Genetic mechanisms of *opa* gene variation in *Neisseria gonorrhoeae*

Nicole Bilek

Imperial College London

Department of Infectious Disease Epidemiology

Thesis submitted for the degree of Doctor of Philosophy of Imperial College London

I, Nicole Bilek, declare that this dissertation is my own original work and that all else has been referenced to the best of my ability.

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed."

Albert Einstein - Nobel Prize for Physics in 1921. 1879-1955.

ABSTRACT

To understand the rates and mechanisms of *Neisseria gonorrhoeae* opacity (*opa*) gene variation, the 11 *opa* genes were amplified independently so that an *opa* allelic profile could be defined for any isolate from the sequences at each locus.

Initial examination of 14 unrelated gonococcal isolates showed that no *opa* alleles were shared. Analysis of closely related isolates showed these typically shared most *opa* alleles and so the mechanisms by which recent changes occurred at individual *opa* loci could be determined. The great majority of changes were due to recombination among existing alleles that either duplicated an *opa* allele present at another locus or resulted in a mosaic of existing *opa* alleles. Single nucleotide changes or the insertion/deletion of a single codon also occurred, but few of these events were assigned to mutation, the majority being assigned to localised recombination. Introduction of novel *opa* genes from co-infecting strains was rare and all but one occurred in the same sexual network. Changes at the eleventh *opa* locus (*opa11*) occurred much more rapidly than at other *opa* loci, almost always differing even between recent sexual contacts. Examination of the neighbouring *pilE* gene showed that changes at *opa11* and *pilE* often occurred together, although this linkage may not be a causal one.

The Opa protein sequences encoded by the *opa* genes were determined and the regions spanning the two hyper-variable regions (HVR1 and HVR2) were analysed. An almost equal number of Opa protein and *opa* nucleotide sequences were detected but there was a limited number of HVR combinations, indicating that a large

proportion of differences are located elsewhere in the protein. Very few novel HVRs were generated by recombination, even at *opa11* where the rate of variation was greatest. Most changes resulted in re-assortment of existing HVRs, most likely due to protein function restriction/benefits.

ACKNOWLEDGEMENTS

My thanks go first and foremost to my supervisors, Prof. Brian G. Spratt and Prof. Cathy A. Ison, not only for the opportunity to do this work but also all their expertises and guidance, their support and encouragement. There were many days when my morale was low but unknowing to them, their contributions lifted my spirit and kept me going.

A very special thank you to my husband Tim (and my son Lukas too) who put up with me, never expecting a clean house or cooked meal on the table but instead provided me with endless amounts love, cuddles, cups of tea and a guilt free environment to finish up my thesis. I'd like to give a word of thanks also to my parents for giving me the opportunity to go to university, a novelty in our family.

Also worthy of my immense gratitude are my friends and colleagues, Daniel Henk, David Aanensen, William Hanage, Daniel Godoy, other members of the laboratory and, in particular, Cynthia Bishop, who supported me patiently over the years, listening to me and who undoubtedly lost many hours of their life to gonorrhoea and *opa* genes.

A note in remembrance of a few special past members of the department with whom I hold special memories of my own, including Steve Dunstan, Klaus Kurtenbach and Richard Suswillo. Especially though to Janie Pryce-Miller, a very good and dear friend I wish had been here to see the day I finished this thesis and could share in a glass of bubbly with me.

CONTENTS

TITLE PAGE.....1

DECLARATION OF ORIGINALITY.....2

ABSTRACT.....3

ACKNOWLEDGEMENTS.....5

List of tables.....10

List of figures.....13

List of abbreviations.....16

Chapter 1: Introduction.....18

1.1 Description of *Neisseria gonorrhoeae*.....18

1.2 Epidemiology of gonorrhoea.....20

1.3 Diagnosis.....22

1.4 Clinical manifestation.....23

1.5 Pathogenesis.....25

1.6 Treatment and antimicrobial resistance.....29

1.7 Gonococcal genetics.....32

1.8 Strain typing.....35

1.9 Project motivation and objectives.....39

Chapter 2: Methods.....48

2.1 Bacterial culture.....48

2.2 DNA preparation.....48

2.3 Amplification of *opa* genes.....48

2.4 Sequencing of each *opa* locus.....49

2.5 NG-MAST and MLST.....51

Chapter 3: Amplification and sequencing of the 11 gonococcal *opa* genes.....52

3.1 Introduction……**52**

3.2 Methods……**54**

3.3 Results……**58**

3.3.1 Preliminary sequencing data……**58**

3.3.2 *opa* gene sequences amongst diverse isolates……**59**

3.4 Discussion……**62**

Chapter 4: Mechanisms of *opa* gene variation amongst closely related isolates in Sheffield……**65**

4.1 Introduction……**65**

4.2 Methods……**67**

4.2.1 Isolates……**67**

4.2.2 *In silico opa*-typing……**67**

4.3 Results……**69**

4.3.1 *opa* gene profiles of isolates with the same sequence type……**69**

4.3.2 *In silico opa*-typing of ST12/*opa*-type 1 and ST261/*opa*-type 4 isolates**.....73**

4.3.3 *opa* gene profiles of isolates from known sexual contacts……**73**

4.3.4 Assigning allelic change in *opa* genes to recombination or point mutation**.....79**

4.4 Discussion……**83**

Chapter 5: Mechanisms of *opa* gene variation amongst closely related isolates in London……**86**

5.1 Introduction……**...86**

5.2 Methods……**87**

5.2.1 Isolates……...……**87**

5.3 Results……..……**88**

5.3.1 General observations…….……**88**

5.3.2 Variations observed in each London network……...……**89**

5.4 Discussion……**96**

Chapter 6: The relationship between variation at *opa11* and at the neighbouring *pilE* gene……**98**

6.1 Introduction……**98**

6.2 Methods……....……**101**

6.2.1 Isolates…….……**101**

6.2.2 *pilE* sequencing……...……**101**

6.3 Results……**103**

6.3.1 Nature of the pilin genes neighbouring the *opa11* locus….……**103**

6.3.2 Variation at *opa11* and at the neighbouring *pilE* gene……**106**

6.4 Discussion……..…**107**

Chapter 7: Stability of *opa* allelic profiles, *pilE* sequences and coding repeats during *in vitro* sub-culturing…….....……**109**

7.1 Introduction……..……**109**

7.2 Methods…….……**110**

7.3 Results……**111**

7.3.1 Changes in *opa* allelic profiles during sub-culture**.....111**

7.3.2 Changes in *pilE* sequences during sub-culture……**111**

7.3.3 Changes in the multiple of CRs during sub-culture……**111**

7.4 Discussion……..……**114**

Chapter 8: Analysis of the hyper-variable regions of Opa proteins from closely related isolates……..…….**116**

8.1 Introduction……..……**116**

8.2 Methods…….……**118**

8.2.1 Isolates…….……**118**

8.2.2 Sequence translation and trimming…….……**118**

8.2.3 HVR matrices…….....**119**

8.2.4 Dendrograms of HVRs**.....119**

8.3 Results……**120**

8.3.1 Amino acid sequences of Opa proteins and their HVRs**.....120**

8.3.2 HVRs within populations**.....120**

8.3.3 Differences in HVRs between sexual contacts and within sexual networks**.....125**

8.3.4 Amino acid sequences of shared HVRs**.....135**

8.3.5 Sharing of HVRs between populations**.....150**

8.4 Discussion.....**152**

Chapter 9: Final discussion.....**157**

Bibliography…….**.....**…….**.....171**

Appendices:

APPENDIX 1: Primer pairs used to amplify each *opa* locus and annealing temperatures…….……**187**

APPENDIX 2: Relevant publications**.....188**

List of tables

Table 3.1 *opa* gene profiles of 14 unrelated isolates of *N. gonorrhoeae*…….....…**60**

Table 3.2 p-values as a measure of diversity for the *opa* alleles of each diverse isolate….....…….....……**61**

Table 4.1a o*pa* gene profiles of 21 ST12/*opa*-type 1 isolates chosen to be representative of the sampling period…….……**71**

Table 4.1b *opa* gene profiles of 27 ST261/*opa*-type 4 isolates chosen to be representative of the sampling period…….……**72**

Table 4.2a *In silico opa*-typing fragments generated with *Taq*I for the 17 *opa* repertoires observed amongst the 21 ST12/*opa*-type 1 isolates...….....……**75**

Table 4.2b *In silico opa*-typing fragments generated with *Taq*I for the 18 *opa* repertoires observed amongst the 27 ST261/*opa*-type 4 isolates……..….....……**76**

Table 4.3a o*pa* gene profiles of ST12 isolates from recent sexual contacts.....**77**

Table 4.3b o*pa* gene profiles of ST261 isolates from recent sexual contacts…...…**78**

Table 5.3 *opa* gene profiles of isolates from individuals in London recently infected with a strain with the same genotype by NG-MAST.....**91**

Table 5.3a MSM cluster ST802.....**91**

Table 5.3b MSM cluster ST825.....**91**

Table 5.3c MSM cluster ST876.....**91**

Table 5.3d Heterosexual cluster ST610.....**92**

Table 5.3e Heterosexual cluster ST709.....**92**

Table 5.3f Heterosexual cluster ST777.....**92**

Table 5.3g Mixed sexual orientation cluster ST584.....**92**

Table 6.1a o*pa11* allele and nature of *pilE* gene of ST12 isolates from recent sexual contacts…….....……**104**

Table 6.1b *opa11* allele and nature of *pilE* gene of ST261 isolates from recent sexual contacts……..……**105**

Table 7.1 *opa* profiles of isolates sub-cultured ten times from single colonies…...**113**

Table 8.1a HVR1-HVR2 amino acid combinations of alleles from the 21 ST12/*opa*-type 1 isolates chosen to be representative of the sampling period……..……**123**

Table 8.1b HVR1-HVR2 amino acid combinations of alleles from the 27 ST261/*opa*-type 4 isolates chosen to be representative of the sampling period.....**124**

Table 8.2a HVR1-HVR2 amino acid combinations of alleles from the ST12 isolates from recent sexual contacts…….....……**126**

Table 8.2b HVR1-HVR2 amino acid combinations of alleles from the ST261 isolates from recent sexual contacts…….....……**127**

Table 8.3 HVR1-HVR2 amino acid combinations of alleles from isolates from individuals in London recently infected with the same genotype by NG-MAST…….....**133**

Table 8.3a MSM cluster ST802……...……**133**

Table 8.3b MSM cluster ST825……..…**133**

Table 8.3c MSM cluster ST876……...……**133**

Table 8.3d Heterosexual cluster ST610……...……**134**

Table 8.3e Heterosexual cluster ST709……...……**134**

Table 8.3f Heterosexual cluster ST777……....……**134**

Table 8.3g Mixed sexual orientation cluster ST584…….....…**134**

Table 8.4 Matrix of HVR1 and HVR2 amino acid sequences of all ST12 alleles.....**137**

Table 8.5 Matrix of HVR1 and HVR2 amino acid sequences of all ST261 alleles.....**138**

Table 8.6 Matrix of HVR1 and HVR2 amino acid sequences of all ST802 alleles.....**140**

Table 8.7 Matrix of HVR1 and HVR2 amino acid sequences of all ST825 alleles.....**141**

Table 8.8 Matrix of HVR1 and HVR2 amino acid sequences of all ST876 alleles.....**142**

Table 8.9 Matrix of HVR1 and HVR2 amino acid sequences of all ST610 alleles.....**143**

Table 8.10 Matrix of HVR1 and HVR2 amino acid sequences of all ST709 alleles….....**145**

Table 8.11 Matrix of HVR1 and HVR2 amino acid sequences of all ST777 alleles.....**147**

Table 8.12 Matrix of HVR1 and HVR2 amino acid sequences of all ST584 alleles….....**148**

Table 8.13 Matrix of all HVR1 and HVR2 amino acid sequences in the Sheffield and London networks.....**151**

List of figures

Figure 1.1 Number of diagnoses of gonorrhoea by sex, GUM clinics, England and Wales: 1925 – 2008.....**20**

Figure 1.2 Gonorrhoea in the news.....**31**

Figure 1.3 Diagrammatic representation of an *opa* gene and the encoded protein...**40**

Figure 3.1 Diagrammatic representation of the single contiguous genome sequence of *N. gonorrhoeae* strain FA1090…….....…**54**

Figure 3.2 Diagrammatic representation of an *opa* gene showing the location and orientation of the sequencing primers on each strand…….....……**57**

Figure 6.1 Region amplified during *opa11* PCR which also amplifies *pilE* and *pilS* loci……...**100**

Figure 8.1 Diagrammatic representation of the recombination event in pair 14.....**129**

Figure 8.2 Diagrammatic representation of the recombination event in pair 17.....**130**

Figure 8.3 Diagrammatic representation of the recombination event in pair 18.…**131**

Figure 8.4a Dendrogram of ST12 HVR1 amino acid sequences based on the number of amino acid differences…….....**....137**

Figure 8.4b Dendrogram of ST12 HVR2 amino acid sequences based on the number of amino acid differences…….....…**137**

Figure 8.5a Dendrogram of ST261 HVR1 amino acid sequences based on the number of amino acid differences…….....…**138**

Figure 8.5b Dendrogram of ST261 HVR2 amino acid sequences based on the number of amino acid differences……**139**

Figure 8.6a Dendrogram of ST802 HVR1 amino acid sequences based on the number of amino acid differences……**140**

Figure 8.6b Dendrogram of ST802 HVR2 amino acid sequences based on the number of amino acid differences……**140**

Figure 8.7a Dendrogram of ST825 HVR1 amino acid sequences based on the number of amino acid differences……**141**

Figure 8.7b Dendrogram of ST825 HVR2 amino acid sequences based on the number of amino acid differences……**141**

Figure 8.8a Dendrogram of ST876 HVR1 amino acid sequences based on the number of amino acid differences……**142**

Figure 8.8b Dendrogram of ST876 HVR2 amino acid sequences based on the number of amino acid differences……**....**……**142**

Figure 8.9a Dendrogram of ST610 HVR1 amino acid sequences based on the number of amino acid differences……**143**

Figure 8.9b Dendrogram of ST610 HVR2 amino acid sequences based on the number of amino acid differences……**....**……**144**

Figure 8.10a Dendrogram of ST709 HVR1 amino acid sequences based on the number of amino acid differences……**145**

Figure 8.10b Dendrogram of ST709 HVR2 amino acid sequences based on the number of amino acid differences……**146**

Figure 8.11a Dendrogram of ST777 HVR1 amino acid sequences based on the number of amino acid differences……**147**

Figure 8.11b Dendrogram of ST777 HVR2 amino acid sequences based on the number of amino acid differences……**147**

Figure 8.12a Dendrogram of ST584 HVR1 amino acid sequences based on the number of amino acid differences……**148**

Figure 8.12b Dendrogram of ST584 HVR2 amino acid sequences based on the number of amino acid differences……**149**

List of abbreviations

°C – degree Celsius

A/S – auxotype/serotype

bp – base pairs

CEA – carcinoembryonic antigen

CEACAM – carcinoembryonic antigen cell adhesion molecule

CR – coding repeat (CTCTT)

DGI – disseminated gonococcal infection

GUM – genitourinary medicine

HSPG – heparan sulfate proteoglycan

HVR – hyper-variable region

HVR1 – hyper-variable region 1

HVR2 – hyper-variable region 2

id – identity

Ig(A/G/M) – immunoglobulin A/G/M

indel – insertion/deletion

kb – kilo base pairs

LOS – lipooligosaccharide

min – minutes

MLEE – multi-locus enzyme electrophoresis

MLST – multi-locus sequence typing

MSM – men who have sex with men

NAAT – nucleic acid amplification test

NG-MAST – *N. gonorrhoeae* multi-antigen sequence typing

NS – non-synonymous

OMP – outer membrane protein

Opa – opacity

Opa^- – Opa non-expressing

Opa^+ – Opa expressing

PCR – polymerase chain reaction

PID – pelvic inflammatory disease

PMN – polymorphonuclear neutrophil/leukocyte

sec – seconds

ST – sequence type

STI – sexually transmitted infection

SVR – semi-variable region

µl – microlitres

x g – times gravity

Chapter 1: Introduction

1.1 Description of *Neisseria gonorrhoeae*

N. gonorrhoeae is the causative agent of the sexually transmitted infection (STI) gonorrhoea and is one of two pathogenic species belonging to the genus *Neisseria*. Apart from *N. gonorrhoeae* and *N. meningitidis*, most members of the genus are commensal species of humans (or animals) which are rarely involved in human disease amongst immuno-competent individuals (115, 152). *N. meningitidis*, usually acquired through the inhalation of respiratory droplets, colonises the nasopharynx but occasionally gains access to the blood leading to septicaemia and/or meningitis (115). *N. gonorrhoeae* is an obligate human pathogen and transmission usually occurs through direct sexual contact but indirect modes of transmission have been reported (142, 152). It is therefore usually isolated from the urinogenital tract, namely the cervix and the urethra, but may also be found at other mucosal surfaces such as the nasopharynx, rectum and eyes (152). During colonisation in the nasopharynx, it may at times co-exist with other human *Neisseria* species.

Studies have described considerable DNA sequence similarity between members of the genus, but it appears particularly high between the two pathogenic species (3, 223). Despite considerable similarity, they have significantly different pathogenicities. Genomic comparisons of *N. meningitidis, N. gonorrhoeae* and *N. lactamica* (a common human commensal) using DNA arrays, indicate that the pathogenic determinants do not appear to be located on large chromosomal islands as would usually be expected (168). However, a variable genetic island has been

reported in gonococci, a version of which carries a serum resistance locus, a cytotoxin gene and is preferentially associated with disseminated infections (56, 78, 200). Many of the genes studied in either of the two pathogenic species have had their homologues described in the other, as have many virulence factors been shown to have their counterparts, e.g. pilin, porin (Por, Protein I) and opacity proteins (Opa, Protein II) (200, 223). Interestingly, there is good evidence that the commensal species also carry virulence genes and may therefore act as reservoirs for virulence alleles (133, 201). Indeed, there is evidence of interspecies recombination occurring between members of this genus but this is likely to less commonly involve *N. gonorrhoeae* due to ecological separation from *N. meningitidis* and the other human commensal species (63, 126, 162, 232, 233).

1.2 Epidemiology of gonorrhoea

Gonorrhoea is the second most common bacterial STI in the UK after Chlamydia and not surprisingly, co-infection with these two bacteria is often reported (173). In 2008, almost 17 000 cases of uncomplicated gonorrhoea were diagnosed in genitourinary medicine (GUM) clinics across the UK (www.hpa.org).

Over the last 80 years, the epidemiology of gonorrhoea has been strongly influenced by key historical events and the incidence of gonorrhoea has therefore varied greatly, as is seen in the graph below showing the numbers of cases reported in England and Wales between 1925 and 2008.

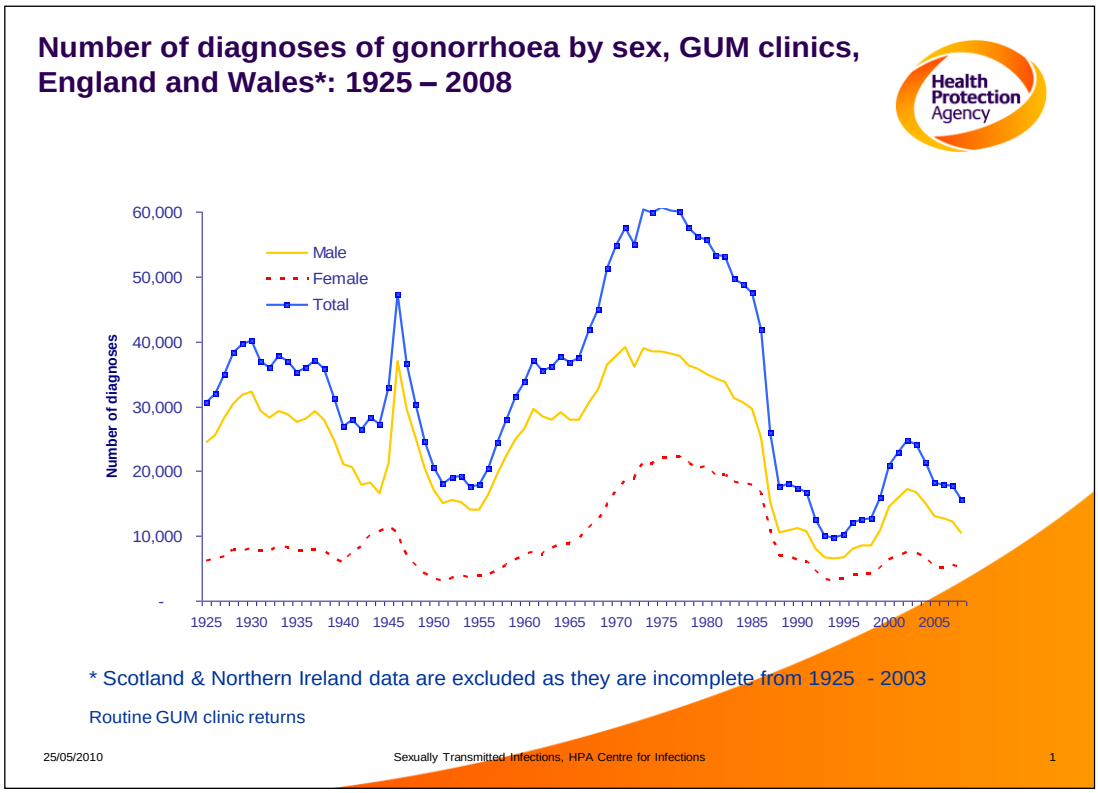


Figure 1.1 Number of diagnoses of gonorrhoea by sex, GUM clinics, England and Wales*: 1925 – 2008. Taken from www.hpa.org.

A sharp rise in cases coincided with the end of the Second World War and was followed by a decline in the 1950s following the introduction of penicillin. A second

peak occurred in the 1970s, thought to reflect the sexual revolution (219). Around 1985, there was a steep decline to the lowest ever recorded levels, especially amongst men who have sex with men (MSM), believed to have occurred in response to the HIV (the cause of acquired immunodeficiency syndrome, AIDS) pandemic, the first fatal STI, and contributed to by better education and access to effective diagnosis and treatment (98, 159, 170, 219). Incidence rose again briefly from the mid 1990's until 2002 and is believed to be due to an increase in unsafe sexual behaviour, probably because of the perception that AIDS is a treatable infection, along with delays in the access to treatment (57, 209, 243). Since then rates have been steadily declining in England and Wales (www.hpa.org).

Rates of gonorrhoea in both the developed and developing world vary greatly, and in many countries data about the incidence and prevalence are not accurate enough to give a clear picture of disease distribution (68). However, the incidence in developing countries, although not accurately known, is greater than in industrialised countries (68). The highest rates of gonorrhoea appear to be found in specific subgroups of the population, i.e. young people, lower socioeconomic groups and, in developed countries, among MSM, clients of commercial sex workers and travellers (98, 219). In England and Wales, the highest rates are found in London and other urban areas, among homosexual/bisexual men and black ethnic minority populations (www.hpa.org).

1.3 Diagnosis

N. gonorrhoeae is a Gram negative coccus. The cells are 0.6 to 1 micometres in diameter and are usually observed as pairs/diplococci (152). Presumptive diagnosis has historically been made by microscopy which has a high sensitivity in urinogenital symptomatic cases, less so in asymptomatic cases (98). In cases where gonorrhoea is presumed, diagnosis is made based on positive identification of *N. gonorrhoeae* at the infected site (40, 152). Culture, accompanied with a Gram stain and a positive oxidase test, is considered presumptive identification (98, 242) and is the method of choice in GUM clinics in the UK as it is inexpensive and sensitive (40), however, there are both biochemical and immunological tests available to provide a more specific identification.

Alternative genetic tests became available in the 1990s, including nucleic acid amplification tests (NAATs) and nucleic acid hybridisation tests, which are becoming increasingly popular because of their sensitivity, although their limitations include cost, inhibition and an inability to provide antibiotic resistance data (242). Comparisons suggest NAATs are highly sensitive, exceeding 90% at genital sites, whereas culture may be less than 75% sensitive for endocervical specimens (40, 229). There are currently no NAATs licensed for rectal and pharyngeal specimens (40), however in these cases, where culture may not always be sufficiently sensitive (164), there is increasing evidence that they have better sensitivity compared to culture and should be considered as the method of choice (4, 5).

1.4 Clinical manifestation

The severity of symptoms associated with an uncomplicated infection with *N. gonorrhoeae* can vary widely to the point of having no symptoms. Asymptomatic infection is more common in women than men (60, 152). When symptoms are present they manifest amongst men as urethral discharge and dysuria (40, 152, 219). For women, symptoms include an increase in or unusual vaginal discharge (98, 152). Other symptoms include occasional dysuria if the urethra is infected and lower abdominal pain. Infection of the endocervix is frequently asymptomatic; in fact, over 50% of cases of gonorrhoea in women are asymptomatic (40, 98, 152, 219). Complicated gonorrhoea may follow when uncomplicated episodes either go untreated or are treated but unsuccessfully. When the infection spreads up to the endometrium and fallopian tubes, the condition becomes known as pelvic inflammatory disease (PID) which may ultimately lead to infertility or ectopic pregnancies. In men, complicated gonorrhoea presents as prostatitis or epididymitis (98, 152, 219).

Symptoms at other sites are largely the same in both men and women. Rectal infection is most common in MSM and is often asymptomatic but may present with a purulent discharge, blood or mucous on the stool, rectal pain, burning or itching (40). In women, rectal infection is often the result of contamination from vaginal discharge and rarely true colonisation (98, 152). Infections of the pharynx are also usually asymptomatic, occasionally mild symptoms of a sore throat are experienced (40). More seriously, infection of the eyes may occur in newborns (opthalmia neonatorum) exposed during passage through the birth canal. If not treated, these can lead to serious damage to the eyes and also blindness (152).

Rarely (1-2%), mucosal infections will lead to gonococcal bacteraemia and result in disseminated gonococcal infections (DGI), the signs of which include, fever, rash and arthritis (152, 219). In these cases, the primary mucosal site is often asymptomatic (7, 177). The ratio of male: female DGI is 1:3 or 1:4, probably explained by the fact that asymptomatic gonococcal infection is more common in women than men (7, 177). Pregnancy and menses may further increase the likelihood of DGI in women due a change in the colony phenotype of gonococci from opaque to transparent, which is more resistant to the bactericidal nature of normal serum. This change has been associated with a few factors, including changes in vaginal pH, cervical mucous and genital flora (7, 177). Certain biological characteristics of gonococcal strains have been associated with DGI, including serum resistance (56), serovar A, lack of outer membrane opacity proteins on the cell surface, a nutritional requirement for arginine, hypoxanthine and uracil (116, 118), sensitivity to penicillin (7, 118, 150, 177), smaller colonies and weak fermentation of glucose (150).

1.5 Pathogenesis

The ability to attach and invade the mucosal surface of the genital tract is what distinguishes gonococci from the other *Neisseria* species (82). In order to establish an infection, the bacteria must adhere to epithelial cells to avoid being swept away by urine (in men) and cervical secretions (in women). Gonococci bind to the microvilli of non-ciliated columnar epithelial cells found on the mucosal surfaces (82, 152). Initial attachment is mediated by pili (hair-like extensions) that are present on the surface of the gonococci (212). CD46 (a receptor), which is present on nearly all human cells except erythrocytes (142), was first identified as a receptor for pili (108, 109); however, there has since been evidence that pilus adhesion can occur independently of CD46 (112, 113, 224). PilC, another phase variable gonococcal protein, has been implicated in both pilus biogenesis and pilus mediated epithelial cell adherence (179). Although initial attachment occurs via pili, initial interactions are probably due to electrostatic forces which act over longer distances. Pili are believed to be able to overcome the electrostatic repulsive forces expected due to the overall negative charge of both gonococcal and host cell surfaces (82, 84). After initial attachment with pili, subsequent and more intimate interactions occur between the Opa proteins present on the bacterial cell surface, and the carcinoembryonic antigen cell adhesion molecules (CEACAM, previously termed the CEA/CD66 family) or the heparan sulphate proteoglycans (HSPG) (81). The Opa proteins and their receptors are discussed in more detail in section 1.9.

