strain-specific regions: | Associated strain-specific region: |
|-----------------------------------------------|-------------------------------------|
| XNG0080 (hypothetical)                        | XNG0080 (hypothetical) hypo-pglB MME|
| XNG0081 (hypothetical)                        | XNG0081 (hypothetical) hypo-pglB MME|
| XNG0528 (hypothetical)                        | XNG0528 (hypothetical) hypo-aspa MME|
| XNG0529 (hypothetical)                        | XNG0529 (hypothetical) hypo-aspa MME|
| XNG0669 (hypothetical)                        | XNG0669 (hypothetical) ribAB-hypo MME|
| XNG0670 (type I restriction enzyme M protein) | XNG0670 (type I restriction enzyme M protein) ribAB-hypo MME|
| XNG0671 (ribA-B)                              | XNG0671 (ribA-B) ribAB-hypo MME     |
| XNG0741 (hypothetical)                        | XNG0741 (hypothetical) ung-hypo MME  |
| XNG1618 (hypothetical)                        | XNG1618 (hypothetical) nuoL-nuoM MME|

| Gene probes hybridizing to MS11 but not FA1090 | Hybridization to non-FA1090 probes may be due to: |
|-----------------------------------------------|-------------------------------------|
| Genes known to be associated with divergence | divergence in MS11 tspB sequence     |
| NMA1797 (tspB)                                | prophage region-associated divergence|
| NMB1119 (hypothetical)                        | prophage region-associated divergence|
| NMB1120 (hypothetical)                        | prophage region-associated divergence|
| GcJ560 (suppressor)                           | prophage region-associated divergence|
| GcJ561 (regulator)                            | prophage region-associated divergence|

| Genes associated with strain-specific regions: | Associated strain-specific region: |
|-----------------------------------------------|-------------------------------------|
| 59 Gonococcal Genetic Island genes            | 59 Gonococcal Genetic Island genes |

**Differences between *N. gonorrhoeae* strains FA1090 and MS11**

The results of hybridization with strains FA1090 and MS11 are shown in Table 1. One of the two strain MS11 pilC genes [31] is sufficiently divergent not to hybridize to any of the four pilC probes on the microarray (X0049, X1797, B0049, B1847). The T-cell stimulating antigen tspB [32] is present in multiple copies in the sequenced neisserial genomes [16,29,30], and the sequences are often divergent. Hybridization of strain MS11 with a meningococcal tspB probe (A1797) is a reflection of that divergence and possible exchange between the species. Additionally, probes to five maf-associated hypothetical genes hybridized to strain FA1090 labelled chromosomal DNA, but not to that from strain MS11. Through analysis of the complete neisserial genome sequences, it has been suggested that the region downstream of mafB, which is itself present in multiple copies, contains silent cassettes of mafB akin to the silent pilS cassettes that are capable of homologous recombination into the expression locus pilE [30]. The regions containing the mafA and mafB genes and
the proposed silent mafB cassettes also contain many hypothetical genes. These features of the maf regions, present in the neisserial genomes in up to four locations [16], are not conserved, and are composed largely of short hypothetical genes of low %G+C separating what are believed to be the silent mafB recombination exchange cassettes [30]. It is these maf-associated low %G+C potential short coding sequences that appear to be divergent between the gonococcal strains, which is also the case between the four sequenced genome strains. Also present in strain FA1090, but absent from strain MS11, are two prophage-associated hypothetical genes. Additionally, there are four prophage-associated genes present in strain MS11, which are absent from strain FA1090. As prophages and prophage-associated genes are horizontally transferred and as each of the four neisserial genome sequences has quite different prophage and prophage-associated gene complements, it is not unexpected that differences in hybridization to these sequences would be found between the gonococcal strains.

All of the other differences between N. gonorrhoeae strain FA1090 and N. gonorrhoeae strain MS11 are associated with MMEs [12], the sites of which could be determined by comparison of the three complete neisserial genome sequences [16,29,30]. One of these MME regions is the GGI (Figure 1), which has been sequenced previously from strain MS11 [21]. The four remaining MMEs showing differences in the presence of strain FA1090 genes