Gonococci also have lipopolysaccharides on the cell surface which mediate binding. These have a highly branched oligosaccharide structure and lack the O-specific side

chain and therefore are sometimes referred to as lipooligosaccharide (LOS) (152). There is evidence that gonococci can also be internalised into epithelial cells via the asialoglycoprotein receptor that binds to LOS (80). LOS is capable of antigenic variation, as well as phase variation of the expression loci via slipped strand mispairing of homopolymeric tracts (156, 197, 200). LOSs are important in pathogenesis because they confer immune serum resistance when sialylated, (197) by preventing complement-dependent killing (167, 241). Sialylation occurs due to host factors and is lost on just one sub-culture *in vitro*. Resistance can be restored when cells are incubated with genital secretions, serum or red or white blood cell extracts (197). Desialylation occurs before the cells are internalised (156). Interactions between Opa proteins and sialic acids (on sialylated LOS) are thought to interfere with the target receptor thus resulting in reduced host-bacterium interactions (148). The gonococcal outer membrane protein (OMP) porin, Por, is also believed to play a role in adhesion and is associated with increased bacterial invasion (137).

Once adhesion has been achieved, cells are pinched off by the mucosal cell membrane into a vacuole via a process called parasite mediated endocytosis (152). Pili however appear to inhibit invasion of adherent gonococci *in vitro* but loss of the piliated phenotype allows invasion of the adherent gonococci (131). Therefore, subsequent invasion via Opa protein interactions may require the loss of pili. Intracellular multiplication occurs within this vacuole which migrates to the basal membrane where the bacteria are released into the sub-epithelial tissues and establish infection (82, 152). Here the LOSs, present on the gonococcal cell surface are shed causing the release of proteases, phospholipases and the production of

proinflammatory cytokines leading to the local inflammatory response and the cell damage associated with symptomatic infection (82, 152). Gram negative bacteria usually induce the release of cytokines when epithelial cells and mononuclear phagocytes (macrophages) are stimulated (180). One study could only detect elevated levels of cytokines when the patient had a concomitant STI (86). The lack of an expected inflammatory cytokine response led to speculation that gonococci have evolved to avoid stimulating the initial epithelial cell cytokine responses (180). Indeed, LOSs are antigenically similar to antigens present on erythrocytes and therefore the similarity may preclude the immune response to these molecules (152, 156, 166, 231).

Most patients produce antibodies to gonococcal LOS, pili, and Opa proteins and antibodies produced against Opa proteins are highly specific (82, 245). Immunoglobulin (Ig) A is the primary protective immunoglobulin at mucosal sites, however evidence of protection in the genital tract is poor (143, 152, 180). IgA against Opa proteins can be detected in vaginal fluids more so than in serum (125) but gonococci produce an IgA protease that cleaves it (105, 180). IgG against Opa proteins can be detected in both serum and vaginal fluids (125) and recent work has suggested it is the dominant immunoglobulin (143, 181). IgG and IgM are considered effective against gonococci in that they bind to the bacteria and thus tag them for opsonisation by macrophages and polymorphonuclear neutrophils/leukocytes (PMNs) and activate the complement system to disrupt the bacterial cell membrane, however, there is recent evidence that strongly suggests gonococci are in fact highly resistant to PMN killing and delay apoptosis to allow intracellular replication (49, 194). Although detectable, in general, antibody

responses appear to be weak in both serum and genital secretions (85, 165, 180, 181). The absence of any protective immunity to infection from prior infections correlates with the observation that patients who reported previous infections with gonorrhoea do not have greater immune responses than those with no previous infection reported (180, 186). The lack of protective immunity is therefore likely due to a complex balance of a less than ideal immune response and the organism's remarkable ability to adapt and evade the immune response via antigenic variation.

1.6 Treatment and antimicrobial resistance

Gonorrhoea remains a disease for which a vaccine has not been found. Many extracellular exposed components have been considered as vaccine candidates but none have yet been successful or are considered suitable. The challenge in finding a vaccine lies not only in finding a structure or target common in all strains but also one that must fulfil several criteria. The ideal target needs to be stable enough to be present in unrelated strains. Unfortunately, many external moieties, for example, pilin or other surface-exposed OMP, are far too variable to fulfil this role and so far the search has been unfruitful. Not only must a good vaccine candidate be consistently present in the organism but most importantly it needs to be strongly immunogenic to ensure that the immune response raised is sufficient to clear the infection. A major hindrance in the search for a vaccine against gonorrhoea is that it appears that gonococci do not elicit a strong enough antibody response to generate any lasting immune response memory (85, 180) and evade the host immune system via highly effective antigenic variation systems.

Due to the lack of a vaccine against gonorrhoea, treatment is the primary means of control. The evidence that gonorrhoea (together with other STIs) facilitates HIV transmission further highlights the need for effective control measures as it is important also as part of HIV prevention strategies (65). Regardless of still being relatively easy to treat, *N. gonorrhoeae* has acquired resistance to antimicrobial agents. The ease with which they acquire antibiotic resistance is concerning and is either plasmid or chromosomally mediated.

Penicillin used to be effective but no longer is due to the acquisition of plasmid-encoded β-lactamases (6, 169) which became disseminated worldwide by the 1980s (98, 176), or less commonly, chromosomally mediated changes of penicillin-binding proteins. Resistance has also occurred to several other classes of antimicrobial agents that were used since penicillin was no longer the antibiotic of choice, including spectinomycin and ciprofloxacin (59, 67, 101, 158, 222). Treatment will therefore sometimes fail due to resistance of the strain. The current treatment regime for gonorrhoea in the UK is a single dose of ceftriaxone or cefixime which is given either by injection or orally in tablet form, respectively. These antibiotics are highly effective in treating susceptible organisms, however the disease is maintained in the population via core groups and continues to be a burden on public health (33, 221), (Figure 1.2). Although the current treatment recommendations are still very effective, there are already signs of resistance to them emerging. In 2007, the first signs of reduced susceptibility to ceftriaxone were observed (173). As part of patient management, patients are advised to avoid sexual contact until their (and their partner(s)) treatment is complete and includes counselling on the long-term health effects of the condition for both themselves and partners.



Figure 1.2 Gonorrhoea in the news. Taken from the BBC website on 30 March 2010. (http://news.bbc.co.uk/1/hi/health/8593366.stm)

1.7 Gonococcal genetics

The genome of *N. gonorrhoeae* consists of a single chromosome approximately two mega base pairs in size (55). The complete genome sequences of at least two strains, namely FA1090 (http://www.genome.ou.edu/gono.html, GenBank Acc. No. AE004969) and NCCP11945 (39), have now been determined, the availability of which has and will continue to make further studies on other gonococcal strains much easier.

Gonococci commonly carry a cryptic 4.2 kilo base pairs (kb) plasmid, but its presence has not been correlated with virulence or any other properties and thus to date it has no known function (176, 200). Gonococci also have conjugative plasmids. The prevalence of the 39kb conjugative plasmid varies geographically and temporally. It does not carry any detectable markers for resistance to antibiotics but it mobilises itself and the smaller β-lactamase plasmids (which encode penicillin resistance) efficiently between strains (176, 200). In addition to penicillin resistance via the β-lactamase plasmids, tetracycline resistance appeared due to the transposition of a tetracycline resistance determinant into the conjugative plasmid (176, 200).

Most relevant though to this project is the fact that gonococci are highly competent for transformation and, unlike most other naturally transformable bacteria, they remain competent throughout their lifecycle (205) which allows more opportunities for transformation to occur. They are highly recombinogenic organisms (139, 161, 162, 232) and as well as via transformation, recombination may also occur within the

cell before cell division occurs (intra-chromosomal recombination) (94). Gonococci are highly autolytic, thereby releasing their DNA into the environment and provide DNA for uptake by other gonococci by transformation (19, 77). Additionally, gonococci have a type IV secretion system that secretes chromosomal DNA into the environment (not directly into a neighbouring cell) that can then be involved in transformation events (56, 77, 78). Gonococci bind both heterologous and homologous DNA but only homologous DNA is used for transformation (19, 77). It is limited by the requirement of genus specific DNA uptake/recognition sequences (62, 70, 77) and it is also thought that to some degree, DNA homology is capable of excluding foreign DNA (89).

There is evidence that pili/associated components also play a role in DNA uptake in transformation (18, 77) because competence is generally restricted to piliated organisms or those that produce the pilin subunit (19, 119). Loss of pili occurs readily *in vitro* and the competence of non-piliated organisms was found to be greatly reduced; however, piliated organisms can be deficient in DNA uptake suggesting that pili alone are not sufficient for transformation (19, 119). Interestingly, the expression of Opa proteins has also been correlated with increased rates of transformation, suggested to occur due to an expected positive charge on the surface of the cell, promoting donor DNA sequestration and thus enhancing more efficient transformation over time (91).

There is no evidence of other para-sexual processes common in other bacterial species, such as phage-mediated transduction, in the movement of chromosomal DNA between gonococci (19, 66, 205). Conjugation has been reported but it does

not appear to facilitate the movement of chromosomal genes between gonococci (19, 66).

Recently, Tobiason *et al.* reported that *N. gonorrhoeae* was polyploid, over and above the diploid state a cell may be in after replication but before cell division occurs (225). If correct, this would have implications for increased recombination potential within a cell because single copy genes become 'multi-gene families', but on separate chromosomes, and the original multi-gene families become super multi-gene families.

Gonococci have several diverse surface proteins that interact with host cells, the expression of which varies widely between different isolates. The types of variation can be categorised into two classes, antigenic variation and phase variation. Antigenic variation results in changes to a protein's primary sequence whereas phase variation has a simple 'on/off' effect (146). These two mechanisms are referred to often in later sections and it has become clear that in addition to the immune evasion advantages they bestow, these two mechanisms also affect other aspects of gonococcal biology including agglutination, charge density, cell tropisms, adhesion and invasion, to name a few. One review eloquently referred to them as cell differentiation mechanisms because of the consequential phenotypic variation within a particular lineage (142).

1.8 Strain Typing

Typing micro-organisms is of great importance for studying disease transmission and may also in some cases provide data to explore population structures. Neutral markers are ideal for obtaining the true relationships among isolates, for investigating population structure and indexing genetic relatedness (192). The questions that are most pertinent to a sexually-transmitted infection are related to very recent events, such as identification of individuals who are within a sexual network or providing supporting evidence for individuals being sexual contacts (including forensic evidence). For addressing such questions, slowly evolving neutral variation is not useful and variation that accumulates very rapidly is required. For *N. gonorrhoeae*, it may therefore be permissible to forgo the requirement of selectively neutral loci since we are mainly interested in short term transmission (160), localised epidemics and circulating strains, to inform local intervention strategies, rather than in comparing strains from different continents and decades.

Early methods for typing gonococci included auxotyping (31, 32) and serotyping (117, 218). Briefly, auxotyping categorises gonococci according to their nutritional requirements (31). All gonococci require cysteine for growth but some strains of *N. gonorrhoeae* require additional amino acids (31). For example, an isolate classed as AU would in addition to cysteine, require arginine and uracil. Auxotyping was useful in helping define the biological characteristics of strains associated with DGI, namely the AHU auxotype (116, 118).

Serotyping is based on reactions with monoclonal antibodies raised to the main OMP Por present on the surface of the cells (117, 182, 218). There are two main classes/serovars of Por according to the scheme, Por1A and Por1B, which are further subdivided into serotypes based on combinations of reactions to a panel of monoclonal antibodies. While each classification system on its own proved useful (100), it quickly became clear that they did not offer sufficient resolution as unrelated isolates often had the same auxotype or serotype and a scheme based on the combination of auxotype and serotype soon emerged (117). Combined auxotyping/serotyping (A/S) improves resolution (182, 230) but in most populations, only a limited number of A/S classes are represented (99) and it does not always distinguish between isolates that are epidemiologically linked and those that are not (160). This led the way for genotypic methods such as plasmid profiling, which allowed the spread of resistance phenotypes to be monitored; pulsed-field gel electrophoresis which resolves large genomic fragments to generate unique profiles (172); and ribotyping, the restriction fragment length polymorphism analysis of ribosomal genes, was also good but was best used in conjunction with serotyping (171). Another method based on fragment length polymorphisms is fluorescent amplified fragment length polymorphism analysis which is a polymerase chain reaction (PCR) based genome sampling technique that results in a (fluorescent) fragment profile for each isolate (163).

The house-keeping genes of gonococci are relatively uniform (204, 232) making applications such as multi-locus enzyme electrophoresis (MLEE) and multi-locus sequence typing (MLST) (203), which target house-keeping genes, unsuitable for discrimination between unrelated isolates in a sexual network. In order to attain

levels of sufficient discrimination, large numbers of house-keeping loci would need to be targeted (160).

Using variation that accumulates rapidly may allow individuals within a sexual network to be identified by the sharing of the same gonococcal strain (160). In this approach, individuals who are not epidemiologically linked should have distinguishable gonococcal isolates using a highly discriminatory typing method, whereas those from sexual contacts, or individuals within a short chain of transmission or sexual network, should have indistinguishable isolates. To this end, *opa*-typing was developed which employs the diversity exhibited by a family of 11 polymorphic *opa* genes (160). Briefly, it is the restriction fragment analysis of internal fragments of all 11 *opa* genes amplified simultaneously by one pair of primers in a single PCR. The restriction endonucleases (four base cutters are used to produce large numbers of fragments) recognise sites/sequences within the 11 *opa* gene fragments and generate a complex pattern of fragments. It is far more discriminatory than phenotypic methods such as A/S typing (96, 240) and is considered to be sufficiently discriminatory such that isolates recovered from individuals which were sexual contacts or were closely linked through chains of transmission frequently have identical *opa*-types (160). Although *opa*-typing data correlate well with contact tracing data, and is also able to distinguish virtually any pair of randomly chosen gonococci (160, 240), it is labour intensive and the data (complex fragment patterns) are not easily comparable between laboratories.

In general, typing of micro-organisms has largely turned to sequence based methods, such as MLST and *por* gene sequencing, in the case of gonococci (129, 140, 238,

239), as they allow precision and high-throughput analyses of bacterial populations. For *N. gonorrhoeae*, the sequence based genotyping method *N. gonorrhoeae* multi-antigen sequence typing (NG-MAST) (136) has become a popular choice internationally. NG-MAST involves sequencing internal fragments of two highly polymorphic loci (rather than house-keeping loci), *porB* (490 base pairs, bp) and *tbpB* (390bp), which encode the OMP Por and the β-subunit of the transferrin-binding protein respectively. Each unique sequence at each locus is assigned an allele number and thus provides each isolate with a two digit allelic profile. Each unique two digit profile is assigned as a sequence type (ST), also an integer. There are high levels of congruence between *opa*-typing and NG-MAST but NG-MAST appears to correlate slightly more closely with the contact tracing data provided by patients (1, 16). Although *opa*-typing appears to allow a finer discrimination between isolates, NG-MAST aims to provide sufficient discrimination such that the results may indicate links between individuals within a sexual network and identify outbreaks due to strains that have public health significance such as antibiotic resistant strains.

Sequence based methods such as NG-MAST have advantages in that they allow unambiguous results that are reproducible and easily compared within and between laboratories. They are not subject to variations in reagent batches (97, 100) or an individual's personal interpretation, as may be common with methods such as serotyping, auxotyping and *opa*-typing. Sequences, profiles and other strain information can be easily stored on and shared/compared via a website.

1.9 Project motivation and objectives

In culture, gonococcal colonies may vary in their degree of opacity (111). Colonies appear to switch from transparent to opaque and vice versa at the same rate (138). This opacity change is based on the variable expression of Opa proteins in the outer membrane (hence the name opacity) which affects the auto-aggregation of cells (123, 211, 213, 214). The Opa proteins in meningococci are very similar to those in gonococci (110), however, their expression does not necessarily correlate to colony opacity in encapsulated meningococci (2).

The mature Opa protein comprises eight β-strands which fold to form β-barrel structures spanning the membrane with four surface exposed loops (Figure 1.3) and as described earlier, they play an important role in adherence and invasion of host cells. They are encoded by a family of 11 independent *opa* genes dispersed throughout the genome (14, 15) designated *opaA* through to *opaK*, named alphabetically in order of increasing electrophoretic mobility of their *Eco*RI and *Sma*I restriction fragments (14). In comparison, *N. meningitidis* is known to have only three or four *opa* genes (2). Most commensal members of this genus do not have *opa* genes or may have one copy, with the exception of *N. flavescens* which has been reported to have two copies and *N. lactamica* which has been reported to have two or three copies (133, 226).

There are regions of considerable sequence similarity among the different *opa* genes, as well as highly variable regions (45) and recognisable genetic elements or motifs (Figure 1.3). Among these are the tandem pentameric repeats (CTCTT) called the

coding repeat (CR) present in the signal sequence encoding region (206). Comparisons of the sequences of *opa* genes have identified three particular regions where differences appear to be largely confined, a semi-variable region (SVR) encoding the N-terminal amino acids and two short hyper-variable regions (HVR1 and HVR2) (94, 206). These observations correlate with peptide mapping experiments which identified common hydrophobic peptides with few, relatively unique, hydrophilic ones (82, 83, 210). Therefore, three of the four surface exposed loops are encoded by the variable regions of the gene and host antibodies will primarily target epitopes in these regions (which over time has driven their variability) (8, 14, 146) while the fourth loop appears conserved (14).

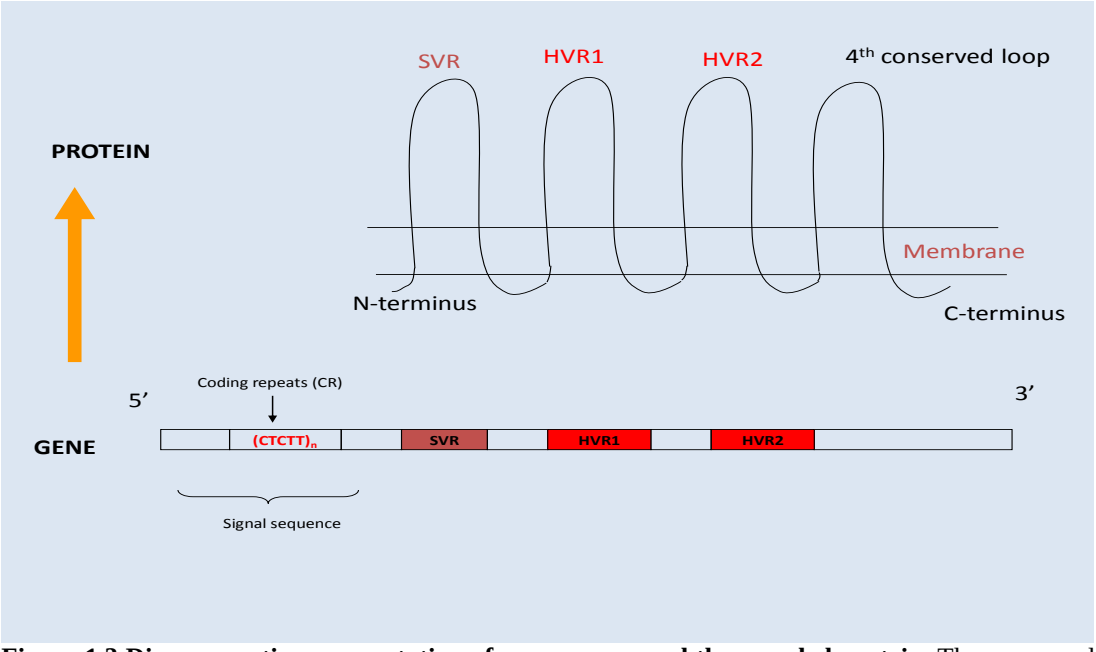


Figure 1.3 Diagrammatic representation of an *opa* gene and the encoded protein. The conserved gene structure is shown indicating the variable and CR regions. The secondary protein structure is also shown indicating the surface exposed loops.

Certain Opa proteins have been associated with different properties such as the degree of adherence to epithelial cells after binding mediated by pili (212) and they may also differ in their cell tropisms (74, 82, 122, 193). Opa proteins bind to either

of two families of receptors, the HSPG receptors ubiquitously expressed on eukaryotic cells (34, 52, 234) or the CEACAM receptors (35, 36, 52, 73, 236). Opa proteins that have surface exposed loops rich in positively charged amino acyl residues tend to bind to the HSPG receptors (142) present on epithelial cells. Binding and invasion are reduced in the presence of competing polyanions e.g. heparin and heparan sulphate (142). There are at least two mechanisms for uptake via HSPG receptors. Either the action of phosphatidylcholine-specific phospholipase C and acid sphingomyelinase or serum factors vitronectin or fibronectin along with integrin receptors (52, 53, 72, 120). The HVR1 of the Opa protein has been shown to be the main region responsible for binding particular HSPG receptors (17). When replaced with HVR1 from different Opa proteins, binding is interrupted (71). The SVR and HVR2 enhance receptor binding but do not represent binding domains themselves (71).

Many Opa protein variants do not in fact bind to HSPG receptors but rather the CEACAM family of receptors present on PMNs and epithelial cells (22, 35, 36, 52, 61, 64, 73, 155, 236, 237). CEACAM receptor proteins can mediate cell-cell adhesion via protein-protein binding interactions at the amino-terminal domain of the receptor (52, 120, 237). The production of cytokines upon infection results in up-regulation of expression of this receptor and may thus result in enhanced Opa-mediated binding and invasion (52, 75, 153). Different CEACAM receptors are expressed on different host cell types (52, 142). Therefore, the differential specificities of the receptors will result in different combinations of receptor and Opa protein and can lead to different host cell responses e.g. binding or uptake (74, 142, 153, 234).

The binding sites of Opa proteins to the CEACAM receptors are relatively conserved, allowing distinct CEACAM receptors to be targeted by one or more Opa proteins (35, 237) More specifically, three residues in the receptor have been identified that are required for optimum Opa receptor activity (23) which co-localise to create an exposed face which has no carbohydrates (120). Different Opa proteins that recognise the same receptor show variable affinities to the receptors with sub-optimal activity suggesting that the binding characteristics in each case are slightly different (23).

Different Opa proteins can therefore differ in their ability to interact with their receptors, and their ability to promote invasion of human cells (22, 122, 236). The effect of Opa expression with respect to serum sensitivity and resistance has been unclear. It has been well documented that Opa$^-$ (non-Opa expressing, transparent colonies) can be cultured from disseminated infections, the cervix of women at the time of menstruation and the urethras of asymptomatic men, suggesting that this phenotype promotes serum resistance (102, 150). There are contradictory reports however which do not take into account other factors such as LOS mediated serum resistance and the differences in Opa expression between different strains (24, 37, 82, 103, 124, 235).

Recently, Opa proteins have been shown to promote persistence of gonococci in the female genital tract in both early and late stages of infection, their phase variable nature allowing them to make use of host factors (43) and the variable expression of Opa proteins is thought to be a determinant in the severity of cases of PID (130). Recent *in vitro* studies suggest that the interaction between the CEACAM receptors

and Opa proteins may also aid immune evasion as they have been shown to suppress the activation of lymphocytes, antibody production and the inflammatory response of epithelial cells (26, 165, 195).

In vitro studies have been complemented by human challenge studies and have proven to be an efficient means of studying gonococcal infection (41, 42). These have demonstrated that Opa expression is important in the early stages of establishing an infection and which Opa proteins are expressed will also vary during the infection (104, 185, 215). When Opa^- variants were used to infect male volunteers, predominantly Opa^+ (Opa expressing) organisms were recovered (104, 215). In human volunteers infected with the same strain of *N. gonorrhoeae*, different Opa proteins are expressed in different individuals (104) which suggests that there are multiple Opa proteins within a strain that can facilitate interactions at a mucosal surface.

Gonococci can exist in different states of Opa expression, expressing none, one or several at any one time (20, 123, 213, 214). Each monocistronic *opa* gene has a single open reading frame which is constitutively transcribed (14). The expression of each is controlled by phase variation, defined as reversible on-off switching (206), reported to occur *in vitro* at a frequency of approximately 1-3 x 10^{-3} per cell per generation (13). Phase variation control mechanisms are well conserved among the *opa* genes of the pathogenic *Neisseria* (207) and numerous other putative phase variable genes have been identified among these species (199). Expression (translation into mature protein) is dependent on the correct multiple of CRs in the signal sequence of the locus, such that the transcript is in the correct reading frame to

prevent premature termination of the protein product (202, 206). Genomic regions that contain such repeats (termed short sequence repeats) are prone to transient misparing during replication (87). Illegitimate recombination (12) and slipped strand mispairing events allow variation in the number of CRs to occur in this region by insertions and deletions (indels) and hence control Opa expression (13, 87, 154, 206). It is also believed that the CRs are able to form triple stranded anti-parallel H structures in which regions of single stranded DNA may be exposed and targeted by single strand nucleases resulting in single strand breaks (11). Repair to these breaks may lead to changes in the number of CR units. Comparisons of Opa expression *in vitro* and in human challenge suggest the infected host may also play a role in the prevalence of some Opa proteins (185). Also, the strength of different promoters present at each locus may have an effect on translation (12). Some *opa* genes found to be expressed at higher rates, correlated with variation in the rate of phase variation from 'off' to 'on' states (12). Although not clearly established, it is likely that during the initial stages of infection, gonococci that express Opa and, more specifically, that express Opa proteins that have tropisms for the encountered mucosal surface (urethral, cervical, rectal, pharyngeal), are selected.

The family of *opa* genes in gonococci is believed to have evolved by various genetic mechanisms. Studies using phylogenetic analyses of *opa* gene sequences from *N. gonorrhoeae* strain MS11 variants, suggested gene duplication, gene replacement and partial non-reciprocal recombination events were important in the evolution of the *opa* gene family (14), with similar conclusions drawn from studies in meningococci (94). Gene duplication involves a complete copy of one *opa* gene being incorporated into a new location on the chromosome, leading to an increase in

the number of *opa* loci. Such an event may have been aided by unique 26bp repetitive sequences, present in all members of the genus *Neisseria* and which occurs in the gonococcal genome with a copy number of approximately 20 (47, 48, 127, 133). An inverted repeat sequence which is identical to the DNA uptake sequence identified by earlier studies follows the termination codon in all the *opa* loci (14, 62, 70).

Gene replacement is the replacement of the coding region of one *opa* locus with that of another by means of recombination, leading to a duplication of the original *opa* sequence although not changing the number of *opa* loci. Partial recombination events within coding regions between the different *opa* genes further increases the diversity of the genes by allowing new combinations of the hyper-variable regions (HVRs) (14, 45).

Gene replacements and partial recombination events also contribute to ongoing diversification of the *opa* gene family. Regions of homologous DNA between the *opa* genes aid recombination events, not only between different *opa* genes in the same strain, but also between different gonococci (14). Concurrent infection with more than one lineage has been reported, at which time recombining DNA is made available by autolysis of co-infecting strain(s) (135), and serves as an opportunity to introduce new hyper-variable sequences by transformation (45, 187). Antigenic variation would be the overall result of any of these recombination events that change the primary amino acid sequence of the protein.