| Genes known to be associated with divergence | Lack of hybridization may be due to: |
|---------------------------------------------|-------------------------------------|
| XNG0049 (pilC1)                            | divergence in FA19 pilC sequence    |
| XNG0211 (hypothetical)                     | maf-associated divergence           |
| XNG0212 (hypothetical)                     | maf-associated divergence           |
| XNG0213 (hypothetical)                     | maf-associated divergence           |
| XNG1848 (hypothetical)                     | maf-associated divergence           |
| XNG0458 (hypothetical)                     | maf-associated divergence           |
| XNG0459 (regulator)                        | maf-associated divergence           |
| XNG0460 (hypothetical)                     | maf-associated divergence           |
| XNG0461 (regulator)                        | maf-associated divergence           |
| XNG0462 (hypothetical)                     | maf-associated divergence           |
| XNG1006 (hypothetical)                     | maf-associated divergence           |
| XNG1509 (hypothetical)                     | maf-associated divergence           |

| Genes associated with strain-specific regions: | Associated strain-specific region: |
|-----------------------------------------------|-----------------------------------|
| XNG0438 (hypothetical)                       | pilin-AzC MME                     |
| XNG0741 (hypothetical)                       | ung-hypo MME                      |
| XNG0528 (hypothetical)                       | hypo-aspA MME                     |
| XNG0529 (hypothetical)                       | hypo-aspA MME                     |
| XNG1114 (hypothetical)                       | uvrA-hypo MME                     |
| XNG1115 (methylase)                          | uvrA-hypo MME                     |
| XNG1441 (hypothetical)                       | nadC-xthA MME                     |

| Genes known to be associated with divergence | Hybridization to non-FA1090 probes may be due to: |
|---------------------------------------------|---------------------------------------------------|
| NMB1541 (lbpB)                              | lbpB is known to be dead in FA1090 (AF072890)     |
| NMA0036 (hypothetical)                      | maf-associated divergence                         |
| NMB2109 (hypothetical)                      | maf-associated divergence                         |
| NMB2110 (hypothetical)                      | maf-associated divergence                         |
| NMA1360 (hypothetical)                      | the presence of a sequenced gene fragment in FA19 |
| NMB1119 (hypothetical)                      | prophage region-associated divergence             |
| NMB1120 (hypothetical)                      | prophage region-associated divergence             |
| GCJD560 (suppressor)                        | prophage region-associated divergence             |
| GCJD561 (regulator)                         | prophage region-associated divergence             |

| Genes associated with strain-specific regions: | Associated strain-specific region: |
|-----------------------------------------------|-----------------------------------|
| NMA2121 (hypothetical)                       | hypo-hypo MME                     |
| 59 Gonococcal Genetic Island genes           | ung-hypo MME                      |

Table 2: Genomotyping results of N. gonorrhoeae strain FA1090 versus N. gonorrhoeae strain FA19.
were investigated further through PCR amplification and sequencing.

The MME flanked by *nuoL* and *nuoM* is shown in Figure 2, from which four newly described neisserial genes were sequenced from strain MS11 (Accession AY386266). These lack any significant similarity to sequences in the public databases. Although these genes are not represented on the Pan-*Neisseria* microarray, it was possible using the microarray data to identify the *nuoL*-*nuoM* region as a target for sequencing due to the lack of hybridization of the labelled strain MS11 DNA to the strain FA1090 probe for XNG1618 or to the alternate *N. meningitidis* strain Z2491 probes for NMA0001 and NMA2230.

The MME flanked by a putative lipopolysaccharide biosynthesis protein and *pglB* is shown in Figure 3. In *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strains Z2491 and FAM18, there are two hypothetical genes (NMA0640 &NMA0641) in the intervening region. In *N. gonorrhoeae* strain MS11, however, there is a short (167 bp) intergenic region (Accession AY386267), similar to that found in *N. meningitidis* strain MC58.

The remaining two MMEs, identified through absence of hybridization of microarray probes with gonococcal strain MS11, failed to amplify either from strain MS11 or from control strain FA1090. The first is flanked by a hypothetical gene and *aspA*, containing three hypothetical genes in *N. gonorrhoeae* strain FA1090, spanning a region of over 7 kb. Of these three genes, two do not hybridize with the FA1090-derived probes, and appear to be absent in strain MS11. The second of these remaining MMEs is flanked by a hypothetical gene and *ribA-B* and in *N. gonorrhoeae* strain FA1090 this region, of approximately 11.5 kb, contains genes encoding a restriction-modification system. The size of these MMEs in strain FA1090 (7 & 11.5 kb), may account for the failure to amplify these regions.
Figure 3
The MME flanked by a putative LPS biosynthesis gene and pgIB. This MME contains two hypothetical genes in N. gonorrhoeae strain FA1090 [16], N. meningitidis strain Z2491 [30], and N. meningitidis strain FAM18; and a short intergenic region in N. meningitidis strain MC58 [29] and N. gonorrhoeae strain MS11 (AY386267). Similar genes are indicated in the same colour.