It is now common knowledge that individuals may experience repeated episodes of gonorrhoea (186) and the reason for this has much to do with the immune response of individuals, or lack thereof, as well as the organism's capacity to vary antigenically. Although Opa proteins are immunogenic (244), the humoral immune response elicited is reported to be only modest (85, 180) and the lack of sufficient anti-gonococcal antibodies generated during infection might also be responsible for the lack of immunity to re-infection (86). The lack of apparent immunity is also thought to be attributable to the variation in their surface antigens which allows them to evade the host immune response (145). However, since the receptor-binding sites involve the surface exposed HVRs of the Opa proteins, non-synonymous (NS) substitutions in these regions of the gene, selected by the need to evade the host immune system, are presumably strongly filtered by the need to maintain receptor-binding activity. The importance of Opa proteins is therefore emphasised by the role they play both in attachment to the host mucosa as well the HVRs being targeted by the host immune system.

Due to the strong selection pressure imposed by the human immune system, the *opa* genes evolve rapidly and consequently are highly variable. It is currently not clear just how rapidly *opa* genes diverge during transmission of gonorrhoea within a community (160). The mechanisms by which changes in *opa* genes occur have been studied in *N. meningitidis* by analysing isolates belonging to the same clonal complex, defined by MLEE (93, 94) or MLST (30) and include NS mutations within HVRs, recombination events that include *opa* gene allele duplication and generation of mosaic alleles, and import of novel *opa* sequences from other strains. While *opa* gene diversity and the general mechanisms involved in *opa* gene variation have been

described in meningococci, there is little information on the rates of change at *opa* loci or the relative contributions of the different mechanisms of *opa* gene variation. The diversity of the larger gonococcal *opa* gene family has been much less well characterised and has been restricted to analysis of the *opa* sequences from strain MS11 (14) and the HVRs from FA1010 (46) and analysis of additional strains with a set of HVR1 and HVR2 DNA hybridisation probes (27). This is largely because previous studies have not been able to look at the changes of the *opa* sequences at the individual *opa* loci since no method has been available to amplify and sequence each individual locus.

In order to tackle these questions it is necessary to use isolates that have been collected from individuals involved in transmission chains such that the changes that are observed are very recent and consequently are the most likely to be readily interpretable. In this work I will describe a method to amplify each of the 11 *opa* genes independently in order to determine the sequences at each locus. By focusing on isolates from known recent sexual contacts in Sheffield and small sexual networks in London I will address the above questions.

Chapter 2: Methods

2.1 Bacterial culture

Isolates were removed from storage at -80°C in glycerol broth and plated onto single-strength Difco GC medium base (Becton Dickinson, Oxford, United Kingdom) supplemented with 1% Vitox (Oxoid, Basingstoke, United Kingdom) and incubated overnight at 37°C under 5% carbon dioxide. Single colonies were picked and sub-cultured, from which DNA lysates were prepared.

2.2 DNA preparation

Bacterial growth was re-suspended in phosphate buffered saline and boiled for 5 minutes (min) to lyse the cells. The boilate was then centrifuged at 2000 times gravity (x g) for 5 min to pellet the cell debris. The supernatant containing the DNA was removed and stored at -20°C until use.

2.3 Amplification of *opa* genes

All loci were amplified by PCR using the Extensor Hi-Fidelity PCR Master Mix, Buffer 1, Reddymix Version (Abgene, Epsom, United Kingdom), according to the manufacturer's instructions. Primers were designed within conserved regions of genes which flank each *opa* gene. The primer pairs used for each *opa* locus, the annealing temperatures and extension times are listed in Appendix 1. Primer design was aided using the program Primer3 (178) and primers were manufactured by Eurogentec (Seraing, Belgium).

The sizes of each PCR product varied depending on the *opa* locus, ranging between approximately 2085bp and 7497bp. 2 microlitres (µl) of each product was electrophoresed with a quantitative size standard (Lambda DNA/*Hind*III marker 2; Fermentas Life Sciences, Burlington, Canada) in 0.8% agarose gel containing SafeView Nucleic Acid Stain (according to the manufacturer's instructions; NBS Biologicals, Huntingdon, Cambridgeshire, United Kingdom) at 15 volts per centimetre for 40 min in 1 X Tris borate EDTA (ethylenediaminetetraacetic acid) buffer and visualised under ultra-violet light.

The remainder of each PCR product was cleaned of excess primers and nucleotides using SureClean (Bioline, Randolph, USA), using the manufacturer's recommendations as guidelines. Briefly, equal volumes of SureClean were added to each reaction, mixed and incubated at room temperature for a minimum of 10 min. Reaction mixtures were then centrifuged at maximum speed (approximately 3000 x g) for 20 min to pellet the products. Pellets were then washed twice with 70% ethanol and dried. Cleaned products were re-suspended in a volume of sterile water to achieve a final DNA concentration of approximately 100 nanograms per microlitre.

2.4 Sequencing of each *opa* locus

Each *opa* gene was then sequenced using the BigDye Terminator Cycle Sequencing Kit version 1.1 (Applied Biosystems, Cheshire, United Kingdom). Each reaction contained 2µl PCR product, 1µl primer (at a concentration of 1 picomole/microlitre) and 2µl BigDye and cycled as follows: Initial denaturation for 3 min at 96°C

followed by 30 cycles of 30 seconds (sec) at 96°C, 10 sec at 50°C and 2 min at 60°C. Reactions were then cooled to 4°C at a rate of 0.1°C per sec. The primers used to sequence each *opa* gene included those designed for *opa*-typing (160) (opa-up 5'-GCGATTATTTCAGAAACATCCG-3' and opa-dn 5'-GCTTCGTGGGTTTTGAAGCG-3') and two primers designed from central conserved regions that sequence outwards towards the start and end of the genes, int1 (5'-GATTATGCCCGTTACAG-3') and int2-GAT/int2-TGC (5'-CC(GAT/TGC)ATAGGGTTTGAA-3'). Sequence off the former determined which of the two alternative second primers were used. This combination of primers provides sequence data off both strands over almost the entire length of the *opa* gene including the HVRs. Short regions at either end could therefore only be sequenced on one strand but with two independent primers. The sequence of the *opa*-typing primers was reverse complemented so that they could also be used in the reverse orientation (Ropa-up 5'-CGGATGTTTCTGAAATAATCGC-3' and Ropa-dn 5'-CGCTTCAAAACCCACGAAGC-3') to explore the immediate flanking regions and also in cases where the sequence at the very start of the gene was required.

All sequencing products were then cleaned of excess primers and BigDye by a sodium acetate/ethanol method. Briefly, 15µl of a 4:11 mixture of 3 molar sodium acetate pH 5.2:sterile deionised water was added to each reaction, as well as 50µl 95% ethanol and mixed well. Reaction mixes were then incubated at -20°C for 1 hour to precipitate the DNA and then centrifuged at approximately 2000 x g for one hour at 4°C to pellet the products. Pelleted products were then washed twice with 70% ethanol, dried and stored at -20°C until further use. When required, clean products were re-suspended in 10µl Hi-Di Formamide (Applied Biosystems) and

separated on an ABI 3700 DNA Analyzer (Applied Biosystems). All sequence data were viewed, edited and assembled using MEGA version 3.1 (121). Sequences were trimmed (in the +1 reading frame) almost immediately after the last CR at the start of the conserved sequence GCAGC and after the last nucleotide of the stop codon. In a very small number of alleles the first G in this conserved sequence was absent and the trimmed sequence began instead with a T. Identical sequences were identified using the non-redundant database tool available at http://linux.mlst.net/nrdb.nrdb.htm and each unique sequence was assigned an allele number. Any sequence that occurred only once (i.e. at one locus in a single isolate) was confirmed by re-amplifying and re-sequencing to confirm the new sequence before assigning an allele number. Alignments were also performed with MEGA (121).

2.5 NG-MAST and MLST

NG-MAST was performed on all isolates as described by Martin *et al.* (136). Briefly, internal fragments of *por* and *tbpB* were amplified, sequenced and assigned allele numbers at www.ng-mast.net that defined the ST. MLST of certain isolates was done as previously described (129) but with some slightly different primers based on the genome sequence of strain FA1090 (Daniel Godoy, personal communication) for both the PCR and sequencing reactions instead of those described. Briefly, internal fragments of seven house-keeping genes were amplified, sequenced and also assigned allele numbers, the combination of which defined the ST.

Chapter 3: Amplification and sequencing of the 11 gonococcal *opa* genes

3.1 Introduction

Prior to complete genome sequencing, information on the location of the 11 gonococcal *opa* genes in the genome had been based on physical macro-restriction studies (55). In 1991, physical maps of two isolates of *N. gonorrhoeae* were published, FA1090 (55) and MS11 (15). These two studies identified the location of important genes, including the *opa* loci. Gonococcal *opa* genes have been designated *opaA-K*, in order of increasing electrophoretic mobility of the restriction fragments on which they are located (14). In order to collect sequence information for the *opa* genes under this nomenclature, cloning technology would be required to isolate individual *opa* genes. In recent years, sequencing technology has advanced to such a degree that it has become routine and relatively inexpensive and has the obvious advantage of providing unambiguous data. This has meant that complete genome sequences of organisms are becoming readily accessible and can be used for a myriad number of applications. The complete genome sequence of strain NCCP11945 (39) and FA1090 is now available (http://www.genome.ou.edu/gono.html, GenBank Acc. No. AE004969). Although the genome sequence of FA1090 has not been fully annotated (http://stdgen.northwestern.edu/stdgen/bacteria/ngon/), the sequence information could be used to design more locus specific experiments such that cloning could be avoided.

In this chapter I describe how I amplified and sequenced each *opa* gene independently from isolates of *N. gonorrhoeae*. The method was applied to a set of unrelated isolates in order to demonstrate its use amongst diverse strains. These results led to the proposed nomenclature for gonococcal *opa* genes.

3.2 Methods

The complete genome sequence of FA1090 was downloaded from http://www.genome.ou.edu/gono.html (GenBank Accession Number AE004969). The single contiguous sequence begins at the gene encoding the chromosome replication initiator protein, *dnaA*. The position of each *opa* gene in the genome sequence was located by identifying sequences unique to gonococcal *opa* genes. These sequences included the primer sequences used for *opa*-typing (160), the variable number CRs, a conserved region between HVR1 and HVR2 and the DNA uptake/recognition sequence frequently found almost directly downstream of the stop codon (62, 70). The 11 *opa* loci were numbered consecutively, 1-11, in their order around the chromosome of this strain (Figure 3.1).

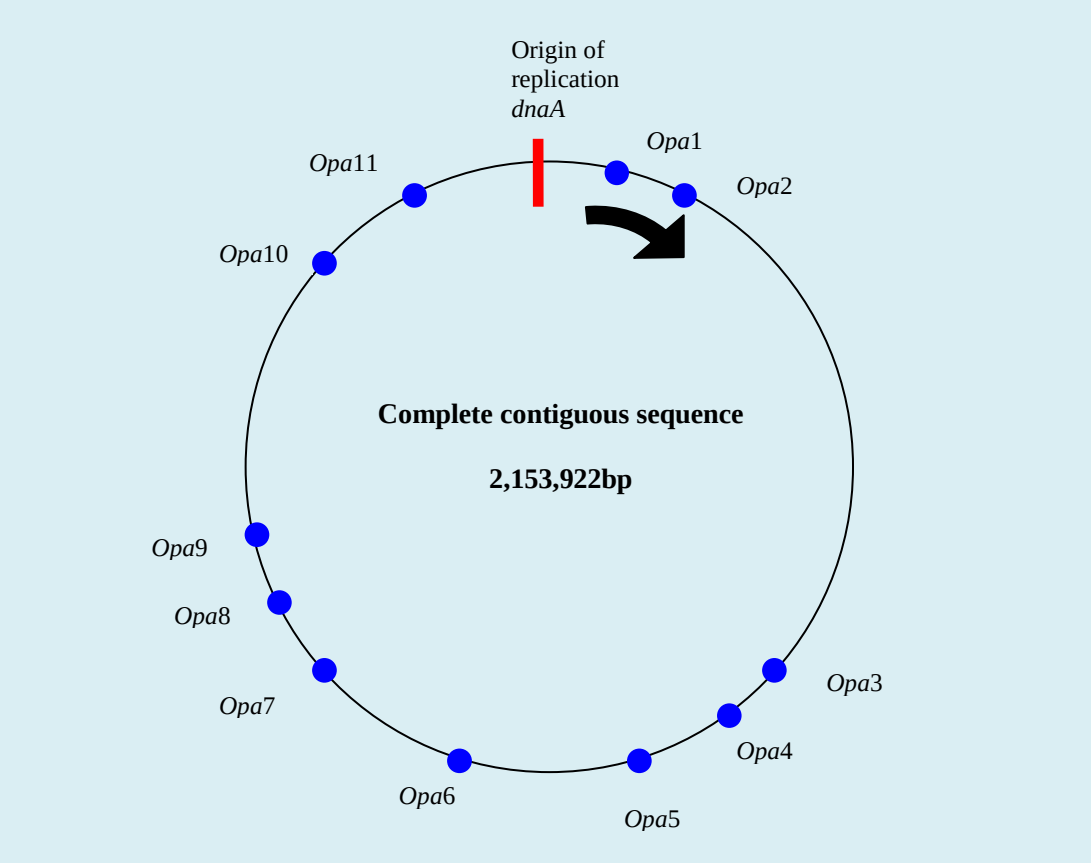


Figure 3.1 Diagrammatic representation of the single contiguous genome sequence of *N. gonorrhoeae* strain FA1090. The approximate location of each *opa* locus in the genome sequence is shown.

Genes common to *N. gonorrhoeae* and *N. meningitidis* appear to be highly conserved with 90% sequence identity typically observed (107). Therefore, the genes that flanked each *opa* locus were also explored and compared to the homologue present in *N. meningitidis* strain MC58 and Z2491 whose sequence information was obtained at http://www.tigr.org/. This allowed identification of the most conserved regions of the flanking genes in which primers pairs could be designed to amplify each *opa* gene independently by PCR (See Chapter 2 and Appendix 1).

Gonococcal *opa* genes are on average 800 – 900bp and so it was not possible to sequence the complete open reading frame in a single pair of reactions for each strand. Instead, it was necessary to sequence separate sections which overlapped and could thereafter be assembled into the complete gene sequence. Each *opa* gene was initially sequenced with the *opa*-typing primers, in both orientations, to explore the regions immediately flanking the *opa* genes, as well as the central regions, in an attempt to identify conserved regions within which further primers could be designed for subsequent sequencing reactions. Figure 3.2 shows the position of the primers and the regions they provide sequence data for.

The above reactions were optimised using a set of unrelated gonococci (all belonging to different STs by NG-MAST) to ensure the primers and parameters could be used on diverse strains. These isolates included strain FA1090 (provided by N. Saunders, University of Oxford) as a control, seven isolates from a study in London (38, 175), three from a study in Sheffield (134), one from a study in Lincoln (unpublished), one from Greece (provided by A. Mavroidi, Imperial College London) and one from Italy (provided by P. Stefanelli, Istituto Superiore di Sanità, Italy).

All sequences were manipulated in MEGA (see Chapter 2), as was MEGA used to calculate the diversity by the proportion of nucleotide differences (p-values).

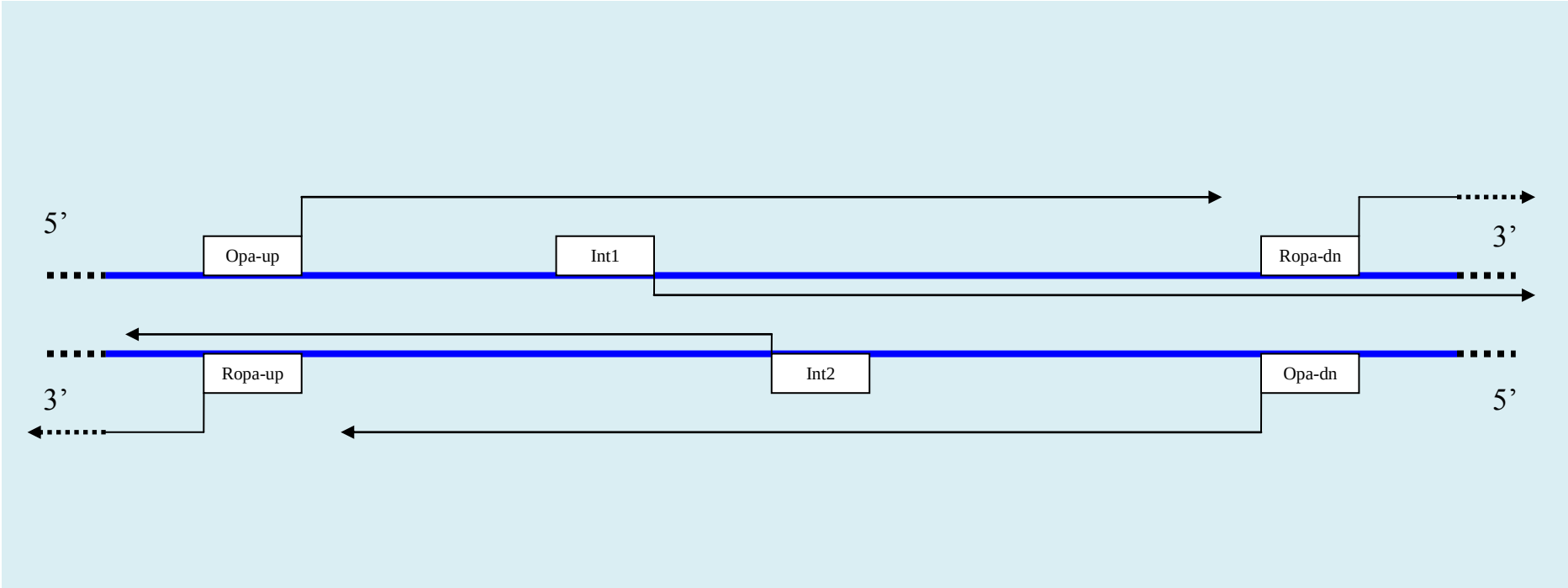


Figure 3.2 Diagrammatic representation of an *opa* gene showing the location and orientation of the sequencing primers on each strand. Blue regions represent the *opa* gene on each strand with black arrows indicating the direction and the regions for which the primers provide sequence data. Not to scale**.**

3.3 Results

3.3.1 Preliminary sequencing data

The sequence data obtained with the *opa*-typing primers for the central region of the *opa* genes allowed identification of two central internal primer sequences (int1 and int2, see Chapter 2). These primers were orientated to sequence outwards towards the start and end of the genes whilst crossing each other to provide overlap (Figure 3.2) and could be used for all gonococcal *opa* genes. The *opa*-typing primers were also used in reverse orientation to explore the regions that immediately flank each *opa* gene in an attempt to identify sequences in which external primers could be designed to sequence inwards and complement the sequence obtained from the internal primers. It was not possible to design unique yet broadly applicable external sequencing primers because the majority of sequence in these regions was either repetitive or not sufficiently conserved amongst the test isolates in order to be used with distinct strains. Moreover, any external sequencing primers would need to be within a certain distance from each *opa* gene to provide the sequence data for the target region and so it would not have been useful to consider the sequence further upstream or downstream. Therefore in order to keep the method simple and suitable for all 11 *opa* loci, both central primers were used in combination with the *opa*-typing primers (in both orientations) in order obtain the complete open reading frame of each *opa* gene.

Comparison of *de novo* FA1090 *opa* gene sequences and those published showed variation in the number of CRs at three loci, but were otherwise identical. Since the

relationship of the two cultures of this strain was unknown, another test isolate, other than FA1090, was re-cultured and the *opa* genes re-analysed. Variation in the multiple of CRs was again observed but occurred at other *opa* loci. The multiple of CRs were therefore adjusted to a multiple of three so that the sequences should be in-frame, however, several of the sequences still had premature termination codons. Closer examination of the leader sequence identified single nucleotide indels in a polyA region just before the CRs that also disrupt the reading frame. In order to avoid problems arising from variation in the polyA tract and the number of CRs in the leader sequence, the sequences were trimmed just after the CRs and after the last nucleotide of the stop codon (see Chapter 2).

3.3.2 *opa* gene sequences amongst diverse isolates

The *opa* gene sequences of the 14 unrelated isolates at all 11 *opa* loci were trimmed and compared and several sequences were found to be identical. The sequences that were identical were not present in different isolates but at different loci within the same isolate. This observation led to the decision to assign allele numbers to each unique sequence, irrespective of the locus at which it is found. In 10/14 (71%) isolates, at least one allele was present at more than one locus, two isolates had the same *opa* allele at three loci and another had the same allele at four loci. Three isolates had multiple copies of two distinct alleles. No pattern was detected that suggested particular loci are more likely to share alleles. Table 3.1 shows the alleles identified at each locus of the 14 diverse isolates.

Table 3.1 *opa* gene profiles of 14 unrelated isolates of *N. gonorrhoeae*

Isolate	ST	*por*	*tbpb*	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
FA1090	773	514	219	8	15	24	27	35	43	52	59	24	73	15
LG0540	610	39	2	204	151	201	196	158	145	193	169	181	176	169
S58	12	8	5	1	10	16	25	29	36	45	54	†	67	29
S89	261	197	11	2	11	17	17	30	37	46	96	62	68	123
LG1930	825	558	21	199	148	199	195	156	142	190	142	179	173	142
LG0159	584	7	27	207	154	171	198	164	147	183	163	184	178	231
L6	302	238	29	7	13	23	13	13	42	51	58	65	71	13
LG0923	709	479	171	205	153	202	197	160	146	194	188	182	177	225
S242	232	98	50	4	12	19	267	32	39	48	56	290	70	#
LG0906	876	59	166	168	150	200	150	157	144	192	185	180	175	168
G181	662	451	192	255	258	262	266	266	275	280	284	289	294	300
LG0617	1063	474	281	257	#	264	270	274	257	282	286	292	296	302
LG1461	1131	759	291	256	259	263	269	273	276	281	285	291	295	301
T0679/05	2707	1605	632	265	260	265	271	265	277	283	288	293	297	303

Coloured alleles indicate the same allele at multiple loci.

#, PCR product smaller than expected and all primers designed from *opa* genes fail to give sequence.

†, 5' region of *opa* gene present but gene truncated.

The diversity within the *opa* repertoire of each was calculated as a measure of the proportion of nucleotide differences (p-value). Table 3.2 shows these p-values which range from 0.114 to 0.147. The diversity calculated for all the *opa* alleles as one group was 0.130.

In isolate S58, only one primer sequence yielded sequence data for the *opa9* locus (opa-up in the 5' region). Although the sequence obtained with this primer was initially that of an *opa* gene, it became unidentifiable sequence approximately 250bp downstream of the initiation codon, indicating that the *opa* gene at this locus was truncated. In another two isolates, S242 and LG0617, the sizes of the PCR products generated for *opa11* and *opa2*, respectively, were considerably less than expected. Since all the *opa* sequencing primers failed to give sequence from the PCR products, it is most probable that no *opa* gene was present at these loci.

Table 3.2 p-values as a measure of diversity for the *opa* alleles of each diverse isolate

Isolate	Sequence Type	Diversity (p-distance)
LG0617	1063	0.104
FA1090	773	0.114
LG0540	610	0.114
S89	261	0.114
LG0159	584	0.114
LG0923	709	0.115
S242	232	0.115
LG1930	825	0.119
S58	12	0.122
LG0906	876	0.123
G181	662	0.123
T0679/05	2707	0.131
L6	302	0.138
LG1461	1131	0.147
Diversity for all isolates		0.130

3.4 Discussion

The nomenclature for *opa* genes was developed before the complete genome sequence of strain FA1090 was available and needs to be rationalised. To name the *opa* genes consecutively according to their location (clockwise) around the chromosome of FA1090 seemed reasonable, however, inversions and large scale genomic events may result in the rearrangement of loci (69, 200, 220). Therefore, in further isolates, the *opa* loci are named according to the homology of flanking genes, rather than a specific order on the chromosome. As the initial amplification of each *opa* gene uses primers in the conserved flanking genes, they can be assigned to the current *opa* locus even if the gene order differs.

The conserved regions of the *opa* genes proved to be useful, allowing primers to be designed that could be used on all loci, irrespective of locus. The key to this approach was to ensure that only one *opa* gene was present in each amplicon. Since the nature of the DNA sequence immediately flanking the genes was not suitable for designing sequencing primers, a combination of primers based on *opa* gene sequences was used. Consequently, short stretches at each end of the *opa* genes were not sequenced on both strands of DNA, however, they were sequenced off two independent primers. This compromise seemed acceptable since sequencing is now very accurate. The HVR1 and HVR2, which were expected to be the most variable, are central and therefore sequenced on both strands.

Strain FA1090 was included as a control in order to validate amplification of the correct *opa* genes. Comparison of the *de novo* FA1090 *opa* gene sequences with that of the *opa* genes in the complete genome sequence, confirmed this, however,

variation in the multiple of CRs was observed. This was not entirely unexpected since it is possible that this strain has been subjected to intensive sub-culturing during which these frame shift mutations would be expected to occur. This was tested in one other isolate and observed to occur again although at different *opa* loci. In order to make comparisons easier, the CRs were adjusted to all have a multiple of three so that the genes would be in-frame, however, a poly A tract which precedes the CRs was, as previously reported (13), also seen to undergo single nucleotide indels which disrupt the reading frame. Therefore, to avoid complicating sequence comparison it seemed necessary to trim the sequences to exclude this region (see Chapter 2). This phenomenon was reported amongst meningococcal *opa* genes and a similar approach to sequence editing was taken (93). Additionally, the fact that the excluded region does not encode any surface exposed loops further ratifies the decision to do so.

Once the way in which the sequence data would be handled was established, a closer look at the sequences was taken. No sequences were shared by any of the unrelated strains, highlighting the diversity in this family of genes, which has also been seen amongst meningococci (30). This diversity is further emphasised by the proportion of nucleotide differences amongst *opa* alleles within the same isolate which are comparable to that of the entire set of unrelated isolates. The values suggest that there may be as much diversity among *opa* alleles within an isolate as between unrelated strains but also that the amount of diversity will vary between strains. The values vary, maybe as a consequence of varying amounts of pressure from the host's immune system.

The most striking feature was the presence of identical sequences at multiple loci, occurring in 71% of these isolates, and has also been reported amongst meningococci (149) and one other gonococcal strain, JS3 (228). These allele duplications could be an inevitable result of gene conversion events but it is tempting to believe that such events are favoured under some circumstances and occur as a result of selection. No pattern was detected that suggested particular loci were more likely to share the same sequence.

For the reason that identical sequences were present at multiple loci, allele numbers were assigned to unique sequences irrespective of the *opa* locus such that this feature could be easily identified. The *opa* profile of an isolate is therefore defined as a series of 11 integers which correspond to the alleles present at each locus, whereas the *opa* gene repertoire refers to the collection of *opa* genes (alleles), irrespective of locus. Two isolates appeared to have lost an *opa* gene and one isolate had a truncated *opa* gene, all at different loci which demonstrates that these types of events commonly occur although we do not know how often they do so *in vivo*.

In conclusion, there is great diversity amongst *opa* genes of unrelated strains. The sequencing method developed will aid further studies on the degree of diversity within strains and the relative contributions of mutation and recombination to the generation of diversity.

Chapter 4: Mechanisms of *opa* gene variation amongst closely related isolates in Sheffield

4.1 Introduction

The previous two chapters describe a method that allows each of the 11 *opa* genes from isolates of *N. gonorrhoeae* to be analysed. Therefore, the next step in investigating the mechanisms of variation was to examine the *opa* gene profiles of a set of appropriate isolates. The lack of any shared alleles amongst the 14 unrelated isolates in Chapter 3 suggests great diversity amongst *opa* sequences in the gonococcal population and therefore, in order to understand the mechanisms that cause variation and the rates of change at each locus, it is necessary to examine gonococci that are closely related. Such strains will have recent shared ancestry so would be expected to have similar *opa* profiles such that any changes that are observed will be recent and can be classified as due to point mutation or recombination events.

If *opa* genes diverge rapidly during transmission, different *opa*-types (by *opa*-typing) should occasionally be seen in isolates from sexual contacts and we would expect most gonococci with the same *opa*-type to be recovered within a short period of time (160). If change in *opa* genes comes about largely by the introduction of new *opa* gene sequences from co-infecting strains, then although sexual mixing between partners may be frequent, the diversity in the *opa* genotypes may be limited due to a lack of mixed infections. Similarly, in a community or dense sexual network in which new strains are frequently introduced (such as a metropolitan city), *opa*-types

may be seen to change more rapidly, due to the increased likelihood of mixed infections.

To this end, advantage was taken of the two dominant strains circulating in Sheffield which persisted over a two year period. These two prevalent strains were previously characterised as *opa*-type 1 and *opa*-type 4 by *opa*-typing, together comprising 41% of all isolates studied (134). Subsequent work on these isolates has shown that they almost always correspond to ST12 and ST261 by NG-MAST, respectively (16) and we can therefore be confident that these isolates are closely related.

4.2 Methods

4.2.1 Isolates

Isolates from cases of gonorrhoea, collected in Sheffield between April 1995 and March 1997, were available from the study of Martin *et al*. (134), along with additional isolates from Sheffield collected up to December 1997 from the same study. Many of these isolates have been characterised by both *opa*-typing (134) and NG-MAST (16) and detailed sexual contact tracing data are available. Initially, a set of isolates from each of the two main strains were analysed that were chosen to be representative of the sampling period. This was followed by the analysis of isolates obtained from mutually named sexual contacts collected within one month of each other. See Chapter 2 for details on bacterial growth and DNA preparation. Amplification and sequencing of *opa* genes was also performed as described in Chapter 2.

4.2.2 *In silico opa*-typing

opa gene sequences were trimmed to the expected size of the fragments amplified during *opa*-typing and submitted to the restrict tool via the EMBOSS interface available through the Imperial College London Bioinformatics Support Service (http://www-bss1.bioinformatics.imperial.ac.uk/codon.bioinformatics/index.html). This generated a list of fragments expected to be generated from each locus when digested with the restriction endonuclease *Taq*I, as used in the original study (134).

The fragment sizes could then be combined to give an indication of the pattern that would be generated via traditional *opa*-typing.

4.3 Results

4.3.1 *opa* gene profiles of isolates with the same sequence type

Initially, the *opa* profile of 21 ST12/*opa*-type 1 and 27 ST261/*opa*-type 4 isolates that were chosen to represent the sampling period were characterised in order to determine whether a consensus *opa* gene profile could be identified. Because these isolates all share the same NG-MAST ST as well as *opa*-type, they were expected to have identical or similar *opa* profiles. Table 4.1a and b show the *opa* allelic profiles for each of these groups. Although these two strains were co-circulating in Sheffield, no alleles were shared between any isolates of the two different STs.

Excepting *opa11*, 57-100% and 63-100% of the alleles at individual loci were the same in the ST12 and ST261 isolates, allowing consensus alleles at each locus, and a consensus *opa* allelic profile, to be identified. At *opa11* almost every isolate had a different allele, some of which corresponded to alleles present at other loci, suggesting variation at this locus is far more rapid than at the other loci. Although a consensus *opa* allelic profile could be identified for both STs, 48% of ST12/*opa*-type 1 and 81% of ST261/*opa*-type 4 isolates departed from the consensus profile in at least one locus, other than *opa11*. As found with the diverse strains in Chapter 3, identical *opa* alleles were observed at multiple loci within many isolates.

Assuming each unique difference from the consensus profile (excluding *opa11* where no consensus allele was identified) is generated only once, eight changes to the ST12/*opa*-type 1 consensus profile were identified. Three changes were due to a

consensus allele at one locus being duplicated whereas the other five generated a new allele. Excluding *opa11,* 19 changes to the ST261/*opa*-type 4 consensus profile were observed; only one appeared to be due to the duplication of a consensus allele, the others being new alleles.

Table 4.1a o*pa* gene profiles of 21 ST12/*opa*-type 1 isolates chosen to be representative of the sampling period

	Locus										
Isolate	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus	1	10	16	25	29	36	45	54	†	67	V
S7	54	25			$_{NS}97^{29}$						~~100~~
S58											29
S78											~~117~~
S81											29
S86											~~118~~
S98											~~120~~^{C}
S100											~~121~~
S149		25			~~128~~	$_{NS}99^{36}$					~~141~~
S151											54
S152		25			~~128~~	$_{NS}99^{36}$					~~140~~^{C}
S154		25			~~128~~	$_{NS}99^{36}$					54
S184											~~101~~
S241											29
S256											~~119~~
S276						$_{NS}99^{36}$					29
S280						~~98~~					~~238~~
S302						$_{NS}99^{36}$					25
S306						$_{NS}99^{36}$					~~253~~
S308					$_{NS}137^{29}$	$_{NS}99^{36}$					~~234~~^{C}
S350											~~102~~^{C}
S427				1		$_{NS}99^{36}$					54

Coloured alleles indicate the presence of a consensus allele at a locus and the same allele at another locus; uncoloured alleles are non-consensus alleles. Colours in Table 4.1a and 4.1b do not correspond to the same alleles.

†, 5' region of *opa* gene present but gene truncated; V, variable.

$_{NS}97^{29}$ The allele number is shown followed by a superscript that denotes the allele from which it differs by a single synonymous (S) or non-synonymous (NS) substitution, or an insertion (INS) or deletion (DEL), all of which were 3bp (one codon).

$_{NS}99^{36}$ As above except the difference from the allele shown by the superscript could be the result of a recombination event or a point mutation.

~~128~~ A diagonal line through the allele number indicates that the allele is a simple hybrid/mosaic of *opa* alleles present within the strain.

~~120~~^{C} An superscripted 'C' denotes a more complex mosaic which may involve multiple recombination events.

Table 4.1b o*pa* gene profiles of 27 ST261/*opa*-type 4 isolates chosen to be representative of the sampling period

Isolate \ Locus	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus	2	11	17	17	30	37	46	96	62	68	V
S3								55^{C}			80
S21											96
S24							$_{DEL\&NS}113^{46}$				96
S26							$_{DEL\&NS}113^{46}$				68
S29											11
S69		75									$_{S}124^{110}$
S70		75									$_{S}124^{110}$
S74		75									125^{C}
S75		75									75
S89											$_{NS}123^{96}$
S94											11
S107											17
S117								17			109
S121							$_{INS}95^{46}$				68
S123							$_{INS}95^{46}$				11
S128								$_{NS}134^{239}$	$_{NS}136^{62}$		127
S129								$_{NS}134^{239}$	$_{NS}136^{62}$		127
S136				111^{C}				135			111^{C}
S137				111^{C}				111^{C}			111^{C}
S140							$_{INS}95^{46}$				108
S156		80			$_{NS}81^{82}$			$_{NS}82^{81}$			80
S196								94^{C}			75
S200							$_{INS}95^{46}$				11
S233							$_{INS}95^{46}$	$_{NS}138^{133}$			17
S243	103			106			$_{INS}95^{46}$				$_{S}110^{124}$
S300								$_{NS}133^{138}$			122
S324		251					$_{INS}95^{46}$				$_{NS}239^{68\&134}$

As in previous table. Colours in Table 4.1a and 4.1b do not correspond to the same alleles.

4.3.2 *In silico opa*-typing of ST12/*opa*-type 1 and ST261/*opa*-type 4 isolates

The numerous differences in the *opa* allelic profiles of the 21 ST12/*opa*-type 1 and 27 ST261/*opa*-type 4 isolates were unexpected given the fact they were assigned to the same *opa*-type and were thus explored further. The *opa* gene sequences of these isolates were subjected to *in silico opa*-typing. Amongst the ST12/*opa*-type 1 isolates there were 17 different *opa* repertoires and 18 amongst the ST261/*opa*-type 4 isolates. Tables 4.2a and b lists the fragments which would be generated by *opa*–typing for each unique repertoire as well as the list of fragments from the consensus profiles which exclude an allele at *opa11* where no consensus was identified. All except one of the ST12/*opa*-type 1 repertoires had the same fragment pattern. The repertoire that was not identical differed only in two fragments by 6bp and 3bp. Amongst the ST261/*opa*-type 4 isolates, nine of the 18 repertoires were different, although these were all minor differences. One repertoire had one fragment which was smaller by 3bp; a further five repertoires had a fragment larger by 3bp. Another five repertoires had an additional fragment, 3bp larger than one already present in all the ST261/*opa*-type 4 repertoires. These differences would not have been detectable by *opa*-typing.

4.3.3 *opa* gene profiles of isolates from known sexual contacts

Considerable *opa* variation was found amongst the closely related isolates with the same ST/*opa*-type circulating in the same city over a 33 month period. Isolates that were even more closely related were therefore examined by focusing on all of the ST12 and ST261 isolates from Sheffield that were obtained within one month of

each other from mutually named sexual contacts. In this case the selection of strains was blind of any available *opa*-typing results, to avoid selection bias for or against differences in their *opa* repertoires. Three triplets and 18 pairs of isolates (45 isolates in total) that fulfilled these criteria were identified and characterised. Tables 4.3a and b show the *opa* profiles of these isolates in relation to the consensus *opa* profile determined for ST12 and ST261.

Despite the use of isolates from recent sexual contacts, 14 of the 18 pairs, as well as all three triplets, differed at *opa11* supporting the previous observation that the rate of variation at this locus is much higher than at other loci. At the other ten *opa* loci, several of the pairs and triplets differed, as expected, from the ST12 or ST261 consensus *opa* allelic profile, but differences in the profile between isolates from recent sexual contacts was only seen in one triplet and three pairs (shown within a box in Table 4.3a and b).

Table 4.2a *In silico opa*-typing fragments generated with *Taq* I for the 17 *opa* repertoires observed amongst the 21 ST12/*opa*-type 1 isolates

Repertoire	12_C*	21_1	12_2	12_3	12_4	12_5	12_6	12_7	12_8	12_9	12_10	12_11	12_12	12_13	12_14	12_15	12_16	12_17
	322	322	322	322	322	322	322	322	322	322	322	322	328	322	322	322	322	322
	319	319	319	319	319	319	319	319	319	319	319	319	322	319	319	319	319	319
	271	271	271	271	271	271	271	271	271	271	271	271	271	271	271	271	271	271
	259	259	259	259	259	259	259	259	259	259	259	259	259	259	259	259	259	259
	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225
	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222
	186	186	186	186	186	186	186	186	186	186	186	186	186	186	186	186	186	186
	154	154	154	154	154	154	154	154	154	154	154	154	154	154	154	154	154	154
	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145
	136	136	136	136	136	136	136	136	136	136	136	136	136	136	136	136	136	136
	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130
	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

C* is the consensus *opa* repertoire (excluding *opa11* where no consensus allele was identified). Fragment sizes shown in number of base pairs.

Table 4.2b *In silico opa* -typing fragments generated with *Taq* I for the 18 *opa* repertoires observed amongst the 27 ST261/*opa* -type 4 isolates

Repertoire	261_C*	261_1	261_2	261_3	261_4	261_5	261_6	261_7	261_8	261_9	261_10	261_11	261_12	261_13	261_14	261_15	261_16	261_17	261_18
	556	556	556	556	556	556	556	556	556	556	556	556	556	556	556	556	556	556	556
	484	484	484	484	484	484	484	484	484	484	484	484	484	484	484	484	484	484	484
				343											343	343	343	343	
	340	340	340	340	340	340	340	340	340	340	340	340	340	340	340	340	340	340	340
	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322	322
	304	304	301	304	304	304	304	304	307	304	304	304	307	304	304	307	307	304	307
	274	274	274	274	274	274	274	274	274	274	274	274	274	274	274	274	274	274	274
	244	244	244	244	244	244	244	244	244	244	244	244	244	244	244	244	244	244	244
	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225	225
	151	151	151	151	151	151	151	151	151	151	151	151	151	151	151	151	151	151	151
	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145	145
	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

As in Table 4.2a.

Table 4.3a *opa* gene profiles of ST12 isolates from recent sexual contacts

		Locus										
Pair/ triplet	Isolate	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST12 Consensus		1	10	16	25	29	36	45	54	†	67	V
Triplet1	**S149**		25			128	$_{NS}99^{36}$					141
	S152		25			128	$_{NS}99^{36}$					140^{C}
	S154		25			128	$_{NS}99^{36}$					54
Triplet 2	S306						$_{NS}99^{36}$					253
	S308					$_{NS}137^{29}$	$_{NS}99^{36}$					234^{C}
	S317					$_{NS}129^{29}$	$_{NS}99^{36}$					233
Triplet 3	**S445**						130^{C}					114
	S463						130^{C}					115^{C}
	S466						130^{C}					67
Pair 1	S78											117
	S86											118
Pair 2	S98											120^{C}
	S100											121
Pair 3	S241											29
	S256											119
Pair 4	S253		25				132	$_{NS}112^{45}$				54
	S257		25				132	$_{NS}112^{45}$				25
Pair 5	S493						131^{C}					116^{C}
	S514						131^{C}					116^{C}
Pair 6	S517						$_{NS}99^{36}$					1
	S522						$_{NS}99^{36}$					25

As in Tables 4.1a and b. Bold isolate numbers indicate the isolate from the individual who named the other two individuals in the triplet as recent sexual contacts. Blocked alleles indicate differences in *opa* gene profiles of isolates from recent sexual contacts.

Table 4.3b *opa* gene profiles of ST261 isolates from recent sexual contacts

Pair/ triplet	Isolate	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST261 Consensus		2	11	17	17	30	37	46	96	62	68	V
Pair 7	S21											96
	S29											11
Pair 8	S24							DEL&NS 113[46]				96
	S26							DEL&NS 113[46]				68
Pair 9	S69		75									S 124[110]
	S70		75									S 124[110]
Pair 10	S74		75									125[C]
	S75		75									75
Pair 11	S89											NS 123[96]
	S107											17
Pair 12	S121							INS 95[46]				68
	S123							INS 95[46]				11
Pair 13	S128								NS 134[239]	NS 136[62]		127
	S129								NS 134[239]	NS 136[62]		127
Pair 14	S136				111[C]				135			111[C]
	S137				111[C]				111[C]			111[C]
Pair 15	S156		80			NS 81[82]			NS 82[81]			80
	S162		80			NS 81[82]			NS 82[81]			83[C]
Pair 16	S192								94[C]			236
	S196								94[C]			75
Pair 17	S266								NS 133[138]			103
	S271								139			252[C]
Pair 18	S254							INS 95[46]	NS 133[138]			108
	S258								NS 133[138]			NS 237[96]

As in Table 4.1a and b. Blocked alleles indicate differences in *opa* gene profiles of isolates from recent sexual contacts.

4.3.4 Assigning allelic change in *opa* genes to recombination or point mutation

Isolates that are very closely related would typically have the same allele at each of the *opa* loci and, in those isolates where the allele at a single locus has changed, one can be confident that this has occurred by a single molecular event and can attempt to determine the nature of this event. Changes away from a consensus allele could be due to one of a number of mechanisms which can be broadly divided into those involving recombination and those due to point mutation.

Several classes of recombinational event can be distinguished, the simplest being a change of the consensus allele to an allele present at another locus in the same isolate, resulting in allele duplication. Recombination can also result in the replacement of only part of the consensus allele with the corresponding region of an allele present at another locus in the isolate, resulting in a simple mosaic allele. In other cases more complex mosaics were observed, where the new allele was composed of parts of two or more alleles present in the same isolate, but which would have required more than one localised recombinational event. The above examples of recombinational changes could occur by gene conversion within a single bacterial cell or, following uptake of *opa* sequences, via natural transformation from sibling cells (these alternatives cannot be distinguished), and are all classified as recombination among existing alleles. In contrast, a new allele could be a sequence that is not already present at another *opa* locus in the isolate, and cannot be formed as a mosaic of the existing alleles in the isolate, and therefore presumably represents the introduction of a novel *opa* allele from a co-infecting strain.

Point mutations in *opa* genes were recognised as alleles that differed from the consensus allele at a single nucleotide site, or in some cases by a single codon indel. Where these substitutions were unique amongst the *opa* genes of a set of isolates, rather than a polymorphism present in other alleles, they were assigned as the result of *de novo* mutation. All of these unique mutations were NS substitutions within HVR1 or HVR2, which correspond to surface-exposed variable loops. Where the single nucleotide changes were not unique, these included both synonymous and NS changes in conserved regions, excepting one within a HVR. As these latter nucleotide changes were present in other *opa* alleles within a set of closely related isolates (i.e. they were polymorphisms), they could have arisen by highly localised recombination that replaced a region of a consensus allele with a small region that differed only at the single nucleotide site, rather than by point mutation. As the appearance of unique point mutations was rare, compared to changes by recombination (see below), most of the changes at sites that were polymorphic were considered more likely to be due to localised recombination rather than point mutation, and were assigned as such.

To evaluate the contribution of mutation and recombination to changes in the *opa* profile, the isolates from recent sexual contacts in Sheffield were initially focused on, to ensure that most allelic changes were due to a single molecular event rather than multiple sequential changes. The problem of multiple sequential changes was particularly problematic at the *opa11* locus, within which changes occurred much more rapidly than at the other *opa* loci. No attempt was therefore made to assign the mechanisms by which allelic changes at *opa11* occurred. Not all isolates of ST12 and ST261 shown in Table 4.1a and b were included as these were selected for being

the same by both NG-MAST and *opa*-typing, and their inclusion could underestimate the contribution of recombination, which would be much more likely to produce changes in *opa*-type than point mutations.

The mechanisms of allelic change among isolates from these sexual contacts were assigned as follows. The changes in triplet 2 (ST12 isolates) occurred at *opa5,* where one isolate had the consensus allele 29 and the other two each had a different allele (Table 4.3a). Alleles 129 and 137 both differed from the consensus allele by single NS substitutions in HVR2. However, the substitution in allele 137 was unique, and thus assigned to mutation, whereas that in 129 was also present in the allele at *opa6* (allele 99), and this change was assigned to a small localised recombinational event rather than point mutation.

The difference in pair 14 (ST261 isolates) occurred at *opa8* and neither had the consensus allele (Table 4.3b). Allele 111 was also present at both *opa4* and *opa11,* and is assumed to be the ancestral allele at *opa8* from which allele 135 arose. Allele 135 is a simple mosaic of allele 111 and allele 46 (the consensus allele at *opa7*) and the allelic change in pair 14 was therefore assigned to recombination.

The two isolates in pair 17 also differed at *opa8* and again neither had the consensus allele (Table 4.3b). Allele 133 is considered to be the ancestral allele in this case as it is present at the same locus in three other ST261 isolates, including pair 18, whereas allele 139 was a unique allele. Allele 139 can be derived as a simple mosaic of allele 133 and allele 17, the consensus allele at both *opa3* and *opa4*, and was thus assumed to have arisen by recombination.

Pair 18 differed at *opa7* (Table 4.3b). One isolate had the consensus allele (allele 46) and the other had allele 95, which differs from the consensus by a codon insertion at the end of HVR1. Since the codon insertion was not found in other alleles, allele 95 almost certainly originally arose from the consensus allele by mutation. However, allele 95 was present at *opa7* in ST261 isolates that were recovered up to eight months before the isolates of pair 18, and the codon insertion is unlikely to have occurred as a de novo mutation in this pair. It is more likely that the derived allele (which could be either allele 46 or 95) was introduced from another isolate of ST261 during co-infection and the allele change within this pair was therefore assigned to recombination.

Therefore, of the five allelic changes in the isolates from recent sexual contacts in Sheffield, one was assigned to point mutation and four to recombination, three of the latter involving recombination among alleles within the same isolate, and the one introduction of an allele from another co-infecting isolate of the same strain.

4.4 Discussion

Allele sharing due to horizontal exchange at times of mixed infections might be expected among strains circulating in the same location at the same time and so it was surprising to find that no alleles were shared by the two dominant strains that were circulating in Sheffield at the same time. A previous study in Sheffield found cases of these two strains had characteristics of endemic infection in that they were concentrated in the city centre and in people with few links outside Sheffield or abroad such that the cases were assumed to be linked by ongoing local transmission (240). It is not clear what kept these two strains separate; most likely, the restricted (or assortative) sexual mixing believed to have contributed to their persistence may also have played a role in confining them to apparent separate subsets of the Sheffield population (134).

The *opa* allelic profiles of ST12/*opa*-type 1 and ST261/*opa*-type 4 isolates were very similar such that a consensus profile could be identified at 10/11 loci. At *opa11*, consensus alleles could not be identified. Duplicated alleles were a common feature as reported in the previous chapter (Chapter 3) and since these do not actually contribute to the antigenic diversity, it may be that the Opa protein encoded is advantageous in function. All these isolates were identical by *opa*-typing which is based on the *opa* gene sequences and the *opa* gene allelic profiles were expected to be close to if not entirely identical. *Opa*-typing should detect differences in *opa* repertoires but not necessarily *opa* profiles as the method uses a combination of all the loci regardless of their position. *In silico opa*-typing was performed using the *opa* gene sequences available. It transpired that the restriction fragment patterns that

would be generated were in fact almost identical. This may be the case if the restriction endonuclease target sequences have not been changed and if the *opa* gene sequences themselves have similarities. The few differences in the patterns that were generated were only ever a few base pairs and these would not have been detectable by traditional *opa*-typing. Morris *et al.* recently reported the ability of *opa*-typing to discriminate within some NG-MAST clusters (151). The results presented here suggest that re-assortative changes in *opa* profiles will result in isolates having very similar *opa*-types. However, this may not be the case if the changes involve accumulation of novel and diverse *opa* gene sequences. This in turn will depend on the nature of the sexual networks in question and the contribution of mixed infections to diversity within NG-MAST clusters.

There were a considerable number of changes from the consensus amongst the ST12/*opa*-type 1 and ST261/*opa*-type 4 isolates, and it was also not always clear which were the ancestral or recently derived alleles. Isolates that were even more closely related were used as the direction of change could usually be identified and assigning mechanisms could be done with more confidence. The differences at *opa11* were not included here as the rate of variation at this locus was enough that the alleles frequently differ even amongst direct sexual contacts. Amongst these sexual contacts, five recent changes were identified, one of which was due to a NS point mutation in a HVR. There were four assigned to recombination, one being an example of importation from a co-infecting isolate and the remaining three differences were assigned as simple recombination events amongst existing alleles. The proportion of recombination to point mutation here certainly suggests the predominance of recombination events however the number of events identified was

too few to make any firm assumptions at this point. In the following chapter, sexual networks in London are explored to add to this analysis.

Chapter 5: Mechanisms of *opa* gene variation amongst closely related isolates in London

5.1 Introduction

The large ST12 and ST261 sexual networks in Sheffield were heterosexual and relatively few strains were circulating within this city during the sampling period. To include both MSM and heterosexual networks from an epidemiological setting in which many more strains are circulating, gonococcal isolates from London were analysed. In a previous study most isolates from cases of gonorrhoea in London between June and November 2004 were analysed by NG-MAST, and groups of individuals that shared the same ST, and had similar behavioural and demographic characteristics, were considered to be within the same sexual network (38, 175).

5.2 Methods

5.2.1 Isolates

Isolates were from the study of Choudhury *et al*. (38, 175). The *opa* allelic profiles of isolates from individuals within three small heterosexual networks, three small MSM networks and one of mixed sexual orientation were examined. To focus on isolates that are likely to be the most closely related, only sexual networks in which all isolates were recovered within a period of three months were included i.e. they had not been identified before or after the time period at the time of selection. To avoid inadvertently choosing networks that were related, the STs chosen had to have distinct *por* and *tbpB* alleles by NG-MAST. All sequencing was done as described in Chapter 2.

5.3 Results

5.3.1 General observations

The *opa* profiles of isolates from each of these London networks are shown in Table 5.3. As observed in other strains, the same consensus allele was present at more than one locus in one of the three heterosexual networks and all three MSM networks. Atypically, there was no consensus allele at *opa8* in the ST610 heterosexual network. Furthermore, in two MSM (ST802 and ST876) and one heterosexual network (ST777), a consensus allele could be identified at *opa11*; the time frame in which these isolates were collected was similar or greater than that in which the isolates from sexual contacts in Sheffield were obtained, which suggests the apparent stability of *opa11* in these London networks is not due to a shorter time between recovery of the isolates.

Isolates of the two dominant strains in Sheffield (see Chapter 4) were not seen to share any *opa* alleles and, similarly, the strains in the three MSM networks in London did not share any alleles. However, six *opa* alleles were shared between two heterosexual and the mixed sexual orientation networks. Five of these were shared between isolates in the ST777 and ST584 sexual networks, which showed substantial similarities in their *opa* profiles that are almost certainly due to recent shared ancestry, although they have diverged sufficiently to have different alleles at both *porB* and *tbpB*. Shared ancestry was supported by the fact that isolates of these two NG-MAST STs are identical by MLST. The consensus alleles at nine of the eleven loci of ST777 are also consensus alleles in ST584 and, in five cases, these shared

alleles are at the same *opa* locus. Similarly, six consensus alleles in the ST584 network were in common with ST777. The other shared allele (allele 183) was found at *opa7* in one isolate of the ST610 network and at the same locus in all isolates of the ST584/777 networks. ST610 and ST584/777 share 5/7 MLST loci and the sharing of allele 183 could be due to relatively distant common ancestry rather than introduction from another strain. The fact that this allele was only present in the most recently recovered isolate of ST610, argues in favour of importation. The changes in the London sexual network isolates were analysed in the same way as the pairs and triplets in Sheffield (Chapter 4) and each network is described in more detail below.

5.3.2 Variations observed in each London network

MSM network ST802 (Table 5.3a) had five isolates, collected over a period of just over two months, and all had the same *opa* profile. This ST does not share any *opa* alleles with the other London strains studied here (neither MSM, heterosexual or mixed sexual orientation), but it does share three alleles with STs circulating among MSM in Sheffield and Lincoln (data not shown). The three *opa* alleles shared by these MSM isolates are present at the same loci, suggesting they share recent ancestry. These isolates were all identical by MLST and also have identical *tbpB* alleles and therefore strongly suggests that the shared alleles are due to shared recent ancestry rather than being due to horizontal exchange.

MSM network ST825 (Table 5.3b) included nine isolates, recovered over a period of 11 weeks, that had the same alleles at eight loci and a consensus allele at two of the other loci. Three changes from the consensus were identified, one of which was due

to the duplication of a consensus allele, and one that could have arisen as a simple mosaic of two consensus alleles, or as a duplication of the allele present in this isolate at *opa11* (both alternatives involving recombination). The third variant allele (allele 220 at *opa8*) has a single NS substitution compared to the consensus allele, which occurs in a conserved region between the SVR and HVR1. This polymorphism is present in other consensus alleles in ST825 and therefore was assigned to recombination rather than point mutation.

The third MSM network, ST876 (Table 5.3c), comprised four isolates, all having identical alleles at nine loci. A consensus allele was identified at the remaining loci, including *opa11*. Only two differences to the consensus profile were observed, both of which were simple mosaics of existing consensus alleles. Recombination was therefore the mechanism involved in generating both changes in this network. Thus, in the three MSM networks all five allelic changes were due to recombination among existing alleles in the same isolate.