Differences between N. gonorrhoeae strains FA1090 and FA19
As with gonococcal strain MS11, divergence is indicated by the genomotyping results for the strain FA19 pilC gene, maf-associated hypothetical genes, and prophage-associated genes (Table 2). Additional points of variation between N. gonorrhoeae strain FA1090 and strain FA19 have been confirmed, by sequencing, to be due to hybridization to sequences not present in strain FA1090.

Cross hybridizations to divergent sequences revealed through sequencing
Probe A0036 was designed to NMA0036, which is within an MME. In N. gonorrhoeae strain FA1090 XNG1638 is in this MME location, probed by X1638. Strain FA19 genomic DNA hybridized both with probe A0036 and X1638, suggesting that both genes, which were believed to be mutually exclusive in this MME, are present in strain FA19. This MME was therefore amplified and sequenced from strain FA19 (data not presented). This sequence shows that the MME region is similar to that present in strain FA1090 and did not contain the MME-associated gene NMA0036, therefore the hybridization of strain FA19 to probe A0036 must be to a sequence outside of this MME. The probe A0036 cross-hybridizes in N. meningitidis strain MC58 at 86% over 132 bp with NMB0656, which is a maf-associated hypothetical gene. In N. gonorrhoeae strain FA19, therefore, this probe may be hybridizing to a maf-associated region, rather than to the MME as initially suspected.

A second probe designed against N. meningitidis strain Z2491, A1360, was found to hybridize the FA19 labelled genomic DNA. It was known that this probe cross-hybridizes slightly to fragments of the meningococcal gene in strain FA1090 (28 bp at 96% and 37 bp at 89%), but not at a level that would be expected to be detected on the Pan-Neisseria microarray and which did not, in fact, generate a microarray hybridization signal. Sequencing of this region from strain FA19 revealed that an additional 106 bp is present in this strain (data not presented), generating a region in strain FA19 that is 91% similar to the 181 bp length of the A1360 probe. Although this sequence homology is to a longer sequence than is in strain FA1090, it is still a fragment and not the complete meningococcal gene in strain FA19. Therefore, this result is due to sequence diversity and cross-hybridization of the probe to a gene fragment, rather than the presence of an entire gene.

Presence of the Gonococcal Genetic Island in strain FA19
Labelled chromosomal DNA from N. gonorrhoeae strain FA19 hybridized to all 59 probes for the GGI, indicating for the first time, the presence of the complete island in this strain (Figure 1).

MME-associated differences between strains FA1090 and FA19
The MME-associated region detected in strain MS11, flanked by a hypothetical gene and aspA, was also detected in strain FA19, although PCR amplification of this region from strain FA19 failed as it did with strains FA1090 and MS11 (see above). The remaining four strain FA19 MMEs, indicated to be divergent through genomotyping, revealed additional differences between the gonococcal laboratory strains (Figures 4, 5, 6, 7), including a large new neisserial gene (Figure 6).

Hypothetical gene XNG1441 is not present in strain FA19. This gene is located between nadC and xthA in strain FA1090, where it disrupts a hypothetical gene identified in N. meningitidis strain MC58, NMB0397 (Figure 4). This meningococcal hypothetical gene is disrupted by different means in all other strains from which this region has been sequenced, including in N. gonorrhoeae strain FA19 (Accession AY386268). Therefore the absence of XNG1441 in strain FA19 is confirmed by sequencing and an additional means of disrupting NMB0397 has been identified.

Hypothetical gene XNG0438 is also missing from strain FA19. In this case the MME containing this gene has conserved features in the meningococcal strains, in the form
of two nearly identical Correia Repeat Enclosed Elements (CREE) [33-35] flanking a variable length intergenic region composed largely of tandem repeats (Figure 5). The Correia Repeat [33,34] is an intergenic sequence element commonly found as an inverted repeat with a characteristic core region, most often being 105 or 153 bp in length and present in over 100 copies in the genome [35]. Although it has not been determined how the CREE is mobilized, it is present in different locations between the genomes and the target site is TA, which is copied upon insertion [35]. The insertion of a CREE can generate a promoter at either its 5' [36] or 3' end [37] and can disrupt the sequence into which it has been inserted. In this case the CREE disrupts the 3' end of the AzlC-related gene that flanks the MME, generating a termination codon for this coding region within one of the CREE. Sequencing from strain FA19 revealed only one CREE in this region, although the AzlC-related gene was still disrupted at the 3' end by divergent sequence (Accession AY386269). Therefore the absence of XNG0438 in strain FA19 has been confirmed by sequencing and an additional variation in the intergenic region of this MME has been identified.

In N. gonorrhoeae strain FA1090, the MME flanked by uvrA and a hypothetical gene contains another hypothetical gene and a cytosine-specific methylase gene (Figure 6). In N. meningitidis strains Z2491 and FAM18, there are three and two hypothetical genes in this location, respectively. The region 3' of uvrA is also the site of chromosomal rearrangement in N. meningitidis strain MC58, relative to the other strains. DNA from strain FA19 does not hybridize to any of these genes within the uvrA-associated MME. PCR and sequencing of this region from N. gonorrhoeae strain FA19 revealed one new coding sequence within this MME (Figure 6; Accession AY386270). This sequence has no homology to complete coding sequences within the public databases, although the first 137 bp (100% DNA identity) of this CDS, 3' of uvrA, is included as unannotated sequence at the end of the GenBank entry for uvrA from library clone pJKD909, derived from N. gonorrhoeae (U134760).

The hybridization of the probe for N. meningitidis strain Z2491 CDS NMA2121, suggested that this gene is present in strain FA19 in an MME flanked by two hypothetical genes (Figure 7). Annotated as a hypothetical protein, this CDS is not present in the other two neisserial genome
sequences, nor in the recently completed *N. meningitidis* strain FA19 genome sequence. This MME was sequenced from *N. gonorrhoeae* strain FA19 to determine if a complete version of NMA2121 was present in this location (Accession AY386271). Sequencing demonstrated that although this gonococcal strain does contain a CDS with homology to the sequence of the meningococcal CDS NMA2121, it is disrupted. A 107 bp CREE is present 44 bp into the CDS, which introduces a termination codon shortly thereafter and generates a frame-shift (Figure 7). The presence of this CREE also generates a potential promoter at the 3' end of the element, similar to those described previously and identified 5' of *uvrB* [37]. This, coupled with a potential ribosome binding site and secondary translational initiation codon, suggests that the NMA2121 homologue may be expressed in *N. gonorrhoeae* strain FA19, despite the apparent interruption of the gene by a CREE.

### Discussion

The Pan-*Neisseria* microarray, coupled with MME analysis of sites of divergence indicated by the microarray analysis, has revealed five new neisserial genes not represented on the microarray, as well as identifying the 59 genes of the GGI in *N. gonorrhoeae* strain FA19. There are also 36 genes that are present within strain FA1090 that are not universally represented in the other main experimental strains, 16 of which are located within MMEs that are identifiable from comparison of the complete genome sequences of gonococcal strain FA1090 and the available complete meningococcal genome sequences. The ability to identify this range of differences is in part due to the unique nature of the Pan-*Neisseria* microarray (Saunders, Davies, et al. – in preparation). Its design incorporates probes to the most conserved regions of genes, having been evaluated in preparation). Its design incorporates probes to the most conserved regions of genes, having been evaluated through the comparison of *N. gonorrhoeae* strain FA1090 sequences [16] with those of *N. meningitidis* strains MC58 [29] and Z2491 [30]. Genes from the two meningococcal genomes that were not optimally probed by the gonococcal probes, had additional probes added to the microarray. The 59 predicted coding sequences of the Gonococcal Genetic Island, as sequenced from *N. gonorrhoeae* strain MS11 [21], are also represented on the microarray. This microarray composition meant that for the three gonococcal laboratory strains evaluated here, it was possible to determine, solely from the microarray results, that 36 strain FA1090 genes are absent from the other gonococcal strains, while 130 genes are present in these strains that are absent from strain FA1090. This second group of genes could not have been identified with a single-genome microarray based upon strain FA1090 alone, emphasizing the strength of this multi-species, multi-genome design strategy. The microarray results additionally indicated regions of the chromosomes likely

---

**Figure 6**

The MME flanked by *uvrA* and a hypothetical gene. This MME contains a gene encoding a putative methylase and a hypothetical gene in *N. gonorrhoeae* strain FA1090 [16], three hypothetical genes in *N. meningitidis* strain Z2491 [30], two hypothetical genes in *N. meningitidis* strain FAM18, and one new neisserial gene of unknown function in *N. gonorrhoeae* strain FA19 (AY386270). This is the point of a chromosomal rearrangement in *N. meningitidis* strain MC58 [29].