Table 5.3 *opa gene* profiles of isolates from individuals in London recently infected with a strain with the same genotype by NG-MAST

Table 5.3a MSM cluster ST802

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST802 Consensus		203	149	167	13	13	143	49	167	63	174	167
LG0983 (M)	31/08/2004											
W0690 (M)	11/10/2004											
LG1497 (M)	18/10/2004											
W0735 (M)	19/10/2004											
LG1682 (M)	03/11/2004											

Table 5.3b MSM cluster ST825

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST825 Consensus		199	148	199	195	156	142	190	166	179	*173*	V
LG1930 (M)	09/09/2004								142			142
W0494 (M)	16/09/2004											148
W0577 (M)	28/09/2004											165[C]
LG1304 (M)	30/09/2004											165[C]
LG1471 (M)	13/10/2004											166
LG1526 (M)	18/10/2004											166
LG1604 (M)	27/10/2004											165[C]
LG1775 (M)	11/11/2004		235									235
LG1846 (M)	25/11/2004								[NS]220[166]			148

Table 5.3c MSM cluster ST876

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST876 Consenus		168	150	200	150	157	144	192	185	180	175	168
W0231 (M)	14/07/2004		210									
LG0893 (M)	20/07/2004											
LG0903 (M)	27/07/2004											221
LG0906 (M)	28/07/2004											

Coloured alleles indicate the presence of a consensus allele at a locus and the same allele at another locus; uncoloured alleles are non-consensus alleles.
Red bold allele numbers contain sequences from an unknown source, presumably as a result of a mixed infection.
Double underlined allele numbers indicate alleles which are seen in another ST.
X, the sequence has an anomaly in the 5' region such that an allele number could not be assigned.

Table 5.3 *continued*

Table 5.3d Heterosexual cluster ST610

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST610 Consensus		204	151	201	196	158	145	193	V	181	176	V
LG0104 (M)	10/06/2004	191						**218**	**186**			222
W0182 (F)	07/07/2004		152^{C}			159			187			170^{C}
W0236 (F)	14/07/2004								$_{S}169^{145}$			$_{S}169^{145}$
LG0540 (M)	22/07/2004								$_{S}169^{145}$			$_{S}169^{145}$
LG1081 (M)	06/09/2004	**209**	$_{S}$**211**^{223}			159		183	187			$_{S}$**223**^{211}

Table 5.3e Heterosexual cluster ST709

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST709 Consensus		205	153	202	197	160	146	194	188	182	177	V
LG0643 (F)	29/07/2004											X
LG0718 (M)	09/08/2004											224
LG0923 (M)	26/08/2004											225
LG1067 (M)	26/08/2004		248^{C}							247		188
LG1062 (F)	06/09/2004											242
LG1569 (M)	29/10/2004											$_{NS}240^{188}$

Table 5.3f Heterosexual cluster ST777

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST777 Consensus		206	147	171	198	161	147	183	154	183	178	171
W0287 (M)	29/07/2004		154		215							
LG0954 (F)	26/08/2004											
LG1082 (F)	08/09/2004		213				$_{S}216^{147}$					
W0457 (M)	10/09/2004											$_{S}227^{171}$
W0656 (M)	05/10/2004											228^{C}

Table 5.3g Mixed sexual orientation cluster ST584

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST584 Consensus		207	154	171	198	164	147	183	163	184	178	V
LG0161 (M)	15/06/2004											230^{C}
LG0159 (F)	15/06/2004											231^{C}
LG0201 (M)	21/06/2004					162			189^{C}			229
LG0214 (M)	21/06/2004											231^{C}
LG0456 (F)	13/07/2004		$_{NS}155^{154}$	$_{NS}214^{171}$					243			241
LG0592 (M)	29/07/2004					162			189^{C}			172^{C}
LG0594 (M)	29/07/2004					162			189^{C}			172^{C}
LG0788 (M)	16/08/2004											154

Heterosexual network ST610 (Table 5.3d) included six isolates recovered over a period of three months, although one was excluded as a complete *opa* profile could not be obtained. A consensus allele was identified at all loci except *opa8* and *opa11* and all isolates had the same allele at five loci. As *opa8* and *opa11* did not have consensus alleles, no attempt was made to explain the origin of the differences in the alleles at these loci. Among the nine loci where there was a consensus allele, two changes were simple mosaics of consensus alleles present in ST610 isolates, whereas one was a complex mosaic (allele 152). Three *opa* alleles (allele 209, 211 and 218, shown in red) were not found in any other isolates that were examined and could not have been generated as mosaics of other alleles in these isolates. These novel alleles are candidates for *opa* alleles acquired from another strain during a mixed infection. Allele 183, which differs from the consensus at *opa7,* was also novel to the ST610 network and could not be formed as a mosaic of other alleles, although it was present as a consensus allele at the same locus in both the ST777 heterosexual network as well as the mixed sexual orientation network, ST584. This allele may have been introduced from another strain during a mixed infection, as described earlier. Thus, among the ST610 isolates, three changes were assigned to recombination among alleles in the same isolate and four to the introduction of alleles from other strains.

Heterosexual network ST709 (Table 5.3e) comprised six isolates. Excepting *opa11*, this network had a consensus allele at all loci and all isolates had the same alleles at eight loci. Only one isolate differed from the consensus profile, differing at both *opa2* and *opa9.* The variant alleles in this isolate were both mosaics of existing consensus alleles, although allele 248 was a complex mosaic.

The third heterosexual network, ST777 (Table 5.3f), included five isolates. This group had a consensus allele at all loci, including *opa11*, with all isolates having the same alleles at seven loci. Six changes to the consensus profile were observed, one of which was due to the duplication of an existing consensus allele, two were simple mosaics of consensus alleles present in this isolate and one was a complex mosaic. Alleles 216 and 227 each had a single synonymous substitution within a conserved region compared to the consensus allele and, as these substitutions were found in other alleles in these isolates, it was assumed they arose by recombination rather than mutation.

The mixed sexual orientation network, ST584 (Table 5.3g), included eight isolates recovered over a two month period. Six were from heterosexuals, one from a bisexual and one from a homosexual. One change to the consensus (at *opa5*) was due to the duplication of a consensus allele at another locus. Two were simple mosaics of alleles present in this network and one was a complex mosaic. Allele 214 at *opa3* differed from the consensus by a single unique NS substitution in HVR2 and was assigned as a point mutation. Allele 155 at *opa2* also had a single NS substitution compared to the consensus allele, in a conserved region that borders HVR2, but this substitution is seen in other alleles of this network and was assigned to recombination rather than point mutation.

Summing the data for the isolates from all seven London networks, there were 26 recent changes, 21 were assigned to recombination among alleles in the same isolate (three of which being allele duplications, the remainder being due to the formation of

mosaic alleles), four to importation of alleles from other strains, all of which were in the same sexual network, and only one to point mutation.

5.4 Discussion

As observed in the Sheffield isolates in the previous chapter, consensus *opa* allelic profiles could be identified in each sexual network but atypically, *opa8* was found to be variable and without a consensus allele in the ST610 sexual network. Furthermore, three other sexual networks, ST802, ST876 and ST777, were found to have a consensus allele at *opa11*. The time frame in which these were collected was similar to that of the Sheffield isolates and so this finding is not due to strains being isolated closer in time. Until we know the reason for the high rate of change at *opa11*, it will be difficult to comment on the reason behind this.

Generally, there appeared to be very little sharing of alleles between the networks, but where it was detected, this was mostly considered to be due to recent shared ancestry as the isolates were usually identical or very similar by MLST. Allele sharing due to horizontal transfer requires mixed gonococcal infections and might be expected among isolates recovered from the same location during the same time period but, excepting the ST610 network, it was not found among isolates from the MSM and heterosexual networks in London, or among isolates of the two major strains that had been co-circulating among heterosexuals in Sheffield for at least 32 months (previous chapter). For reasons that are unclear, the small ST610 heterosexual cluster was atypical as it included two isolates that appeared to have acquired multiple novel alleles from other strains including one consensus allele from another ST studied. Although 5/7 MLST loci were shared between the STs in question, this shared allele was present only in the most recent isolate. This along

with the other importation events in this ST strongly suggests that importation was responsible.

Having analysed the sexual networks from London in the same way as the sexual contacts from Sheffield, there were 26 changes identified in total. 25 were assigned to recombination, including three duplication events and 4 importation events. There was only one change by point mutation observed. These values further support the results from the Sheffield data in the previous chapter that strongly suggest recombination is the primary mechanism that brings about change to *opa* allelic profiles.

The results presented here complement the results from Sheffield in the previous chapter and will be discussed in more depth in combination with them in Chapter 9.

Chapter 6: The relationship between variation at *opa11* and at the neighbouring *pilE* gene

6.1 Introduction

The *pilE* locus is the expression site for the major pilin subunit and is known to have one of the highest rates of antigenic variation measured (50, 114). Several inactive/silent loci (*pilS*) which lack the 5' conserved region required for expression are located in various locations around the chromosome (76, 141, 190). Antigenic variation of pilin occurs frequently via transformation and/or gene conversion events that replace *pilE* sequences with those from *pilS* loci (76, 87, 114, 147, 216, 246). These recombination events that generate new *pilE* sequences involve the exchange of single and multiple variable minicassettes via interspersed short conserved sequences, as well as exchange of complete silent copies (79, 114). There is also evidence that recombinational cross over points may also occur within the minicassettes (95, 190). Antigenic variation in pili is accompanied by changes in the binding characteristics of isolates and so there may exist several receptors for pili on different human cells, one of which has been identified as CD46 (108, 109, 112, 113, 147, 224).

In addition to pilin antigenic variation, it is also subject to phase variation. This may occur by deletion of the 5' region of *pilE* during a recombination event, and also internal region deletions (87, 92, 189), unequal recombination between the expressed and silent loci that result in long (L) pilin subunits that cannot form pilus structures or, incorrect processing of propilin at amino acid position 40 rather than position 1

which results in short (S) pilin units that are soluble (87, 190). Furthermore, phase variation of the pilus biogenesis protein PilC, via a run of G residues in the region encoding the signal peptide, in turn results in phase variation of pili (106). Therefore, it is important to note that pilus phase variation does not necessarily result in two distinct phases (on and off) but rather a spectrum of variants, including S-pilin and L-pilin, that affect the piliated state of the cell and which in turn influences not only colony morphology but also transformation efficiency and epithelial cell adherence (128).

The expression of *pilE* also appears to be under regulatory control mechanisms via an activator and repressor encoded by two loci, *pilA* and *pilB* respectively, that are located in close proximity to *pilE* (the first of two copies) in strain MS11 (147).

Variation occurs much more rapidly at *opa11* than at other *opa* loci and the pilin expression locus, *pilE*, is located adjacent to *opa11* in strain FA1090 (55). These two genes, along with a silent pilin gene, *pilS*, and a partial pilin gene consisting of portions of interrupted pilin genes are clustered together over a 3kb region (Fig. 6.1) (79). The variable regions of *pilE* were therefore sequenced to determine whether variation at one locus is linked to that at the other.

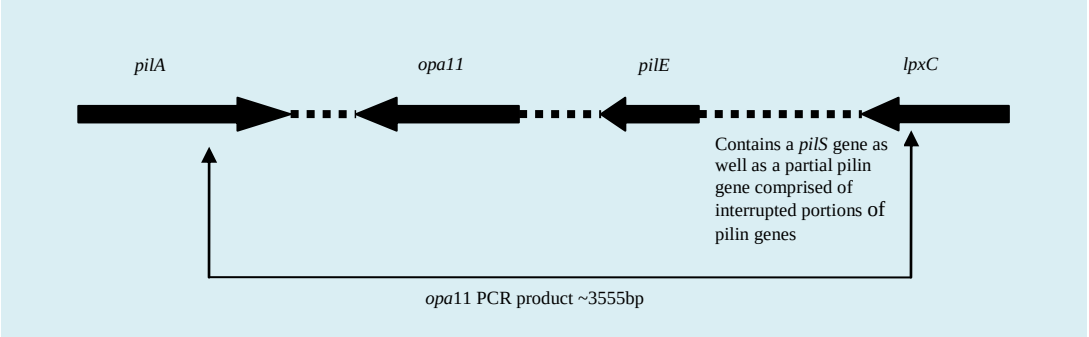


Figure 6.1 Region amplified during *opa11* PCR which also amplifies *pilE* and *pilS* loci. Loci are designated as they have been annotated at http://stdgen.northwestern.edu/stdgen/bacteria/ngon/, except for *opa11* which has not been annotated.

6.2 Methods

6.2.1 Isolates

The variable regions of *pilE* were sequenced for the pairs and triplets of isolates obtained from recent sexual contacts in Sheffield as described in Chapter 4.

6.2.2 *pilE* sequencing

The PCR product from strain FA1090 that contains the *opa11* gene also includes the downstream pilin expression locus, *pilE*, one silent pilin gene, *pilS*, as well as a partial pilin gene comprised of interrupted portions of pilin genes (79). Since *pilS* genes lack the 5' conserved region required for expression (76), a sequencing primer could be designed in this 5' region of the *pilE* gene (pilE1 5'-ATCGAGCTGATGATTGT-3') to sequence (off one strand) the variable cassettes of the expression locus (but not those of *pilS*). The *pilE* sequences were trimmed to begin with the conserved sequence TTGGCGGCA and to end after the last nucleotide of the stop codon with allele numbers assigned to each unique sequence. Frame shifts occur frequently in *pilE* which generate premature stop codons (92), therefore, to avoid changes in the downstream regions going unnoticed, sequences were trimmed to the stop codon identified based on alignments of *pilE* genes that were full length and in frame rather than the early stop codons generated by such frame shifts. When sequence could not be obtained off this primer, another was used, designed from a conserved sequence present at the 3' end of both the *pilE* and *pilS* genes (pilE6 5'-AGCACCTGCCGTCAAC-3'). In the case of strain FA1090, this primer yields a mixed sequence as priming occurs from *pilE*, *pilS* and the partial

pilin gene portions. However, in isolates where the pilE1 primer did not give sequence, the pilE6 primer usually gave unambiguous sequence, rather than mixed sequence. This latter combination suggests that in these isolates a *pilE* gene is absent at this position and that the unambiguous sequence is from a *pilS* gene. For the isolates in which *pilE* appears to be absent, the region from the *opa11* locus, through the region carrying the *pil* genes, to the end of the PCR product was sequenced to confirm this absence. Sequencing was carried out as described previously for *opa* loci.

6.3 Results

6.3.1 Nature of the pilin genes neighbouring the *opa11* locus

Amongst the 45 isolates that comprised the pairs and triplets of isolates from known sexual contacts in Sheffield, 29 had a *pilE* gene for which an allele could be assigned. Two isolates had *pilE* genes but an allele could not be assigned as there were regions of mixed sequence and this presumably represents a change occurring within culture (denoted P in Table 6.1a). Both were repeated from the PCR and yielded the same results. This phenomenon is discussed in more detail in Chapter 7. Eleven isolates did not have a *pilE* gene at this locus (denoted Abs in Table 6.1b) which was confirmed with sequence and is supported by the smaller size of the PCR products. In two further isolates, no sequence was obtained with the pilE1 primer yet mixed sequence was obtained with the pilE6 primer. For these two isolates, the PCR product size was smaller than expected, yet larger than those isolates in which the absence of *pilE* was confirmed with sequence. Therefore, a partial deletion probably occurred involving the 5' region of the *pilE* gene.

In five pairs, one isolate had a *pilE* gene while in the other, the *pilE* gene was absent or at least partially deleted. Since these isolates are from mutually named sexual contacts and their relatedness has been demonstrated by both *opa*-typing as well as NG-MAST, it is probable that these deletions are very recent events.

Table 6.1a o*pa11* allele and nature of *pilE* gene of ST12 isolates from recent sexual contacts

Pair/ triplet	Isolate	*opa11*	*pilE*
Triplet1	**S149**	141	14
	S152	140[C]	23
	S154	54	14
Triplet 2	S306	253	21
	S308	234[C]	26
	S317	233	25
Triplet 3	**S445**	114	14
	S463	115[C]	22
	S466	67	40
Pair 1	S78	117	27
	S86	118	28
Pair 2	S98	120[C]	14
	S100	121	30
Pair 3	S241	29	29
	S256	119	14
Pair 4	S253	54	P
	S257	25	14
Pair 5	S493	116[C]	24
	S514	116[C]	14
Pair 6	S517	1	54
	S522	25	P

As in previous tables. Bold isolate numbers indicate the isolate from the individual who named the other two individuals in the triplet as recent sexual contacts. Blocked alleles indicate differences between isolates from recent sexual contacts. P, *pilE* present but allele could not be assigned, see text. Abs, *pilE* absent at this location in the genome, see text. Partial, 5' region of *pilE* absent.

Table 6.1b *opa11* allele and nature of *pilE* gene of ST261 isolates from recent sexual contacts

Pair/ triplet	Isolate	*opa11*	*pilE*
Pair 7	S21	96	33
	S29	11	46
Pair 8	S24	96	31
	S26	68	32
Pair 9	S69	$_{S}124^{110}$	Abs
	S70	$_{S}124^{110}$	42
Pair 10	S74	125^{C}	Abs
	S75	75	Abs
Pair 11	S89	$_{NS}123^{96}$	Abs
	S107	17	Abs
Pair 12	S121	68	36
	S123	11	Partial
Pair 13	S128	127	Abs
	S129	127	37
Pair 14	S136	111^{C}	Abs
	S137	111^{C}	Abs
Pair 15	S156	80	Abs
	S162	83^{C}	Abs
Pair 16	S192	236	43
	S196	75	Partial
Pair 17	S266	103	34
	S271	252^{C}	44
Pair 18	S254	108	Abs
	S258	$_{NS}237^{96}$	45

As in Table 6.1a.

6.3.2 Variation at *opa11* and at the neighbouring *pilE* gene

Pairs (or triplets) with a *pilE* allele in one isolate, and an apparent absence of *pilE* (or one where an allele could not be assigned) at the other, were classed as having different *pilE* genes (Tables 6.1a and b). Changes between sexual contacts typically occurred at both loci and only one pair of isolates (pair 14) had no changes at either *pilE* or *opa11*. In almost all pairs or triplets where differences in *pilE* were observed they were accompanied by differences in *opa11,* suggesting the two loci have similar rates of variation. Three pairs (5, 9 and 13) had changes at *pilE* but no change at *opa11,* suggesting that changes at the former locus are not linked to changes at the latter. However, there were no clear examples of changes at *opa11* without changes at *pilE*, although three pairs (10, 11 and 15) had differences at *opa11* whilst not appearing to have a *pilE* gene next to *opa11*.

6.4 Discussion

Bhat *et al.* previously reported one *opa* locus (*opaE* in strain MS11) to be more variable than others and this was located next to *pilE* and thus corresponds to *opa11* in this nomenclature (14). Genetic linkage of *pil* and *opa* has been reported although the meaning of any linkage has not been explained (147). The sequences of the adjacent *pilE* gene were examined for each of the pairs and triplets from sexual contacts in Sheffield and showed that most of the isolates in the pairs and triplets had different *pilE* sequences. Indeed, these results are not entirely surprising since gonococcal isolates from sexual contacts have previously been reported to frequently have different pilus types (245), and also, the pili expressed in different anatomical sites in the same patient are antigenically distinct as are those expressed by the inoculum in human challenge compared to those expressed later in the experimental infection (147, 191, 217).

A change at *pilE* was almost always accompanied by a change in *opa11*. There were only three pairs where a change in *pilE* was not accompanied by one in *opa11* and no clear examples of change in *opa11* without change in *pilE*. This implied that the two loci have similarly high rates of variation, but does not provide evidence for a direct causal relationship.

While looking at the neighbouring *pilE* gene, it was found that it is not always present. There were examples of its deletion, both complete and partial. Since pilin is also important in the early stages of infection for attachment, it is probable that the gene is present elsewhere in the genome of those isolates. Strain MS11 has been

reported to have two copies (15, 144) of the locus but it is not known how common this is amongst other gonococcal strains. The expression of pili is routinely switched off in culture (19) and the deletion of this locus is one way in which this occurs, in addition to the usual phase variation of pili (87, 190). It is possible that isolates that are confined to laboratory propagation may lose the expression locus permanently as pili are not required in culture. This would be the only other possible scenario under which a complete deletion of *pilE* could be permissible without a secondary expression locus elsewhere in the genome.

If a relationship between *opa11* and *pilE* exists, then *opa11* should vary at a rate similar to that of the other *opa* loci if the adjacent *pilE* gene was not present at this location. However, the absence of some of these *pilE* loci may have occurred recently, therefore, the slower rate of variation may not be obvious. In order to determine whether an association exists, it would be necessary to examine isolates in which the *pilE* locus is consistently absent at this location.

Chapter 7: Stability of *opa* allelic profiles, *pilE* sequences and coding repeats during *in vitro* sub-culturing

7.1 Introduction

In previous chapters, the possibility of changes occurring during sub-culture has been mentioned and deferred to this chapter. These changes refer to the CRs, known to change rapidly, as well as both the *opa* allelic profiles and *pilE* sequences. Given the fact that these genes and regions are known to vary rapidly, change in culture will undoubtedly occur at times. What is not known is how frequently this occurs at these loci in gonococci in culture. A simple sub-culture experiment was performed to evaluate this and is described below.

7.2 Methods

Twenty isolates, including at least one isolate of each ST in the Sheffield and London clusters, were sub-cultured ten times, always from single colonies. A fresh DNA lysate was prepared from each sub-culture. The *opa* profiles obtained on the first and last sub-cultures were compared. At loci where differences in the *opa* profile were detected, the alleles for the intervening sub-cultures were determined to identify at which sub-culture the change occurred. The *pilE* allele and the multiple of CRs for the first and last sub-cultures were also determined for comparison.

7.3 Results

7.3.1 Changes in *opa* allelic profiles during sub-culture

No differences in the *opa* allelic profiles were detected between the 1st and 10th sub-cultures in 18/20 isolates. The differences in the other two isolates both occurred at *opa11*, one on the 7th and the other on the 9th sub-culture (Table 7.1). Both of the new *opa* alleles were mosaics of existing alleles in their profiles and therefore involved recombination events.

7.3.2 Changes in *pilE* sequences during sub-culture

In the two isolates where *opa11* changed, *pilE* also changed, at the same sub-culture in one isolate and a different sub-culture in the other. An additional six isolates had differences in *pilE* whilst *opa11* remained unchanged, suggesting *pilE* changes more rapidly than *opa11 in vitro*.

7.3.3 Changes in the multiple of CRs during sub-culture

The number of CRs were also compared to estimate how rapidly these change during sub-culture. Ignoring the two *opa11* loci where allelic changes were observed, nine of the 20 (45%) isolates showed variation in the number of CRs during sub-culture. In some isolates, this occurred at more than one *opa* locus. A total of 18 of the 218 (8%) *opa* loci that could be compared varied, 50% (9/18) of which were CR

expansions and 33% (6/18) contractions by a multiple of one repeat. The three other loci expanded or contracted by a greater multiple (2-3 repeats).