**Figure 7**

The MME flanked by two hypothetical genes. This MME contains an intergenic sequence in *N. gonorrhoeae* strain FA1090 [16], and a hypothetical gene in *N. meningitidis* strain Z2491 [30]. In *N. gonorrhoeae* strain FA19 (AY386271) there is a homologue of the meningococcal CDS, which has been disrupted by a CREE 44 bp into the gene. The CREE in this location introduces a termination codon and generates a frame-shift. However, it also generates a potential promoter 5' of a potential ribosomal binding site and secondary initiation codon internal to NMA2121. This is the point of a chromosomal rearrangement in *N. meningitidis* strains MC58 [29] and FAM18. Similar genes are indicated in the same colour.
to contain divergent sequence, analysis of which revealed five new genes.

In strain MS11, 17 probes designed to strain FA1090 genes did not hybridize, including nine genes associated with five MMEs. Sequencing of two of these MMEs confirmed the absence of the strain FA1090 genes and identified four new genes in strain MS11. One of the five MMEs is the Gonococcal Genetic Island, the sequence of which is known from strain MS11. The remaining two MMEs (hypo-aspA and ribAB-hypo) may simply be too large to amplify using the techniques described here, but may therefore represent the location of additional genes present in strain MS11 and absent in FA1090, and vice versa. There are 19 strain FA1090 probes that do not hybridize to strain FA19 DNA, with seven of these being contained in five MMEs, one of which is the Gonococcal Genetic Island. Sequencing of three of the MMEs confirmed the absence of the strain FA1090 genes and identified one new neisserial gene in strain FA19. The remaining MME (hypo-aspA) was also identified in strain MS11. No differences were observed between N. gonorrhoeae strains FA1090 and F62. Discounting those genes that are likely to display differences due to divergence (pitC, lbpB, tspB, maf-associated, prophage-associated, and the NMA1360 gene fragment) a total of 80 genes were found to contribute to inter-strain gene-complement differences, of which 59 comprise the Gonococcal Genetic Island and one represents the hypothetical strain FA1090 gene that is present in its place (XNG0741; Figure 1). This does not take into account those genes that may be found in strains MS11 and FA19 in the large hypo-aspA MME and in strain MS11 in the large ribAB-hypo MME.

Although this approach to strain comparison, utilizing a multi-species, multi-genome microarray is novel, powerful, and has identified differences between these strains, there are some additional forms of divergence that this approach cannot address. This means that although we have an indication of the significant similarity of strains FA1090 and F62, which will facilitate the use of the available genome sequence in this strain, they cannot be assumed to be identical. There may be other differences between these and the other strains such as: (a) the presence of additional genes that are not currently on the microarray, although these cannot be associated with the absence of MME-associated genes as they are in strains MS11 and FA19; (b) divergence outside of the probe regions used for construction of the microarray, which were selected due to their maximal conservation across the species (≥90% ID over a probe of 150–450 bp), therefore the divergence in the fbp iron acquisition genes (strain F62 sequence: U33937) is not detected because the probes hybridize to the conserved regions of the genes at ≥98% over the length of the probe; (c) differences in transcriptional regulation, post-transcriptional regulation, and phase variation, for example as in pgtA, which is phase variable in strain FA1090 but not in strain F62 where the gene lacks the phase variation-mediating poly-G tract [38]. Despite these caveats this study clearly shows some of the significant advantages and power of this microarray design approach in its application to genotyping, and provides a new basis upon which to interrelate the data obtained from the many studies performed using these strains.

**Conclusions**

The multi-strain, multi-species microarray design strategy can be used to identify the presence and absence of genes in a more informative fashion than arrays based upon single genomes, or even single species. This approach can be used to identify those genes that are trans-species, rather than species associated, as demonstrated from limited complete genome sequenced examples. This approach is also able to identify chromosomal discontinuities which when coupled to MME analysis can identify new sequences and genes not present in any of the sequences used for microarray design. This study identified many differences, which are likely to account for at least some of their recorded differences in behaviour between the gonococcal experimental strains. This included the identification of five new neisserial genes. The recently completed N. meningitidis serogroup C strain FAM18 complete genome sequence ftp://ftp.sanger.ac.uk/pub/pathogens/nm1 identified 35 new neisserial genes, other than those associated with a large prophage region. This array-based approach is therefore a comparatively cheap and efficient way of finding new genes within a species for which existing genomes are available. The range of differences that have been identified in this study emphasize the utility of a pan-species microarray, with the design characteristics used, in the investigation of diverse bacterial populations.

**Methods**

**Bacterial strains and growth conditions**

N. gonorrhoeae strains FA1090, F62, MS11, and FA19, were propagated on GC agar (Difco Laboratories) containing the Kellogg supplement and ferric nitrate [39] at 37°C under 5% (v/v) CO₂. These strains were kindly supplied by Prof. William M. Shafer (FA1090, F62, FA19) and Prof. Richard F. Rest (FA1090, F62, MS11).

**DNA extraction, labelling, and microarray hybridization**

Chromosomal DNA was extracted using the Aquapure Genomic DNA Isolation Kit (BIORAD). FluoroLink™ Cy3-dCTP and FluoroLink™ Cy5-dCTP (Amersham Pharmacia Biotech) were incorporated into 10 μg of chromosomal DNA using random hexamer primers (Invitrogen) and DNA polymerase I, Klenow fragment (Bioline, UK). Labelled DNA:DNA probe microarray hybridizations
were conducted in 4 × SSC, 0.2875% SDS at 65°C overnight. The pan-Neisseria microarray was designed, generated and printed as will be described fully elsewhere (Saunders, Davies, et al. – in preparation). This microarray design contains 2704 probes directed to the most conserved regions of the 6294 annotated potential coding sequences from N. gonorrhoeae strain FA1090 [16], N. meningitidis strains MC58 [29] and Z2491 [30], and additional sequence from N. gonorrhoeae strain MS11 including the Gonococcal Genetic Island [21]. Each probe of 150–450 bp was selected to target within the most conserved regions of the 6294 annotated potential coding sequences from N. gonorrhoeae strain FA1090 [16], N. meningitidis strains MC58 [29] and Z2491 [30], and additional sequence from N. gonorrhoeae strain MS11 including the Gonococcal Genetic Island [21]. Each probe of 150–450 bp was selected to target within the most conserved (≥ 90% identical) sequence region of each coding sequence between the genomes against which it was designed. In this way, the pan-Neisseria microarray achieves maximal hybridization to orthologous genes in different strains and species. Each probe was also assessed for cross-hybridization within the same genome. If any significant cross-hybridization was detected, being assessed through sequence homology of 70% over 40 bp, the probe was redesigned to exclude the potential for cross-hybridization with other genes. In this way, the pan-Neisseria microarray eliminates false positive results due to the hybridization to non-representative probes. As an example of the design detail and probe hybridization fidelity, the pilC gene from N. gonorrhoeae strain FA1090 XNG0049 does not hybridize to either of the pilC probes designed against N. meningitidis strain MC58 pilC genes NMBo049 and NMBo1847. While the gonococcal gene is 78.8% (over 3300 bp) and 82.3% (over 3293) identical, with respect to each of the meningococcal genes, the probes for the meningococcal genes fall below the 70% over 40 bp cross-hybridization design threshold and no signal from these probes is detected in control experiments when the microarray is hybridized with N. gonorrhoeae strain FA1090 DNA. Microarray slides were scanned on an Axon GenPix 4000 B scanner using GenPix Pro v 3.0.6.90 and analyzed using GACK [40] downloaded from http://falkow.stanford.edu/whatwedo/software.

PCR amplification and sequencing
PCR amplification from chromosomal DNA was performed using BIO-X-ACT™ Long DNA Polymerase (BIO-LINE) or Taq DNA polymerase (Invitrogen) according to the manufacturers’ instructions. Primers used are listed in additional file 1 Table S1. PCR amplification of two regions, hypo-aspA and ribAB-hypo, failed even after alternate primers were designed. Automated sequencing used ABI Prism® BigDye™ Terminator Cycle Sequencing version 3.0 (Applied Biosystems) and was resolved on an ABI Prism® 3100 DNA Sequencer (Applied Biosystems).

Bioinformatics
ACEDB [41] was used to analyze the complete genome sequences of the Neisseria spp., as described previously [42]. MMEs were identified through comparison of the available genomes in the regions of the chromosome indicated as divergent from the microarray data. Sequence traces were analyzed using Trev [43]. Sequences were aligned and assembled using the Wisconsin Package from GCG (Accelrys). Homology searches were performed using BLAST [44] against the EMBL databases, accessed through the Oxford University Bioinformatics Centre.

Accession numbers for sequence data
The sequence of the MME region containing four new gonococcal genes between nuoL and nuoM from N. gonorrhoeae strain MS11, accession number AY386266, and that containing one new gonococcal gene adjacent to wrA from N. gonorrhoeae strain FA19, accession number AY386270, have been submitted to GenBank. Additional MME region sequences submitted to GenBank are: that adjacent to pglB from N. gonorrhoeae strain MS11 (AY386267); between nadC and xthA from N. gonorrhoeae strain FA19 (AY386268); between a type IV pilin-related genes and an AzIC-related protein gene from N. gonorrhoeae strain FA19 (AY386269); and between two hypothetical protein genes from N. gonorrhoeae strain FA19 (AY386271).