Table 7.1 *opa* profiles of isolates sub-cultured ten times from single colonies

Isolate: sub-culture	ST	Locus *opa1* $_{CR}$	*opa2* $_{CR}$	*opa3* $_{CR}$	*opa4* $_{CR}$	*opa5* $_{CR}$	*opa6* $_{CR}$	*opa7* $_{CR}$	*opa8* $_{CR}$	*opa9* $_{CR}$	*opa10* $_{CR}$	*opa11* $_{CR}$	*pilE*
S350: sub1	12	1 $_{13}$	10 $_{11}$	16 $_{4}$	25 $_{19}$	29 $_{10}$	36 $_{13}$	45 $_{8}$	54 $_{7}$	*	67 $_{10}$	102 $_{9}$	1
S350: sub2												102 $_{9}$	1
S350: sub3												102 $_{9}$	1
S350: sub4												102 $_{9}$	1
S350: sub5												102 $_{9}$	1
S350: sub6												102 $_{9}$	1
S350: sub7												102 $_{9}$	1
S350: sub8												102 $_{9}$	1
S350: sub9												249 $_{7}$	13
S350: sub10	12	1 $_{13}$	10 $_{11}$	16 $_{4}$	25 $_{19}$	29 $_{10}$	36 $_{14}$	45 $_{8}$	54 $_{7}$	*	67 $_{10}$	249 $_{7}$	13
S280: sub1	12	1 $_{10}$	10 $_{11}$	16 $_{4}$	25 $_{19}$	29 $_{12}$	98 $_{15}$	45 $_{7}$	54 $_{7}$	*	67 $_{12}$	244 $_{14}$	2
S280: sub10		1 $_{10}$	10 $_{11}$	16 $_{4}$	25 $_{20}$	29 $_{12}$	98 $_{16}$	45 $_{7}$	54 $_{7}$	*	67 $_{12}$	244 $_{14}$	14
S427: sub1	12	1 $_{13}$	10 $_{13}$	16 $_{4}$	25 $_{20}$	29 $_{7}$	99 $_{13}$	45 $_{8}$	54 $_{7}$	*	67 $_{7}$	245 $_{13}$	3
S427: sub10		1 $_{13}$	10 $_{13}$	16 $_{4}$	25 $_{20}$	29 $_{7}$	99 $_{13}$	45 $_{8}$	54 $_{7}$	*	67 $_{7}$	245 $_{13}$	15
S89: sub1	261	2 $_{14}$	11 $_{7}$	17 $_{2}$	17 $_{14}$	30 $_{12}$	37 $_{8}$	46 $_{11}$	96 $_{10}$	62 $_{13}$	68 $_{21}$	123 $_{7}$	Abs
S89: sub10		2 $_{14}$	11 $_{7}$	17 $_{2}$	17 $_{14}$	30 $_{12}$	37 $_{8}$	46 $_{11}$	96 $_{10}$	62 $_{13}$	68 $_{21}$	123 $_{7}$	Abs
S254: sub1	261	2 $_{13}$	11 $_{10}$	17 $_{2}$	17 $_{9}$	30 $_{11}$	37 $_{8}$	95 $_{9}$	133 $_{7}$	62 $_{14}$	68 $_{18}$	133 $_{6}$	Abs
S254: sub10		2 $_{13}$	11 $_{10}$	17 $_{2}$	17 $_{9}$	30 $_{11}$	37 $_{8}$	95 $_{9}$	133 $_{7}$	62 $_{14}$	68 $_{18}$	133 $_{6}$	Abs
S324: sub1	261	2 $_{15}$	11 $_{7}$	17 $_{2}$	17 $_{10}$	30 $_{11}$	37 $_{8}$	95 $_{9}$	96 $_{8}$	62 $_{14}$	68 $_{22}$	96 $_{9}$	4
S324: sub10		2 $_{16}$	11 $_{7}$	17 $_{2}$	17 $_{10}$	30 $_{11}$	37 $_{7}$	95 $_{9}$	96 $_{8}$	62 $_{14}$	68 $_{22}$	96 $_{12}$	16
LG1846: sub1	825	199 $_{9}$	148 $_{23}$	199 $_{19}$	195 $_{15}$	156 $_{4}$	142 $_{8}$	190 $_{7}$	220 $_{10}$	179 $_{8}$	173 $_{35}$	148 $_{9}$	5
LG1846: sub10		199 $_{10}$	148 $_{22}$	199 $_{19}$	195 $_{14}$	156 $_{4}$	142 $_{8}$	190 $_{7}$	220 $_{10}$	179 $_{8}$	173 $_{34}$	148 $_{9}$	Abs
LG1775: sub1	825	199 $_{9}$	235 $_{25}$	199 $_{20}$	195 $_{17}$	156 $_{4}$	142 $_{8}$	190 $_{6}$	166 $_{10}$	179 $_{10}$	173 $_{26}$	235 $_{25}$	6
LG1775: sub10		199 $_{9}$	235 $_{26}$	199 $_{19}$	195 $_{17}$	156 $_{4}$	142 $_{8}$	190 $_{6}$	166 $_{10}$	179 $_{10}$	173 $_{26}$	235 $_{25}$	17
LG1304: sub1	825	199 $_{8}$	148 $_{22}$	199 $_{17}$	195 $_{15}$	156 $_{4}$	142 $_{8}$	190 $_{6}$	166 $_{10}$	179 $_{10}$	173 $_{18}$	165 $_{8}$	7
LG1304: sub10		199 $_{8}$	148 $_{22}$	199 $_{17}$	195 $_{15}$	156 $_{4}$	142 $_{8}$	190 $_{6}$	166 $_{10}$	179 $_{10}$	173 $_{18}$	165 $_{8}$	18
LG1682: sub1	802	203 $_{8}$	149 $_{4}$	167 $_{13}$	13 $_{14}$	13 $_{11}$	143 $_{9}$	49 $_{10}$	167 $_{7}$	63 $_{5}$	174 $_{7}$	167 $_{10}$	Abs
LG1682: sub10		203 $_{8}$	149 $_{4}$	167 $_{13}$	13 $_{14}$	13 $_{11}$	143 $_{9}$	49 $_{10}$	167 $_{7}$	63 $_{5}$	174 $_{7}$	167 $_{10}$	Abs
LG0906: sub1	876	168 $_{5}$	150 $_{13}$	200 $_{8}$	150 $_{9}$	157 $_{18}$	144 $_{19}$	192 $_{9}$	185 $_{11}$	180 $_{2}$	175 $_{14}$	168 $_{13}$	Abs
LG0906: sub10		168 $_{5}$	150 $_{13}$	200 $_{8}$	150 $_{9}$	157 $_{19}$	144 $_{19}$	192 $_{9}$	185 $_{11}$	180 $_{2}$	175 $_{14}$	168 $_{13}$	Abs
LG0893: sub1	876	168 $_{5}$	150 $_{14}$	200 $_{8}$	150 $_{9}$	157 $_{18}$	144 $_{17}$	192 $_{9}$	185 $_{11}$	180 $_{2}$	175 $_{9}$	168 $_{8}$	Abs
LG0893: sub10		168 $_{5}$	150 $_{14}$	200 $_{8}$	150 $_{9}$	157 $_{18}$	144 $_{17}$	192 $_{9}$	185 $_{11}$	180 $_{2}$	175 $_{9}$	168 $_{8}$	Abs
LG0540: sub1	610	204 $_{5}$	151 $_{16}$	201 $_{11}$	196 $_{7}$	158 $_{15}$	145 $_{5}$	193 $_{19}$	169 $_{10}$	181 $_{8}$	176 $_{16}$	169 $_{8}$	Partial
LG0540: sub10		204 ^{5}	151 $_{16}$	201 $_{11}$	196 $_{7}$	158 $_{15}$	145 $_{5}$	193 $_{19}$	169 $_{10}$	181 $_{8}$	176 $_{16}$	169 $_{8}$	Partial
LG0923: sub1	709	205 $_{8}$	153 $_{13}$	202 $_{11}$	197 $_{22}$	160 $_{11}$	146 $_{8}$	194 $_{16}$	188 $_{7}$	182 $_{8}$	177 $_{6}$	225 $_{8}$	Abs
LG0923: sub10		205 ^{8}	153 $_{13}$	202 $_{11}$	197 $_{22}$	160 $_{11}$	146 $_{8}$	194 $_{>15}$	188 $_{7}$	182 $_{8}$	177 $_{6}$	225 $_{8}$	Abs
LG1067: sub1	709	205 $_{8}$	248 $_{14}$	202 $_{8}$	197 $_{13}$	160 $_{11}$	146 $_{8}$	194 $_{18}$	188 $_{7}$	247 $_{8}$	177 $_{7}$	188 $_{7}$	8
LG1067: sub10		205 ^{8}	248 $_{14}$	202 $_{8}$	197 $_{13}$	160 $_{11}$	146 $_{8}$	194 $_{19}$	188 $_{7}$	247 $_{8}$	177 $_{7}$	188 $_{7}$	8
LG0643: sub1	709	205 $_{8}$	153 $_{16}$	202 $_{8}$	197 $_{13}$	160 $_{11}$	146 $_{8}$	194 $_{18}$	188 $_{7}$	182 $_{8}$	177 $_{7}$	X $_{8}$	9
LG0643: sub10		205 ^{8}	153 $_{16}$	202 $_{8}$	197 $_{13}$	160 $_{11}$	146 $_{8}$	194 $_{18}$	188 $_{7}$	182 $_{8}$	177 $_{7}$	X $_{8}$	9
LG1062: sub1	709	205 $_{8}$	153 $_{13}$	202 $_{9-10}$	197 $_{14}$	160 $_{11}$	146 $_{8}$	194 $_{19}$	188 $_{7}$	182 $_{9}$	177 $_{7}$	242 $_{13}$	Partial
LG1062: sub10		205 ^{8}	153 $_{11}$	202 $_{8}$	197 $_{14}$	160 $_{11}$	146 $_{8}$	194 $_{19}$	188 $_{7}$	182 $_{10}$	177 $_{7}$	242 $_{13}$	Partial
LG0954: sub1	777	206 $_{13}$	147 $_{14}$	171 $_{12}$	198 $_{16}$	161 $_{40}$	147 $_{12}$	183 $_{4}$	154 $_{15}$	183 $_{7}$	178 $_{12}$	X $_{13}$	10
LG0954: sub10		206 $_{13}$	147 $_{14}$	171 $_{12}$	198 $_{16}$	161 $_{40}$	147 $_{12}$	183 $_{4}$	154 $_{15}$	183 $_{7}$	178 $_{12}$	X $_{13}$	10
LG0214: sub1	584	207 $_{17}$	154 $_{11}$	171 $_{11}$	198 $_{22}$	164 $_{2}$	147 $_{6}$	183 $_{4}$	163 $_{18}$	184 $_{4}$	178 $_{7}$	246 $_{11}$	11
LG0214: sub10		207 $_{17}$	154 $_{11}$	171 $_{11}$	198 $_{22}$	164 $_{2}$	147 $_{6}$	183 $_{4}$	163 $_{17}$	184 $_{4}$	178 $_{7}$	246 $_{11}$	11
LG0456: sub1	584	207 $_{17}$	155 $_{6}$	214 $_{6}$	198 $_{7}$	163 $_{19}$	147 $_{11}$	183 $_{4}$	243 $_{13}$	184 $_{1}$	178 $_{8}$	241 $_{16}$	12
LG0456: sub2												241 $_{16}$	19
LG0456: sub3												F $_{16}$	19
LG0456: sub4												241 $_{16}$	19
LG0456: sub5												241 $_{16}$	19
LG0456: sub6												241 $_{16}$	19
LG0456: sub7												250 $_{16}$	19
LG0456: sub8												250 $_{16}$	19
LG0456: sub9												250 $_{16}$	19
LG0456: sub10	584	207 $_{17}$	155 $_{6}$	214 $_{6}$	198 $_{7}$	163 $_{19}$	147 $_{11}$	183 $_{4}$	243 $_{13}$	184 $_{1}$	178 $_{8}$	250 $_{16}$	19

Profiles as in previous tables. *CR*, CR multiple identified. Changes in *opa* and *pilE* alleles are blocked in red. Subscript indicates multiple of CRs, changes indicated by coloured cells. F, failure of primers such that allele could not be assigned. A diagonal line across the allele number indicates that it is hybrid/mosaic of alleles present in the *opa* gene repertoire. X, the sequence has an anomaly in the 5' region such that an allele number could not be assigned.

7.4 Discussion

The *opa11* and *pilE* genes were found to vary during sub-culture. The two changes that occurred at *opa11* generated mosaic alleles and is therefore clear evidence that mosaics can arise by recombination between the alleles within an isolate. Eight changes were observed at *pilE,* indicating that *pilE* changes more rapidly than *opa11 in vitro*.

The multiples of the CRs also varied on sub-culture with 45% of the sub-cultured strains showing changes in the multiple of CRs in at least one of their *opa* loci. This is consistent with the high frequency of CR phase variation *in vitro* (1-3 x 10^{-3} per cell per generation) reported by Belland *et al.* (13). The fact that there were sequences which were identical apart from the multiple of CRs during serial sub-culture supported the choice to assign alleles irrespective of this region. A similar approach has been used for meningococcal *opa* genes (93). The majority (83.3%) of changes in the CRs were by a multiple of one, not surprising considering strand slippage and mispairing events are believed to be involved (13, 154, 206). They would not be expected to result in larger multiples (206), therefore the three loci where multiples of two and three were observed are mostly likely due to more than one event occurring between the first and last sub-culture. Consequently, it is not possible to know which of the *opa* genes were being expressed *in vivo* once isolates have been cultured from a clinical specimen.

Stern *et al.* (208) reported the ability to isolate colony variants that had undergone changes in their state of expression of pilin or Opa but not both. Concurrent changes

in both the *pilE* and *opa11* genes were observed in this study however, an association with concurrent changes in expression cannot be deduced due to the phase variable nature of both gene families.

Chapter 8: Analysis of the hyper-variable regions of Opa proteins from closely related isolates

8.1 Introduction

Two dimensional models of Opa proteins predict the presence of four surface exposed loops and several studies have mapped epitopes recognised by monoclonal antibodies to the second and third loops, HVR1 and HVR2 (93, 132). These two loops therefore constitute the major antigenic portions of the protein that would be detected by the host's immune system but are also involved in binding to host cell receptors. Specific combinations of HVR1 and HVR2 have been shown to be necessary for successful adhesion to receptors (25, 51). As such, many studies on the diversity of gonococcal Opa proteins have focused on the amino acid sequences of these two regions and have found that recombination events can result in both identical and diverse combinations of HVRs at different *opa* loci (93, 94).

The analysis of *opa* gene nucleotide sequences presented in previous chapters shows that the majority of recent changes observed were due to recombination events rather than point mutation. While the secondary structures and the location of the HVRs tell us which regions are surface exposed, little is known about the tertiary structure of Opa proteins and the extent of the interactions between the surface exposed loops. It is very likely that certain HVRs and combinations thereof will be effective at binding to host receptors, or to particular host receptors (e.g. differential binding to urethral, cervical, rectal mucosa), whereas others may be less effective or non-functional. Besides their ability to bind to host receptors, Opa proteins, being surface exposed,

are targeted by the host antibody response (244). This imposes selection for changes in the Opa proteins. The fact that Opa protein expression can be switched to occur from different loci during an infection via the phase variation system (202, 206), further confounds the scenario in which the effect of changes in the protein on function must be balanced with the need for immune evasion.

8.2 Methods

8.2.1 Isolates

Sequence data from all the Sheffield isolates that were ST12 and ST261 were included here, as were all the sequence data from isolates from the London sexual networks. No additional isolates from the diverse set were included (Chapter 3), nor were any additionally identified alleles from the sub-culturing experiment (Chapter 7).

8.2.2 Sequence translation and trimming

The primary nucleotide sequences obtained in previous chapters were translated in MEGA version 3.1 (121). The HVR1 and HVR2 were then identified for further analysis. They begin and end at the amino acid sequences of each of the HVRs described by Callaghan *et al.* (30). The trim points used by this author were chosen in order to maintain consistency through literature and is similar to the regions defined by Malorny *et al.* (132). Each unique HVR1 and HVR2 amino acid sequence was given a unique identity (id) number such that the combination of HVR1 and HVR2 amino acid sequences can be represented by a hyphenated 2 digit number (e.g. 1-2 indicates the HVR1 amino acid sequence is id one and the HVR2 amino acid sequence is id two).

8.2.3 HVR matrices

For each sexual network, matrices were drawn with the HVR1 id on the y-axis and HVR2 id on the x-axis. Each cell indicates the count of a particular HVR combination. This format allows the combinations and sharing of HVRs to be readily visualised.

8.2.4 Dendrograms of HVRs

Dendrograms based on the number of amino acid differences were drawn for each set of HVR1 and HVR2 amino acid sequences, for each of the sexual networks, using MEGA version 3.1 (121).

8.3 Results

8.3.1 Amino acid sequences of Opa proteins and their HVRs

Amongst the ST12 and ST261 isolates from Sheffield and the seven sexual networks in London (n= 106 isolates), there were 171 unique *opa* alleles across all 11 *opa* loci. When translated, these *opa* sequences equated to 168 unique amino acid sequences. Within these 168 distinct amino acid sequences, there were 63 unique HVR1 and 63 HVR2 sequences, but only 98 unique HVR1-HVR2 combinations out of a possible 3969 (63 HVR1 x 63 HVR2). The remaining 70 (42%) amino acid sequences have one of these 98 HVR combinations but differed in regions outside the HVRs. For this reason, alleles with the same HVR1 and HVR2 amino acid sequences are not always encoded by the same *opa* allele.

8.3.2 HVRs within populations

Tables 8.1, 8.2 and 8.3 (pages 123-124, 126-127 and 133-134) show the HVR1 and HVR2 combinations of the ST12 and ST261 isolates from Sheffield and all the sexual networks from London that were examined in the previous two chapters. The colours have not been changed and still indicate the presence of consensus *opa* alleles.

From these tables, it is apparent that different consensus alleles (based on nucleotide sequence) can be comprised of identical HVR combinations (based on amino acid sequences). There are examples in all networks studied but one, ST709, of different

consensus alleles that have identical HVR combinations. They include the following: ST12 (Table 8.1a and 8.2a) *opa3* and *opa4*, ST261 (Table 8.1b and 8.2b) *opa8* and *opa10*, ST802 (Table 8.3a) *opa2* and *opa7*, ST825 (Table 8.3b) *opa1/3* and *opa10*, *opa2* and *opa5*, ST876 (Table 8.3c) *opa5* and *opa6*, ST610 (Table 8.3d) *opa2* and *opa5*, *opa7* and *opa9*, ST777 (Table 8.3f) *opa5* and *opa8* and ST584 (Table 8.3g) *opa2* and *opa8*. The obvious question is how these alleles differ if not in the HVRs. This will be different in each case but it is worth describing at least a couple of examples. The alleles at *opa3* (allele 16) and *opa4* (allele 25) in ST12 differ in the sequence encoding the SVR with an additional NS difference between the SVR and HVR1. In the second example, in ST261, the alleles at *opa8* (allele 96) and *opa10* (allele 68) have the same HVR combination but also differ in the SVR and by a single amino acid residue at the very end of the gene close to the stop codon. HVR1 and HVR2 combinations present in consensus alleles are referred to as consensus HVR combinations.

Similarly, different consensus alleles can have only one HVR amino acid sequence in common, as seen in consensus alleles 1 and 10 amongst ST12 isolates (Table 8.1a), which share the same HVR1 amino acid sequence (which is also found in consensus alleles 16 and 25) but are associated with different HVR2 amino acid sequences.

Amongst the non-consensus alleles, there are examples of novel combinations of HVR1 and HVR2, as seen in isolate S78 (Table 8.1a) at *opa11* where the allele (allele 117) has the consensus HVR1 amino acid sequence from the *opa5* locus and the consensus HVR2 amino acid sequence from the *opa6* locus. Non-consensus

alleles are also seen to have consensus HVR combinations. One example of this is seen in three isolates, S149, S152 and S154 (Table 8.1a and 8.2a) at *opa5* where a non-consensus allele (indicated by not being coloured, allele 128) has the consensus HVR combination (13-16) of allele 29. In this example there are six synonymous nucleotide differences clustered closely between HVR1 and HVR2. This allele is believed to be a simple recombinant of allele 29 and the consensus allele at locus *opa8* (allele 54) which has all of these synonymous substitutions and shares the identical HVR1.

There are also examples of new non-consensus HVR combinations being generated, as in isolate S7 at *opa5* (Table 8.1a). This isolate has the consensus HVR1 id 13 but a non-consensus HVR2, id 47 (the latter being found at *opa*11). Because the exact relationship between these isolates is not known, a more detailed explanation of the changes is given in the next section, based on the differences identified between the isolates from sexual contacts.

Brief attention should also be drawn to the HVRs at *opa11*. Whilst the locus is highly variable, the proteins encoded frequently have the same HVR combinations as the consensus alleles or they have new combinations of the individual HVRs. More novel HVRs were expected to be identified at this locus.

Table 8.1a HVR1-HVR2 amino acid combinations of alleles from the 21 ST12/*opa* -type 1 isolates chosen to be representative of the sampling period

	Locus										
Isolate	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus allele	1	10	16	25	29	36	45	54	†	67	V
Consensus HVR combination	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	V
S7	13-1	1-14	1-14	1-14	13-**47**	22-24	28-29	13-1		24-40	22-**47**
S58	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
S78	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
S81	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
S86	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
S98	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	22-1
S100	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-1
S149	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
S151	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-1
S152	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
S154	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-1
S184	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	**40**-40
S241	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
S256	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-1
S276	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
S280	1-1	1-9	1-14	1-14	13-16	**40**-24	28-29	13-1		24-40	24-40
S302	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-14
S306	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-1
S308	1-1	1-9	1-14	1-14	13-**53**	22-24	28-29	13-1		24-40	13-9
S350	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
S427	1-1	1-9	1-14	1-1	13-16	22-24	28-29	13-1		24-40	13-1

Coloured cells indicate the presence of a consensus allele at a locus and the same at another locus. Uncoloured cells indicate loci where a non-consensus allele was identified. The numbers refer to the combination of HVR1 and HVR2. Colours in Table 8.1a and 8.1b do not correspond to the same allele. †, 5' region of *opa* gene present but gene truncated. V, variable. Bold numbers indicate non-consensus HVRs.

Table 8.1b HVR1-HVR2 amino acid combinations of alleles from the 27 ST261/*opa* -type 4 isolates chosen to be representative of the sampling period

Isolate	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus allele	2	11	17	17	30	37	46	96	62	68	V
Consensus HVR combination	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	V
S3	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
S21	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-22
S24	2-2	8-10	13-15	13-15	20-22	23-19	**53-48**	35-22	38-39	35-22	35-22
S26	2-2	8-10	13-15	13-15	20-22	23-19	**53-48**	35-22	38-39	35-22	35-22
S29	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
S69	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-10
S70	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-10
S74	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
S75	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
S89	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	**55**-22
S94	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
S107	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	13-15
S117	2-2	8-10	13-15	13-15	20-22	23-19	29-30	13-15	38-39	35-22	29-30
S121	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	35-22
S123	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	8-10
S128	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-**51**	38-**52**	35-22	29-30
S129	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-**51**	38-**52**	35-22	29-30
S136	2-2	8-10	13-15	**52**-15	20-22	23-19	29-30	29-30	38-39	35-22	**52**-15
S137	2-2	8-10	13-15	**52**-15	20-22	23-19	29-30	**52**-15	38-39	35-22	**52**-15
S140	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	13-15
S156	2-2	8-10	13-15	13-15	35-22	23-19	29-30	35-22	38-39	35-22	8-10
S196	2-2	8-10	13-15	13-15	20-22	23-19	29-30	13-15	38-39	35-22	8-10
S200	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	8-10
S233	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	13-15	38-39	35-22	13-15
S243	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	20-15
S300	2-2	8-10	13-15	13-15	20-22	23-19	29-30	13-15	38-39	35-22	20-15
S324	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	35-22

As in Table 8.1a.

8.3.3 Differences in HVRs between sexual contacts and within sexual networks

In the earlier chapters, several sexual contacts comprising pairs, triplets and small networks in London were defined and recombination was identified as the main mechanism behind changes in their *opa* gene profiles. I now look briefly at what these genetic changes mean in terms of the HVRs. Tables 8.2a and b and 8.3a-g show the pairs and triplets in Sheffield and the small sexual networks in London respectively. The differences between the sexual contacts are again boxed by bold lines (in Tables 8.2 a and b), however the changes that occurred are discussed in more detail below rather than indicated on the tables.

Among the Sheffield and London sexual networks, 31 changes of *opa* alleles were identified, the majority of which (n=24), were assigned to localised recombination between existing alleles, including three allele duplications. Five were assigned to importation from another strain and only two due to point mutation. Of these 31 changes, 17 retained or resulted in regeneration of a consensus HVR combination (or the HVR combination of the assigned ancestral allele if not a consensus one e.g. see pair 17 below), three of which were because of allele duplications. A further five changes resulted in new combinations of HVRs present in consensus HVR combinations. There were only nine instances (29%) where a change of *opa* allele resulted in the introduction of new HVRs to that strain (i.e. not necessarily a novel one as it may have been seen elsewhere in another strain) five of which were in *opa* alleles that were considered to have been imported from another strain.

Table 8.2a HVR1-HVR2 amino acid combinations of alleles from the ST12 isolates from recent sexual contacts

		Locus										
Pair/ triplet	Isolate	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus allele		1	10	16	25	29	36	45	54	†	67	V
combination		1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	V
Triplet1	**S149**	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
	S152	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
	S154	1-1	1-14	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-1
Triplet 2	S306	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-1
	S308	1-1	1-9	1-14	1-14	13-**53**	22-24	28-29	13-1		24-40	13-9
	S317	1-1	1-9	1-14	1-14	13-**49**	22-24	28-29	13-1		24-40	22-1
Triplet 3	**S445**	1-1	1-9	1-14	1-14	13-16	13-**50**	28-29	13-1		24-40	24-40
	S463	1-1	1-9	1-14	1-14	13-16	13-**50**	28-29	13-1		24-40	24-40
	S466	1-1	1-9	1-14	1-14	13-16	13-**50**	28-29	13-1		24-40	24-40
Pair 1	S78	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
	S86	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-24
Pair 2	S98	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	22-1
	S100	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-1
Pair 3	S241	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	13-16
	S256	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-1
Pair 4	S253	1-1	1-14	1-14	1-14	13-16	13-24	**46**-29	13-1		24-40	13-1
	S257	1-1	1-14	1-14	1-14	13-16	13-24	**46**-29	13-1		24-40	1-14
Pair 5	S493	1-1	1-9	1-14	1-14	13-16	**54-50**	28-29	13-1		24-40	**54**-40
	S514	1-1	1-9	1-14	1-14	13-16	**54-50**	28-29	13-1		24-40	**54**-40
Pair 6	S517	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-1
	S522	1-1	1-9	1-14	1-14	13-16	22-24	28-29	13-1		24-40	1-14

As in Tables 8.1a and b except bold isolate numbers indicate the isolate from the individual who named the other two individuals in the triplet as recent sexual contacts. Blocked cells indicate differences in *opa* gene profiles of isolates from recent sexual contacts.

Table 8.2b HVR1-HVR2 amino acid combinations of alleles from the ST261 isolates from recent sexual contacts

		Locus										
Pair/ triplet	Isolate	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus allele		2	11	17	17	30	37	46	96	62	68	V
combination		2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	V
Pair 7	S21	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-22
	S29	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
Pair 8	S24	2-2	8-10	13-15	13-15	20-22	23-19	**53-48**	35-22	38-39	35-22	35-22
	S26	2-2	8-10	13-15	13-15	20-22	23-19	**53-48**	35-22	38-39	35-22	35-22
Pair 9	S69	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-10
	S70	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	35-10
Pair 10	S74	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
	S75	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	8-10
Pair 11	S89	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	55-22
	S107	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-22	38-39	35-22	13-15
Pair 12	S121	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	35-22
	S123	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-22	38-39	35-22	8-10
Pair 13	S128	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-51	38-**31**	35-22	29-30
	S129	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-51	38-**31**	35-22	29-30
Pair 14	S136	2-2	8-10	13-15	**52**-15	20-22	23-19	29-30	29-30	38-39	35-22	**52**-15
	S137	2-2	8-10	13-15	**52**-15	20-22	23-19	29-30	**52**-15	38-39	35-22	**52**-15
Pair 15	S156	2-2	8-10	13-15	13-15	35-22	23-19	29-30	35-22	38-39	35-22	8-10
	S162	2-2	8-10	13-15	13-15	35-22	23-19	29-30	35-22	38-39	35-22	8-10
Pair 16	S192	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-15	38-39	35-22	2-2
	S196	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-15	38-39	35-22	8-10
Pair 17	S266	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-15	38-39	35-22	2-2
	S271	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-15	38-39	35-22	35-22
Pair 18	S254	2-2	8-10	13-15	13-15	20-22	23-19	**49**-30	35-15	38-39	35-22	13-15
	S258	2-2	8-10	13-15	13-15	20-22	23-19	29-30	35-15	38-39	35-22	35-22

As in Table 8.1a and b. Blocked cells indicate differences in *opa* gene profiles of isolates from recent sexual contacts.

Some of the above observations are described now in a little more detail. In triplet 2 (Table 8.2a), two NS substitutions were identified at *opa5*. One was assigned to mutation, the other to localised recombination. Both however resulted in new HVR2 amino acid sequences, each only differing from the consensus HVR2 by a single amino acid residue.

In pair 14 (Table 8.2b), a change was identified at *opa8*. The allele present in isolate S137 (allele 111, HVR combination 52-15), was assigned as the ancestral allele in this pair as it is present in both isolates at *opa4* and *opa11*. However, at *opa8*, isolate S136 carries allele 135 which has HVR combination 29-30 and is the consensus HVR combination at *opa7* (allele 46). The important thing to note is that as the cell for isolate S136 at *opa8* in Table 8.2b is uncoloured, the allele (allele 135) is not identical to the consensus allele at *opa7* (allele 46), they only share the identical HVR combination. The two alleles at *opa8* only have the 5' region of the gene in common. It appears that a localised recombination event occurred between the allele that was present at this locus and the allele present at *opa7* (Figure. 8.1). The first cross over point probably occurred before the HVRs and the other after, such that both HVR1 and HVR2 were replaced.

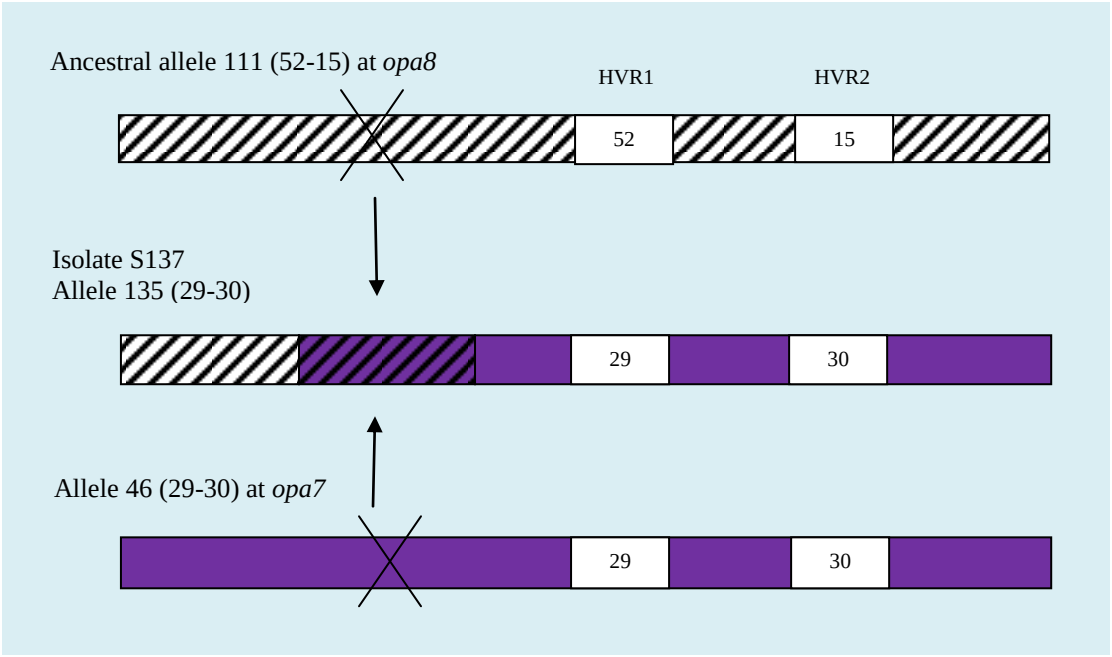


Figure 8.1 Diagrammatic representation of the recombination event in pair 14. The HVR combination is shown in brackets. Allele 111 (hashed) was identified as the ancestral allele at *opa8.* A localised recombination event between this allele and allele 46 (purple, the consensus allele at *opa7*) could result in the generation of allele 135 in isolate S137 such that the new allele has the identical HVR combination as that of allele 46. The coloured and hashed area indicates the region of the genes that are identical and the cross indicates that the cross over point probably occurred within this region of the gene. The other cross over point probably occurred after the HVRs.

In pair 17, a change also occurred at *opa8* which was assigned to localised recombination between the allele identified as the ancestral allele at this locus (allele 133, HVR combination 35-15), present in this and other isolates (although not a consensus allele), and the consensus allele present at *opa3* and *opa4* (allele17). It did not however result in a change to the HVR combination. The first cross over point probably occurred before the HVRs and the second further upstream such that HVR1 and HVR2 were not exchanged (Figure. 8.2).

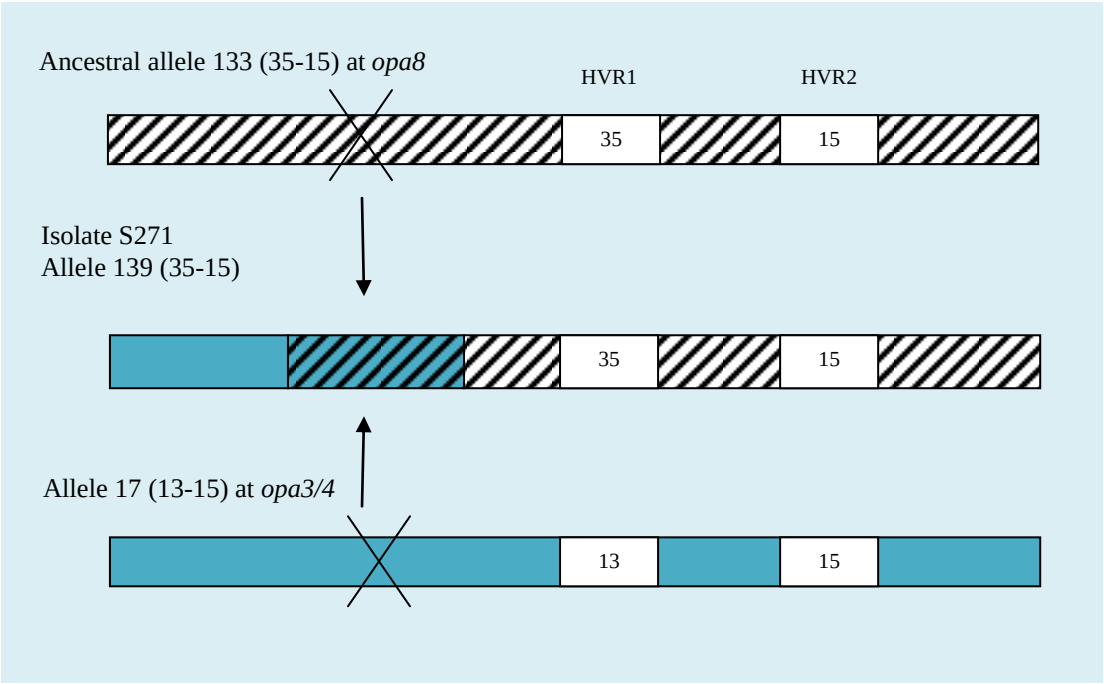


Figure 8.2 Diagrammatic representation of the recombination event in pair 17. The HVR combination is shown in brackets. Allele 133 (hashed) was identified as the ancestral allele at *opa8.* A localised recombination event between this allele and allele 17 (blue, the consensus allele at *opa3* and *opa4*) could result in the generation of allele 139 in isolate S271 such that the new allele has the identical HVR combination as that of allele 133 The coloured and hashed area indicates the region of the genes that are identical and the cross indicates that the cross over point probably occurred within this region. The other cross over point probably lies further upstream.

In pair 18, the change occurred at *opa7* (consensus allele 46), and a recombination event was assigned as the mechanism that introduced a non-consensus allele (allele 95). This non-consensus allele differs to the consensus allele by only a single amino acid in HVR1 and (the allele) is present in other isolates of this NG-MAST ST but is not seen in this pair. Therefore the cross over points could have occurred in conserved regions either side of HVR1 such that only HVR1 was exchanged (Figure. 8.3). Alternatively, a whole gene replacement event could have occurred in which the cross over points were at either end or outside the regions of the *opa* genes. It is not possible to distinguish in this case.

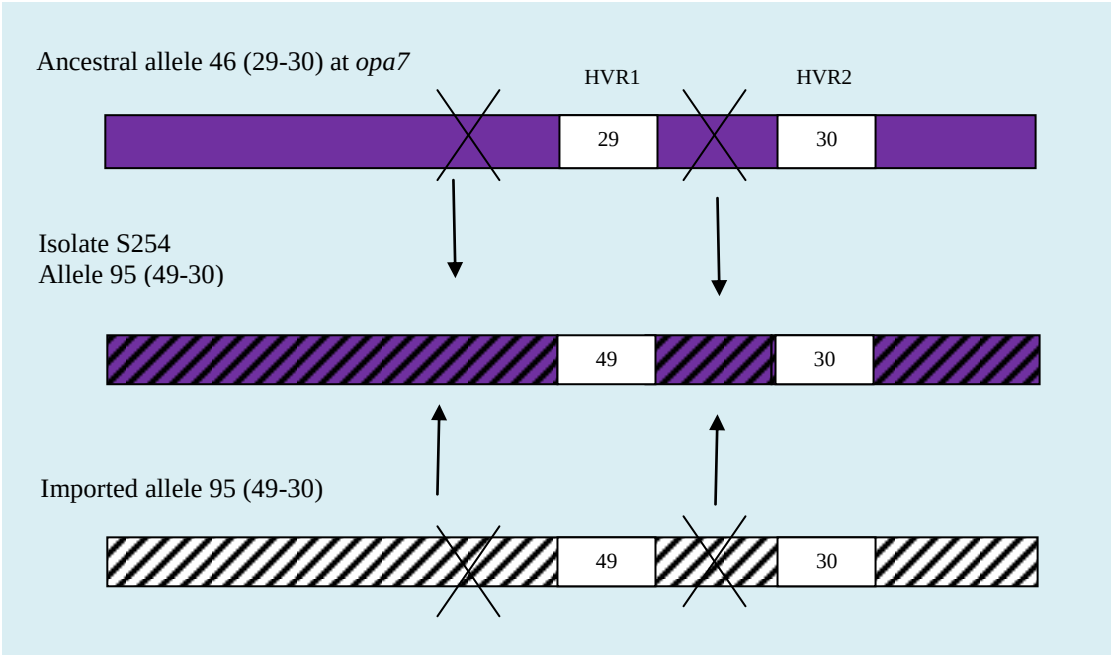


Figure 8.3 Diagrammatic representation of the recombination event in pair 18. The HVR combination is shown in brackets. Allele 46 (purple) was identified as the ancestral allele at *opa7*. A recombination event between this allele and allele 95 (hashed, previously seen at this locus in other isolates of ST261) could result in the re-generation of allele 95 such that the new allele has the HVR1 of allele 95. Allele 95 from isolate S254 is coloured and hashed as these two alleles are identical except for the HVR1s which differ by a single amino acid indel. The cross indicates where the cross over points could have occurred, however, it is also possible that an allele duplication event occurred via a whole gene replacement recombination event.

Amongst the small sexual networks in London, there are several more events such as those described above in Sheffield and it is not necessary to repeat the process of explaining how these probably arose. Only examples of previously uncharacterised events are now described. For example, in MSM cluster ST825 (Table 8.3b), there is an example of a whole gene replacement event in isolate LG1930 at *opa8*. In this case neither of the HVRs were shared and so the cross over points would have had to have occurred beyond the points of any differences, probably in the smaller conserved regions at either end of the *opa* gene.

The only other changes worthy of further explanation is that of the four suspected imported alleles in heterosexual cluster ST610. In isolate LG1081, there are novel alleles at *opa1* and *opa2*. In these cases, none of the alleles in question have any of

the HVRs present in this network. Here again, in order to have incorporated different HVRs, the cross-over points would probably have had to have occurred at the endpoints of the gene where the sequences are far more conserved between different *opa* genes. In the case of isolate LG1081, allele 183 at *opa7* is seen in other STs and was assigned as probably due to importation. Its identification at the same locus suggests that the recombination cross over points could have occurred beyond the start and end points of the *opa* gene, depending on the degree of sequence similarity in these regions between different strains.

Table 8.3 HVR1-HVR2 amino acid combinations of alleles from isolates from individuals in London recently infected with the same genotype by NG-MAST

Table 8.3a MSM cluster ST802

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		203	149	167	13	13	143	49	167	63	174	167
Consensus HVR combination		5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18
LG0983 (M)	31/08/2004	5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18
W0690 (M)	11/10/2004	5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18
LG1497 (M)	18/10/2004	5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18
W0735 (M)	19/10/2004	5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18
LG1682 (M)	03/11/2004	5-32	32-32	67-18	10-5	10-5	25-26	32-32	67-18	15-37	10-43	67-18

Table 8.3b MSM cluster ST825

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		199	148	199	195	156	142	190	166	179	173	V
Consensus HVR combination		71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	V
LG1930 (M)	09/09/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	57-54	18-72	71-54	57-54
W0494 (M)	16/09/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	12-58
W0577 (M)	28/09/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	66-65
LG1304 (M)	30/09/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	66-65
LG1471 (M)	13/10/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	66-65
LG1526 (M)	18/10/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	66-65
LG1604 (M)	27/10/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	66-65
LG1775 (M)	11/11/2004	71-54	57-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	57-58
LG1846 (M)	25/11/2004	71-54	12-58	71-54	85-82	12-58	57-54	83-65	66-65	18-72	71-54	12-58

Table 8.3c MSM cluster ST876

		Locus										
Isolate (gender)	Isolation date	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		168	150	200	150	157	144	192	185	180	175	168
Consensus HVR combination		68-66	60-59	87-19	60-59	58-55	58-55	38-39	58-19	76-73	73-69	68-66
W0231 (M)	14/07/2004	68-66	60-59	87-19	60-59	58-55	58-55	38-39	58-19	76-73	73-69	68-66
LG0893 (M)	20/07/2004	68-66	60-59	87-19	60-59	58-55	58-55	38-39	58-19	76-73	73-69	68-66
LG0903 (M)	27/07/2004	68-66	60-59	87-19	60-59	58-55	58-55	38-39	58-19	76-73	73-69	68-66
LG0906 (M)	28/07/2004	68-66	60-59	87-19	60-59	58-55	58-55	38-39	58-19	76-73	73-69	68-66

As in Table 5. The numbers refer to the combination of HVR1 and HVR2. Double underlined allele numbers indicate alleles which are seen in another ST. Bold numbers indicate non-consensus HVRs. Red bold numbers contain sequences from an unknown source, presumably as a result of a mixed infection. X, an allele number could not be assigned due to a short stretch of sequence in the 5' region which could not be read and which is probably due to secondary structure in the template.

Table 8.3 *continued*

Table 8.3d Heterosexual cluster ST610

Isolate (gender)	Isolation date	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		204	151	201	196	158	145	193	V	181	176	V
Consensus HVR combination		81-79	61-60	62-60	80-78	61-60	59-56	77-74	V	77-74	12-63	V
LG0104 (M)	10/06/2004	80-79	61-60	62-60	80-78	61-60	59-56	**29-89**	**59-77**	77-74	12-63	80-79
W0182 (F)	07/07/2004	81-79	62-60	62-60	80-78	61-63	59-56	77-74	12-63	77-74	12-63	**69-67**
W0236 (F)	14/07/2004	81-79	61-60	62-60	80-78	61-60	59-56	77-74	59-56	77-74	12-63	59-56
LG0540 (M)	22/07/2004	81-79	61-60	62-60	80-78	61-60	59-56	77-74	59-56	77-74	12-63	59-56
LG1081 (M)	06/09/2004	**91-87**	**92-87**	62-60	80-78	61-63	59-56	**79-75**	12-63	77-74	12-63	**92-87**

Table 8.3e Heterosexual cluster ST709

Isolate (gender)	Isolation date	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		205	153	202	197	160	146	194	188	182	177	V
Consensus HVR combination		89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	V
LG0643 (F)	29/07/2004	89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	X
LG0718 (M)	09/08/2004	89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	89-85
LG0923 (M)	26/08/2004	89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	86-83
LG1067 (M)	26/08/2004	89-85	**93**-83	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	82-80
LG1062 (F)	06/09/2004	89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	86-61
LG1569 (M)	29/10/2004	89-85	63-61	88-82	86-83	65-64	12-57	84-81	82-80	78-30	74-70	82-80

Table 8.3f Heterosexual cluster ST777

Isolate (gender)	Isolation date	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		206	147	171	198	161	147	183	154	183	178	171
Consensus HVR combination		90-71	1-14	70-61	73-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61
W0287 (M)	29/07/2004	90-71	64-62	70-61	1-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61
LG0954 (F)	26/08/2004	90-71	1-14	70-61	73-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61
LG1082 (F)	08/09/2004	90-71	64-62	70-61	73-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61
W0457 (M)	10/09/2004	90-71	1-14	70-61	73-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61
W0656 (M)	05/10/2004	90-71	1-14	70-61	73-78	64-62	1-14	79-75	64-62	79-75	75-71	70-61

Table 8.3g Mixed sexual orientation cluster ST584

Isolate (gender)	Isolation date	Locus *opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
Consensus alleles		207	154	171	198	164	147	183	163	184	178	V
Consensus HVR combination		73-86	64-62	70-61	73-78	11-12	1-14	79-75	64-62	73-76	75-71	V
LG0161 (M)	15/06/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	64-62	73-76	75-71	64-62
LG0159 (F)	15/06/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	64-62	73-76	75-71	64-62
LG0201 (M)	21/06/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	79-75	73-76	75-71	73-76
LG0214 (M)	21/06/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	64-62	73-76	75-71	64-62
LG0456 (F)	13/07/2004	73-86	64-62	70-**88**	73-78	64-62	1-14	79-75	64-62	73-76	75-71	73-76
LG0592 (M)	29/07/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	79-75	73-76	75-71	64-**68**
LG0594 (M)	29/07/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	79-75	73-76	75-71	64-**68**
LG0788 (M)	16/08/2004	73-86	64-62	70-61	73-78	11-12	1-14	79-75	64-62	73-76	75-71	64-62

8.3.4 Amino acid sequences of shared HVRs

The HVR combinations are summarised in a matrix format as appears in Tables 8.4-8.12 (see methods). The HVR combinations with the highest counts primarily represent the consensus HVR combinations (highlighted in green) but, as the other regions of the gene are not taken into account here, they also include the non-consensus alleles which have a consensus HVR combination. Combinations that have a much lower count are those of non-consensus allele HVR combinations. These may be comprised of the following: HVRs present in consensus HVR combinations as well as those that are not i.e. new combinations of consensus HVRs (e.g. Table 8.4 HVR 13-9), combinations with one HVR from a consensus allele combination (e.g. Table 8.4 HVR 40-24), or no HVRs from consensus HVR combinations (e.g. Table 8.4 HVR 54-50). HVRs were often seen to occur in more than one combination within a sexual network and so it was important to examine the degree of similarity, or lack thereof, between these HVRs. A dendrogram based on the number of amino acid differences was drawn for each set of HVR1 and HVR2 amino acid sequences from each sexual network (Figure 8.4 to 8.12). These dendrograms, along with the matrices described above and shown in Table 8.4 to 8.12, were used to determine whether the HVRs that were shared in consensus HVR combinations were very similar or very different, i.e. If HVR1 id1 is associated with HVR2 id 2 and HVR2 id 3, are these two HVR2 sequences very similar or different. There were many shared HVRs that differed by only one or a few amino acids but also many which had several differences. For example in ST12, Table 8.4 and Fig 8.4b, the consensus HVR1 (y axis) id 1 is associated with three consensus HVR2 (x-axis), id 1, 9 and 14. HVR2 id 9 and id 14 differ only by a single amino acid but are both very different to HVR2 id 1. If only the consensus allele HVR combinations are

considered and compared, a total of 17 comparisons (eight HVR1 and nine HVR2) were made. 11 (65%) had very different amino acid sequences and six (35%) differed by only a few (1-3) amino acid residues.

Table 8.4 Matrix of HVR1 and HVR2 amino acid sequences of all ST12 alleles

Count	HVR2										
HVR1	1	9	14	16	24	29	40	47	49	50	53
1	34	25	70								
13	37	1		33	6			1	1	3	1
22	2				23			1			
24							35				
28						29					
40					1		1				
46						2					
54							2			2	

The coloured cells indicate HVR and HVR2 combinations present in consensus alleles.

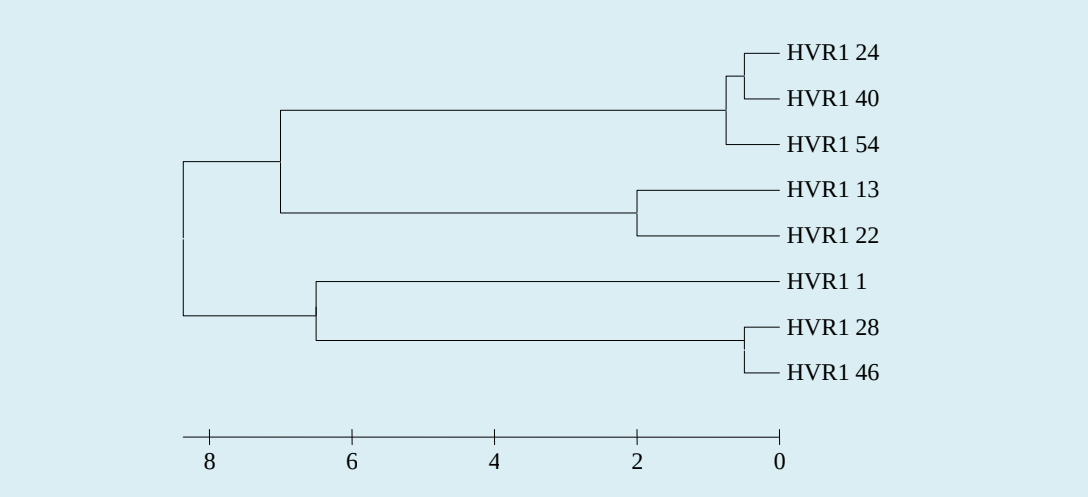


Figure 8.4a Dendrogram of ST12 HVR1 amino acid sequences based on the number of amino acid differences.

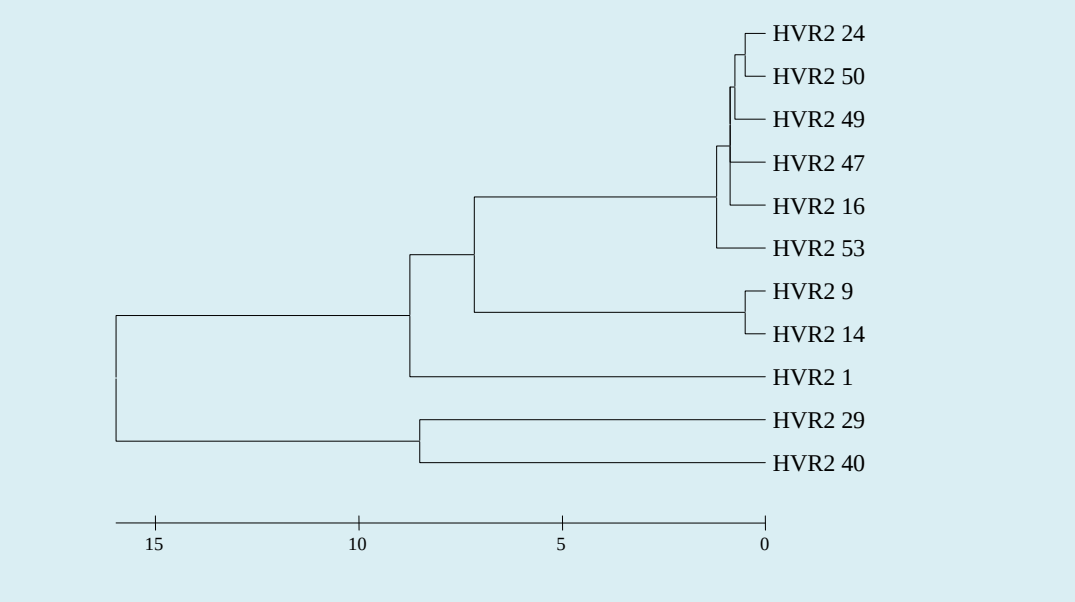


Figure 8.4b Dendrogram of ST12 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.5 Matrix of HVR1 and HVR2 amino acid sequences of all ST261 alleles

Count HVR2

HVR1	2	10	15	19	22	30	31	39	48	51
2	35									
8		43								
13			72							
20			2		31					
23				33						
29						27				
35		2	5		62					2
38							2	31		
49						8				
52			5							
53									2	
55					1					

As in previous table.

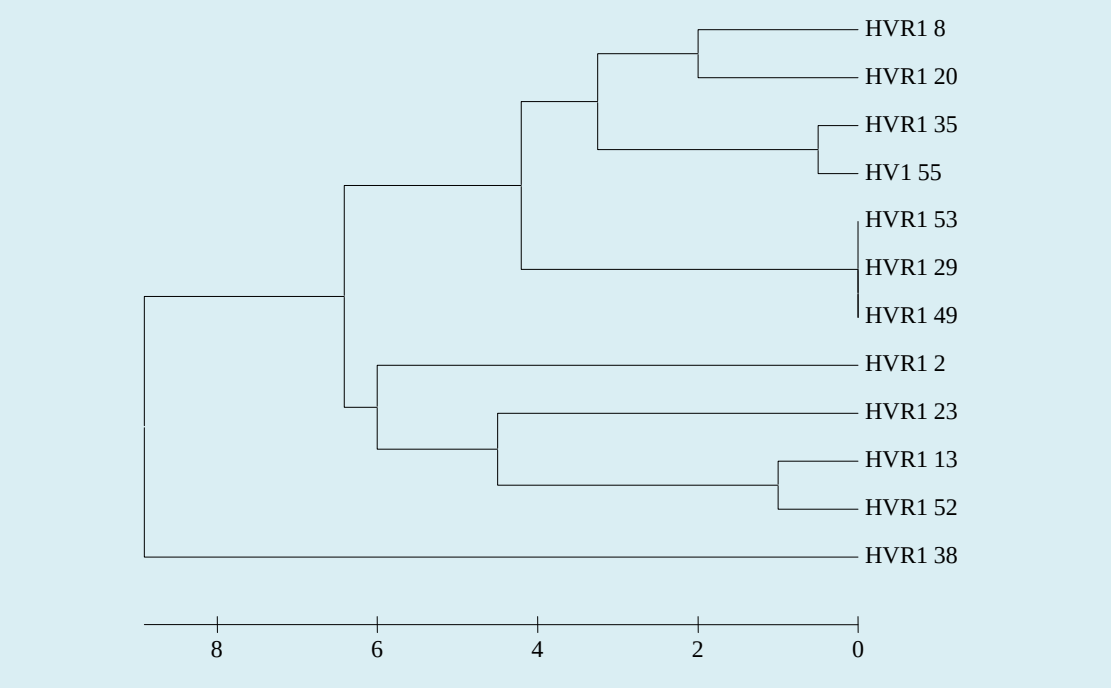


Figure 8.5a Dendrogram of ST261 HVR1 amino acid sequences based on the number of amino acid differences.

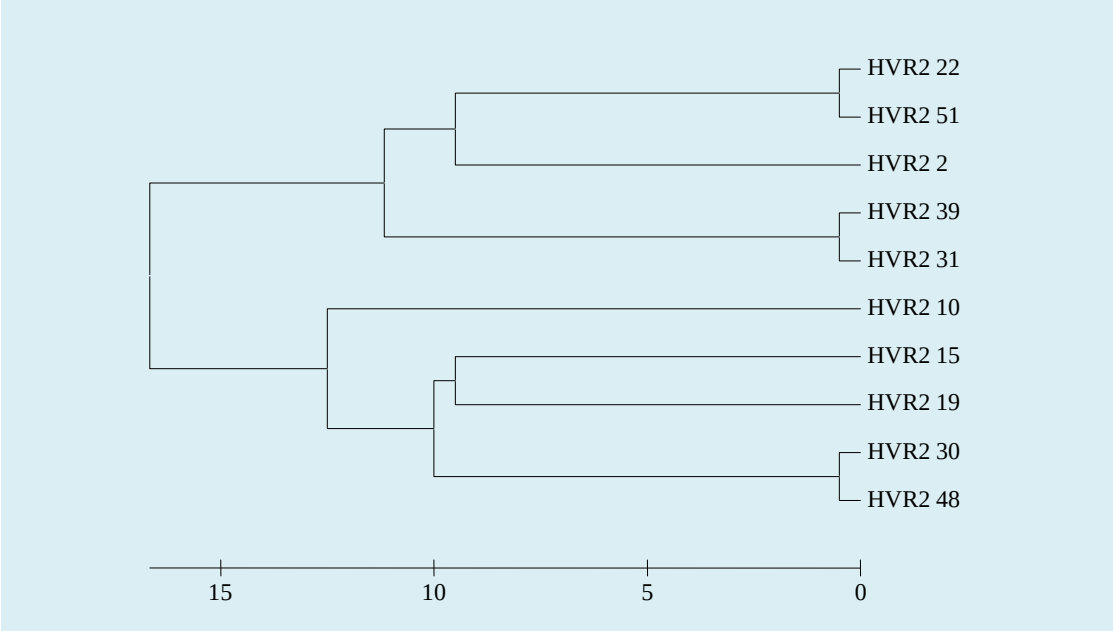


Figure 8.5b Dendrogram of ST261 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.6 Matrix of HVR1 and HVR2 amino acid sequences of all ST802 alleles

Count HVR2

HVR1	5	18	26	32	37	43
5				5		
10	10					5
15					5	
25			5			
32				10		
67		15				

As in previous table.