Authors’ contributions
LS is responsible for all experimental manipulations, analysis of microarray and sequencing data, identification of MMEs, writing of the manuscript, and assisted in probe design for the pan-Neisseria microarray. JD is responsible for the production of the pan-Neisseria microarray slides used in this study and provided support in the preparation of the manuscript. NS conceived of the pan-Neisseria microarray and its use in this kind of study, was the primary designer of the probes for the microarray, supervised this work, and provided support in the preparation of this manuscript.

Additional material

Additional File 1
Table S1. Oligonucleotide primers used in this study. A Table of the oligonucleotide primers used in this study for the amplification and sequencing of the MMEs presented from N. gonorrhoeae strain MS11 and N. gonorrhoeae strain FA19.

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2164-5-23-S1.pdf]

Acknowledgements
The authors would like to thank Dr. Sarah Butcher for continued bioinformatics support and Julian Robinson from the Sir William Dunn School of Pathology sequencing service. NJS was supported by a Wellcome Advanced Research Fellowship. LASS is supported by a Wellcome Trust Project.
Grant. JKD is the recipient of a Project Grant from the Canadian National Health and Medical Research council. The meningococcal genome sequence data were produced by the N. meningitidis serogroup C strain FA118 Sequencing Group at the Wellcome Trust Sanger Institute, and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/nm/. The N. gonorrhoeae genome sequence was obtained from the University of Oklahoma, http://www.genome.uiuc.edu/ongo.html, the Gonorococal Genome Sequencing Project which was supported by USPHS/NIH grant #AI-33899, GenBank accession number AE004969 [16].

References
1. Spratt BG, Smith NH, Zhou J, O'Rourke M, Feil E: The population genetics of the pathogenic Neisseria. Society for General Microbiologists Volume: 17, Heath AB, Bygraves JA, Maiden MC, Ringrose JWP, Wellington EHM and Saunders Jr. SGM Press, 1995:143-160.
2. Carbonetti NH, Simnad VI, Seifert HS, So M, Sparling PF: Genetics of protein I of Neisseria gonorrhoeae: construction of hybrid porins. Proc Nat Acad Sci U S A 1988, 85:6841-6845.
3. O'Rourke M, Spratt BG: Further evidence for the non-clonal population genetics of Neisseria gonorrhoeae: extensive genetic diversity within isolates of the same serogroup type. Microbiology 1994, 140 (Pt 6):1285-1290.
4. De La Fuente L, Vazquez JA: Genetic structure of Neisseria gonorrhoeae populations: a non-clonal pathogen. J Gen Microbiol 1993, 139 (Pt 11):2603-2611.
5. O'Rourke M, Spratt BG: Correlation of cellular activity with the presence of a strain-specific transposable element. Antimicrob Agents Chemother 1997, 41:2457-2462.
6. Saunders NJ, Snyder LAS: Genetic structure of Neisseria gonorrhoeae: pilC1 and pilC2 are each part of a family of comparable size. Mol Microbiol 2002, 40:1067-1077.
7. Connell TD, Black WJ, Kawula TH, Barritt DS, Dempsey JA, Kvernes NM, Spratt BG: An analysis of sequential homologous repeats in Neisseria meningitidis and Neisseria gonorrhoeae. Mol Microbiol 2000, 37:1853-1868.
8. Joiner KA, Puentes SM, Warren KA, Scales RA, Judd RC: Complete genome sequence of Neisseria gonorrhoeae: pilC1 and pilC2 are each part of a larger duplication of the genome of Neisseria gonorrhoeae that is composed of two variable genetic islands. J Bacteriol 1998, 180:6274-6286.
9. Nachamkin I, Cannon JG, Mittler RS: Monoclonal antibodies against Neisseria gonorrhoeae: production of antibodies directed against a strain-specific cell surface antigen. Infect Immun 1981, 32:641-648.
10. Bihrmaier A, Romling U, Meyer TF, Tunneller B, Gibbs CP: Physical and genetic map of the Neisseria gonorrhoeae strain MS11-N198 chromosome. Mol Microbiol 1991, 5:2529-2539.
11. Dillard JP, Seifert HS: A variable genetic island specific for Neisseria gonorrhoeae is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. Mol Microbiol 2001, 41:263-277.
12. Swanson J: Studies on gonococcal infection. IV. PilC: their role in attachment of gonococci to tissue culture cells. J Exp Med 1973, 137:571-589.
13. La Scolea LJ, Dul MJ, Young FE: Stability of pathogenic colony types of Neisseria gonorrhoeae in liquid culture by using the parameters of colony morphology and desoxyribonucleic acid. J General Microbiol 1975, 119:165-170.
14. Daly JA, Lee TJ, Spitznagel JK, Sparling PF: Gonococci with mutations to low-level penicillin resistance exhibit increased sensitivity to the oxygen-independent bactericidal activity of human polymorphonuclear leukocyte granule extracts. Infect Immun 1992, 52:826-832.
15. Jarvis GA, Li J, Swanson KV: Invasion of human mucosal epithelial cells by Neisseria gonorrhoeae upregulates expression of intercellular adhesion molecule 1 (ICAM-1). Infect Immun 1999, 67:1149-1156.
16. Xu XD, Harwig SS, Shafer WM, Lehrer RI: Protegrin structure and activity against Neisseria gonorrhoeae. Infect Immun 1997, 65:636-639.
17. Joiner KA, Puentes SM, Warren KA, Scales RA, Judd RC: Complete binding on serum-sensitive and serum-resistant transformants of Neisseria gonorrhoeae: effect of presensitization with a non-bacterial monoclonal antibody. Microb Pathog 1989, 6:343-350.
18. Olesky M, Hobbs M, Nicholas RA: Identification and analysis of amino acid mutations in porin IB that mediate intermediate-level resistance to penicillin and tetracycline in Neisseria gonorrhoeae. Antimicrob Agents Chemother 2002, 46:2811-2820.
19. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Gwinn ML, Jones M, Skelton J, Whitehead S, Mahillon J, Achtman M, Fraser CM, Moxon ER, Rappuoli R, Venter JC: Complete genome sequence of Neisseria meningitidis. Nature 2000, 404:529-536.
20. Parkhill J, Achtman M, Jones MR, Parkhill N, Squizzato S, Berriman M, Durbin R, Fraser CM, Rappuoli R, Barrell BG, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parkay DS, Blair E, Cottone T, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Yamashita J, Gill J, Sciarato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 2000, 287:1809-1815.
21. Thrall D, Tettelin H, Parkhill J, Jones MR, Parkhill N, Berriman M, Durbin R, Fraser CM, Rappuoli R, Barrell BG, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parkay DS, Blair E, Cottone T, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Yamashita J, Gill J, Sciarato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 2000, 287:1809-1815.
22. Thrall D, Tettelin H, Parkhill J, Jones MR, Parkhill N, Berriman M, Durbin R, Fraser CM, Rappuoli R, Barrell BG, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parkay DS, Blair E, Cottone T, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Yamashita J, Gill J, Sciarato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 2000, 287:1809-1815.
23. Thrall D, Tettelin H, Parkhill J, Jones MR, Parkhill N, Berriman M, Durbin R, Fraser CM, Rappuoli R, Barrell BG, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parkay DS, Blair E, Cottone T, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Yamashita J, Gill J, Sciarato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 2000, 287:1809-1815.
36. Snyder LAS, Shafer WM, Saunders NJ: Divergence and transcriptional analysis of the division cell wall (dcw) gene cluster in Neisseria spp. Mol Microbiol 2003, 47:431-442.
37. Black CG, Fyfe JA, Davies JK: A promoter associated with the neisserial repeat can be used to transcribe the uvrB gene from Neisseria gonorrhoeae. J Bacteriol 1995, 177:1952-1958.
38. Banerjee A, Wang R, Supernavage SL, Ghosh SK, Parker J, Ganesh NF, Wang PG, Gulati S, Rice PA: Implications of phase variation of a gene (pgtA) encoding a pilin galactosyl transferase in gonococcal pathogenesis. J Exp Med 2002, 196:147-162.
39. Kellogg D.S., Jr., Peacock W.L., Jr., Deacon WE, Brown L, Pirkle CI: Neisseria gonorrhoeae. I. Virulence genetically linked to clonal variation. Journal of Bacteriology 1963, 85:1274-1279.
40. Kim CC, Joyce EA, Chan K, Falkow S: Improved analytical methods for microarray-based genome-composition analysis. Genome Biol 2002, 3:RESEARCH0065.
41. Durbin R, Thierry-Mieg JT: A C. elegans DataBase. Documentation, code and data available from http://www.acedb.org, 1991.
42. Snyder LA, Butcher SA, Saunders NJ: Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic Neisseria spp. Microbiology 2001, 147:2321-2332.
43. Bonfield JK, Beal KF, Betts MJ, Staden R: Trev: a DNA trace editor and viewer. Bioinformatics 2002, 18:194-195.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403-410.