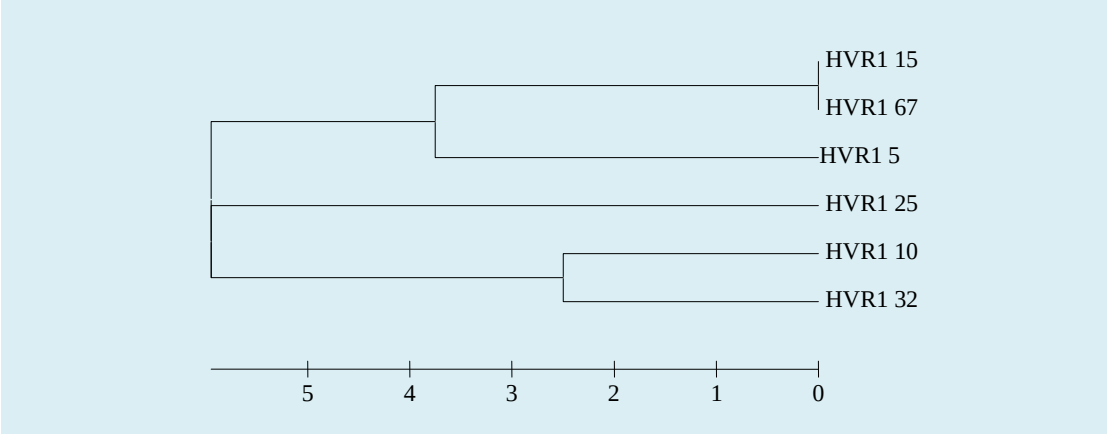


Figure 8.6a Dendrogram of ST802 HVR1 amino acid sequences based on the number of amino acid differences.

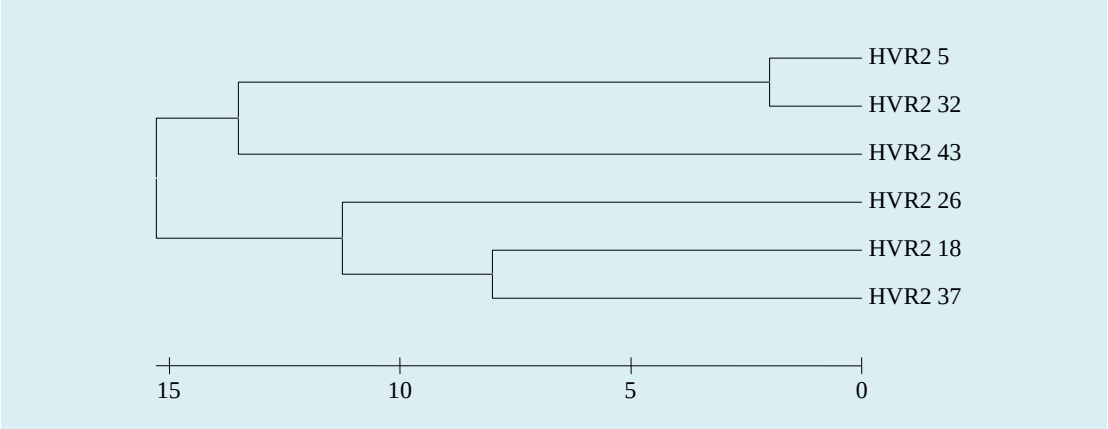


Figure 8.6b Dendrogram of ST802 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.7 Matrix of HVR1 and HVR2 amino acid sequences of all ST825 alleles

Count HVR2

HVR1	54	58	65	72	82
12		19			
18				9	
57	11	2			
66			13		
71	27				
83			9		
85					9

As in previous table.

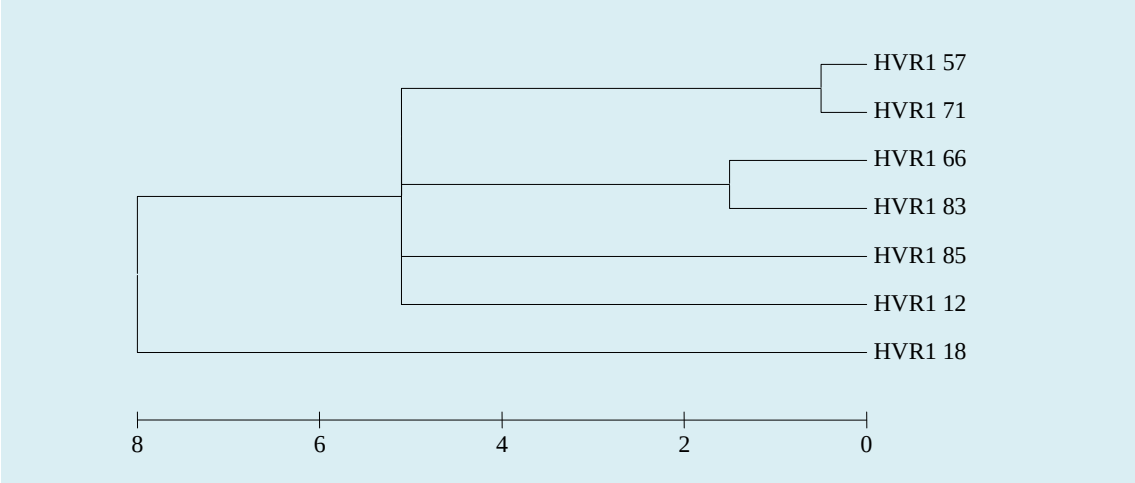


Figure 8.7a Dendrogram of ST825 HVR1 amino acid sequences based on the number of amino acid differences.

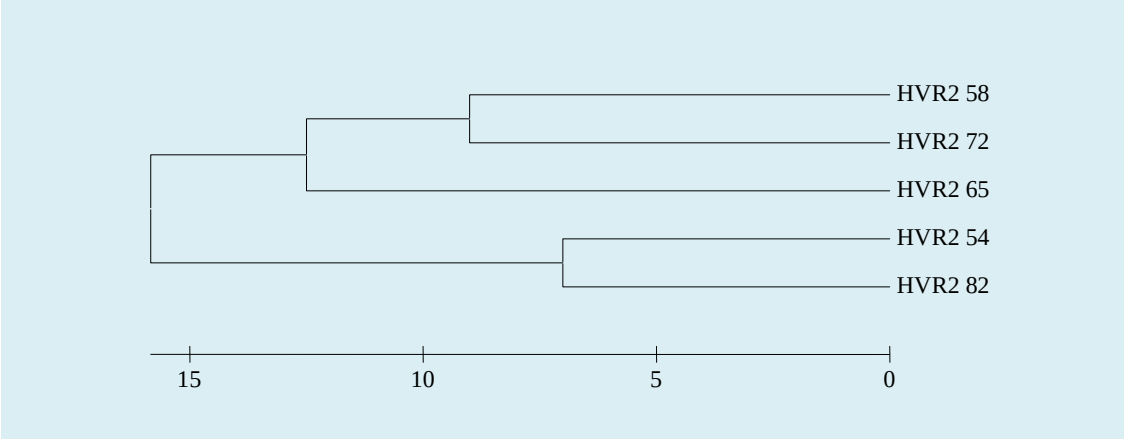


Figure 8.7b Dendrogram of ST825 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.8 Matrix of HVR1 and HVR2 amino acid sequences of all ST876 alleles

Count HVR2

HVR1	19	39	55	59	66	69	73
38		4					
58	4		8				
60				8			
68					8		
73						4	
76							4
87	4						

As in previous table.

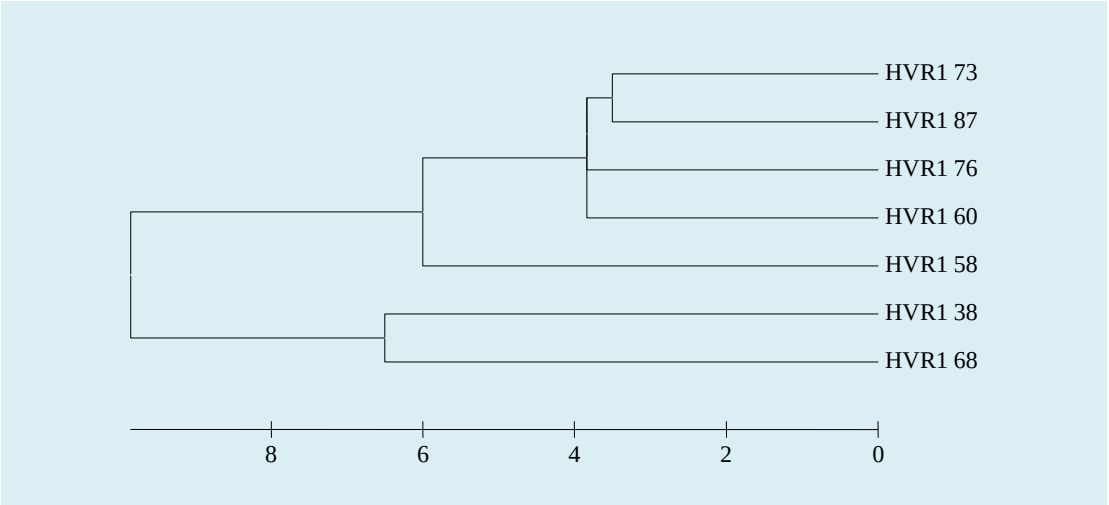


Figure 8.8a Dendrogram of ST876 HVR1 amino acid sequences based on the number of amino acid differences.

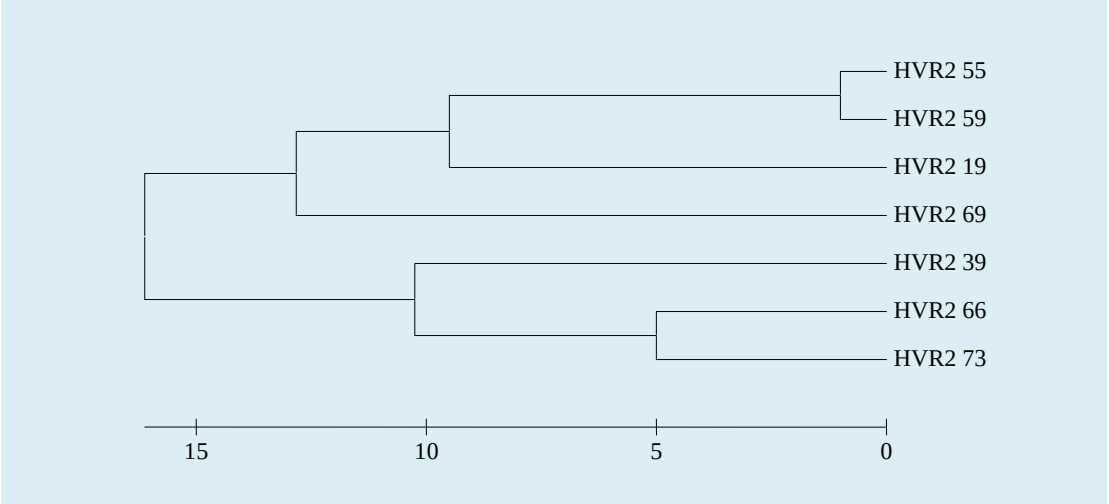


Figure 8.8b Dendrogram of ST876 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.9 Matrix of HVR1 and HVR2 amino acid sequences of all ST610 alleles

Count HVR2

HVR1	56	60	63	67	74	75	77	78	79	87	89
12			7								
29											1
59	9						1				
61		6	2								
62		6									
69				1							
77					8						
79						1					
80								5	2		
81									3		
91										1	
92										2	

As in previous table.

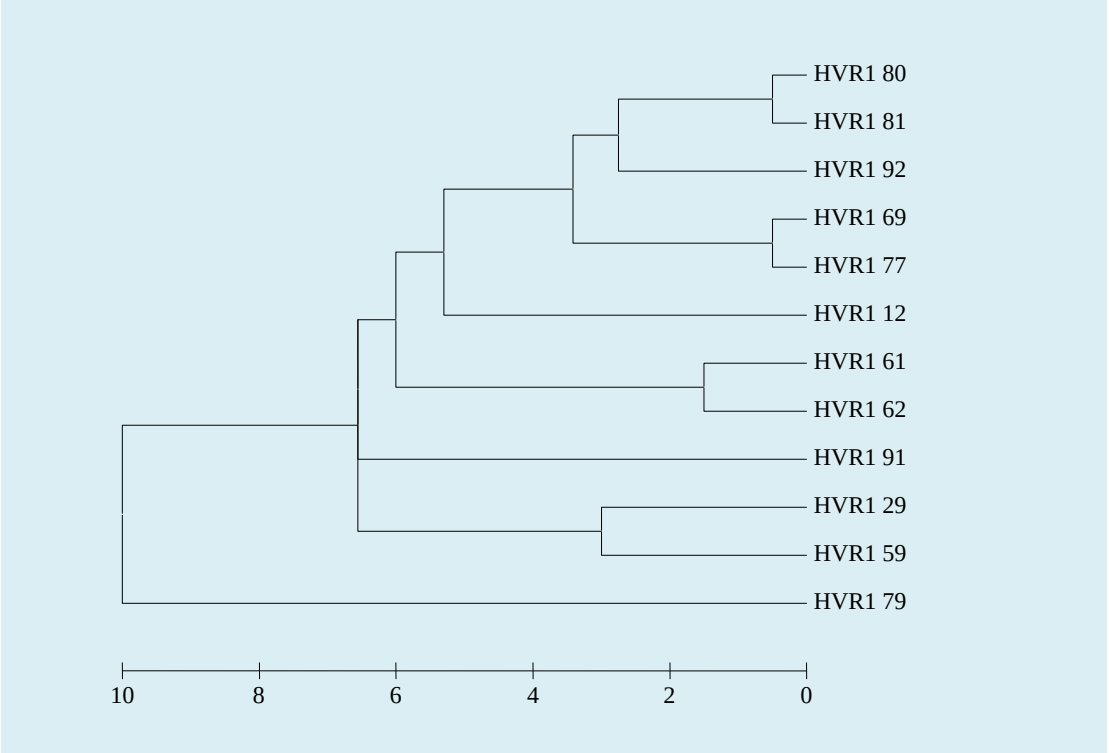


Figure 8.9a Dendrogram of ST610 HVR1 amino acid sequences based on the number of amino acid differences.

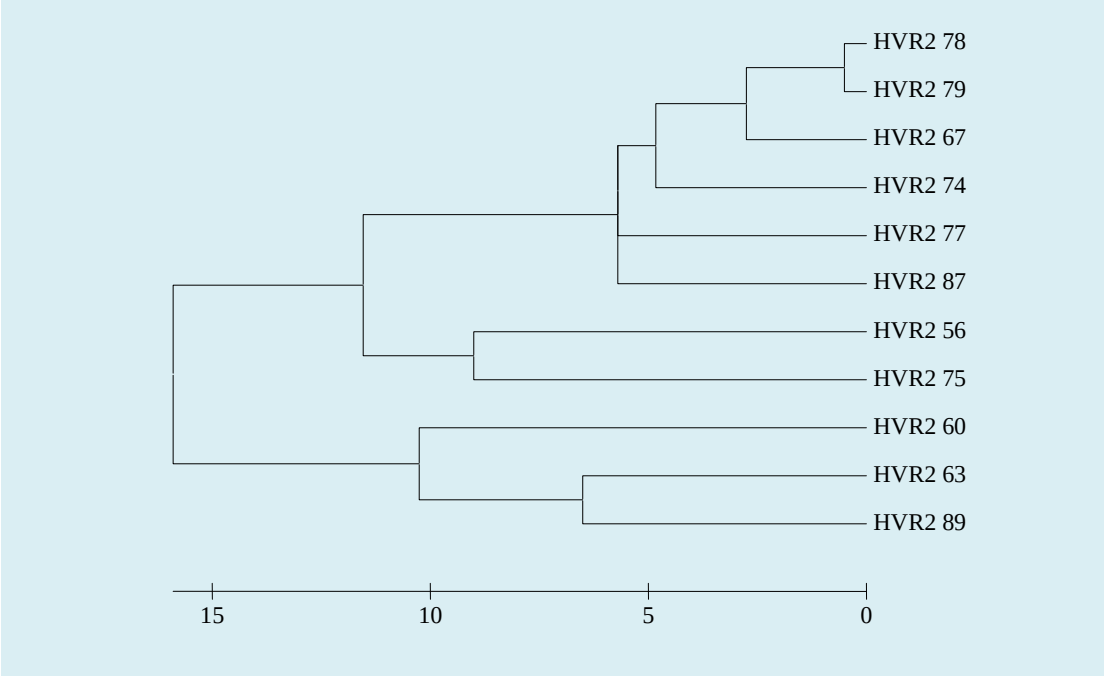


Figure 8.9b Dendrogram of ST610 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.10 Matrix of HVR1 and HVR2 amino acid sequences of all ST709 alleles

Count HVR2

HVR1	30	57	61	64	70	80	81	82	83	85
12		6								
63			5							
65				6						
78	6									
82						8				
84							6			
86			1						7	
88								6		
89										7
93									1	
74					6					

As in previous table.

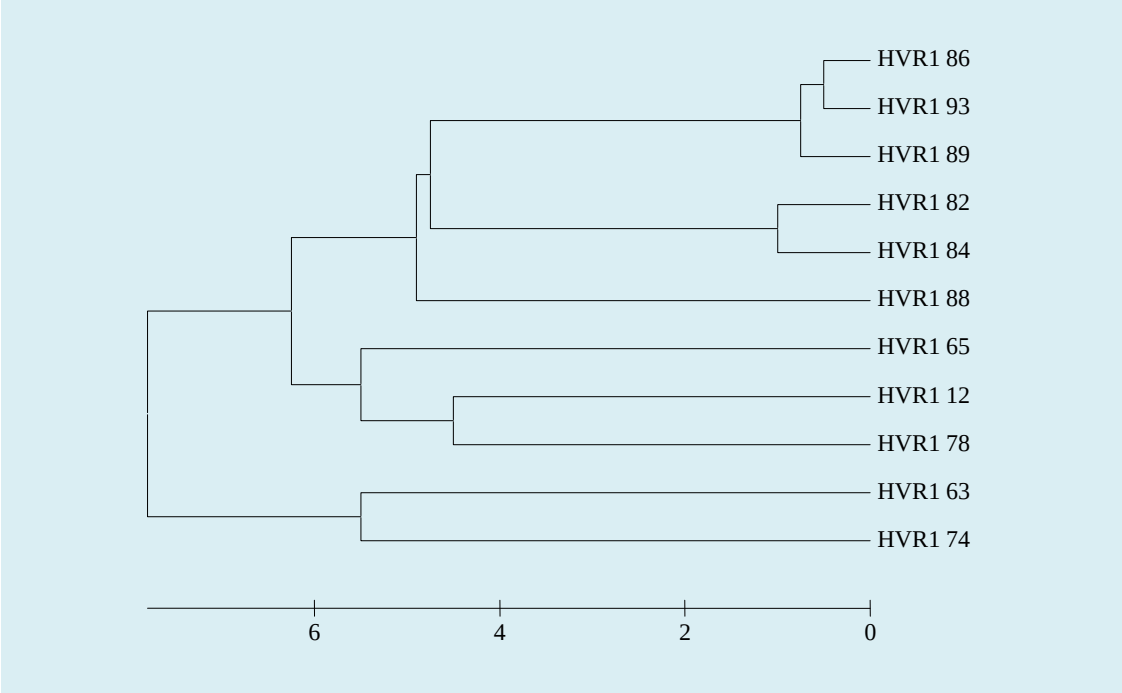


Figure 8.10a Dendrogram of ST709 HVR1 amino acid sequences based on the number of amino acid differences.

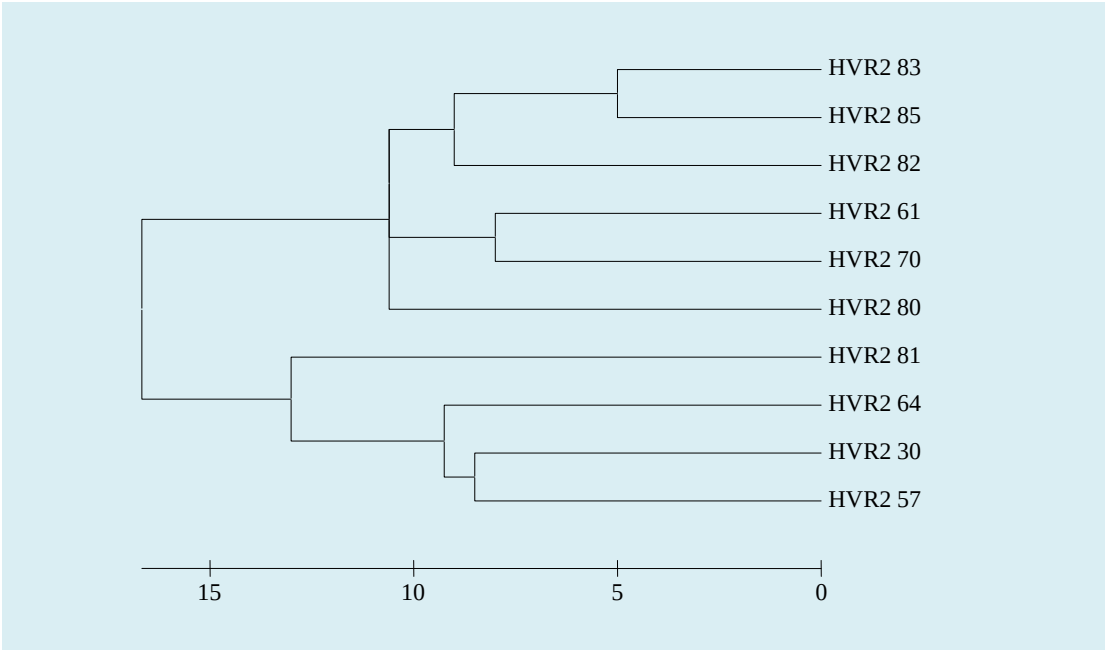


Figure 8.10b Dendrogram of ST709 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.11 Matrix of HVR1 and HVR2 amino acid sequences of all ST777 alleles

Count HVR2

HVR1	14	61	62	71	75	78
1	8					1
64			12			
70		10				
73						4
75				5		
79					10	
90				5		

As in previous table.

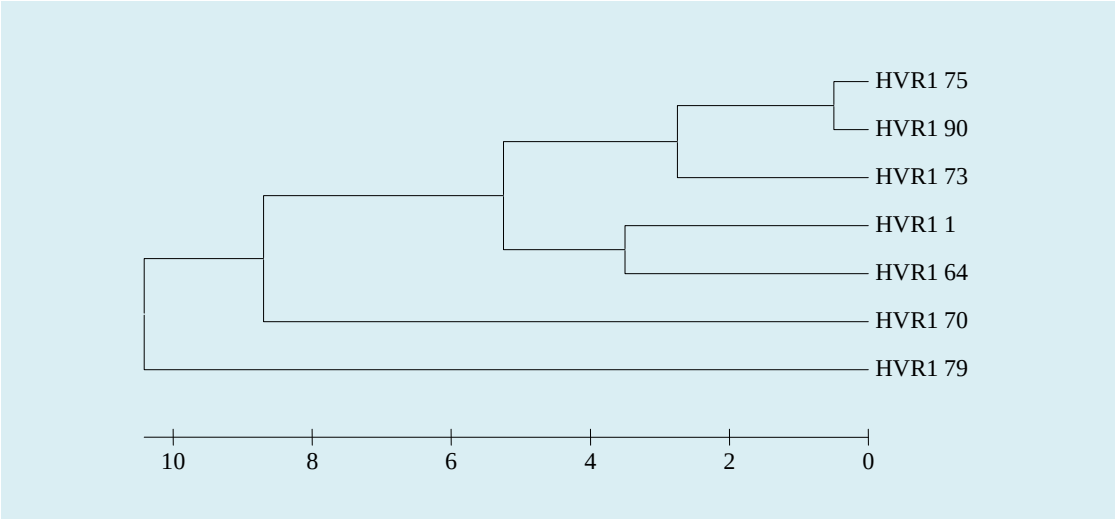


Figure 8.11a Dendrogram of ST777 HVR1 amino acid sequences based on the number of amino acid differences.

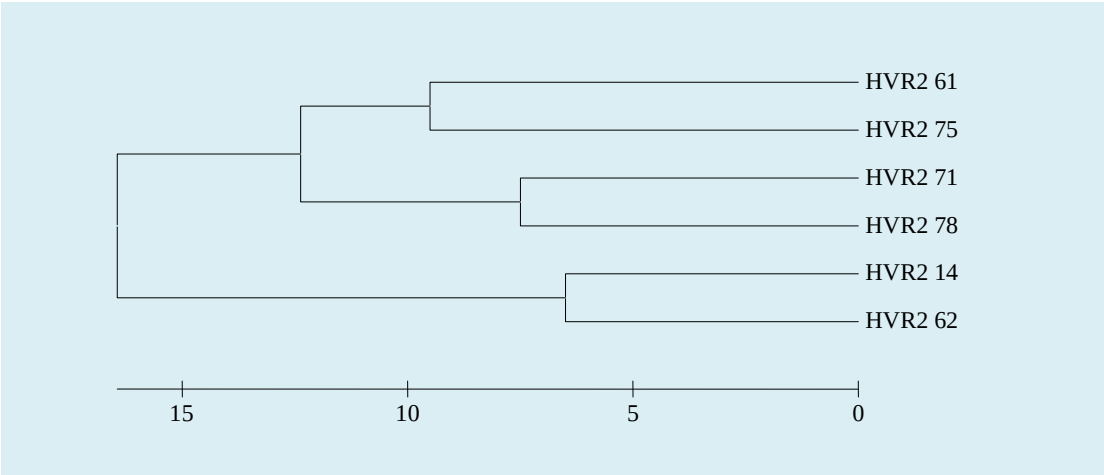


Figure 8.11b Dendrogram of ST777 HVR2 amino acid sequences based on the number of amino acid differences.

Table 8.12 Matrix of HVR1 and HVR2 amino acid sequences of all ST584 alleles

Count HVR2

HVR1	12	14	61	62	68	71	75	76	78	86	88
1		8									
11	7										
64				18	2						
70			7								1
73								10	8	8	
75						8					
79							11				

As in previous table.

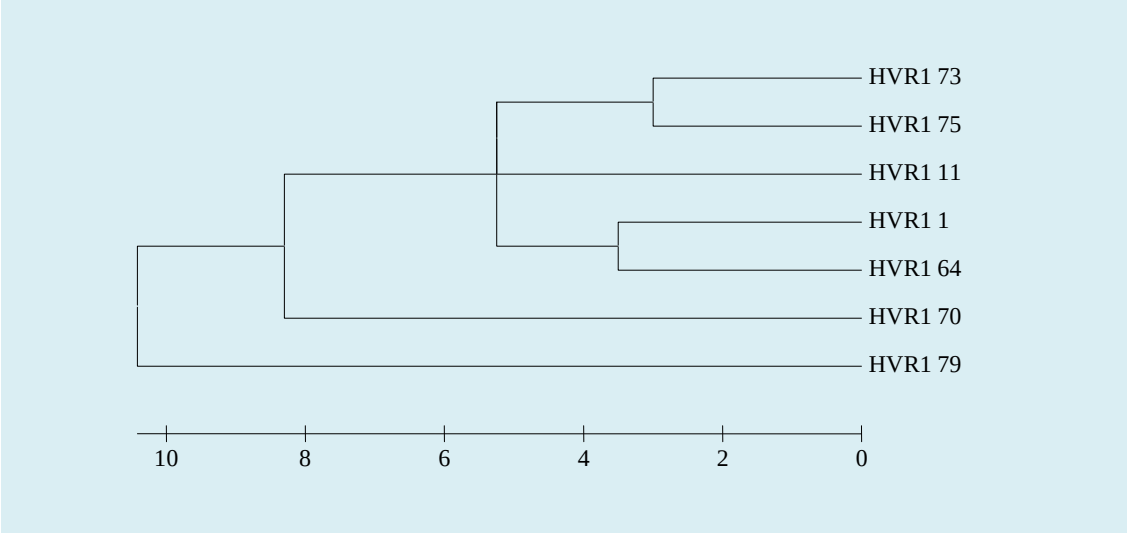


Figure 8.12a Dendrogram of ST584 HVR1 amino acid sequences based on the number of amino acid differences.

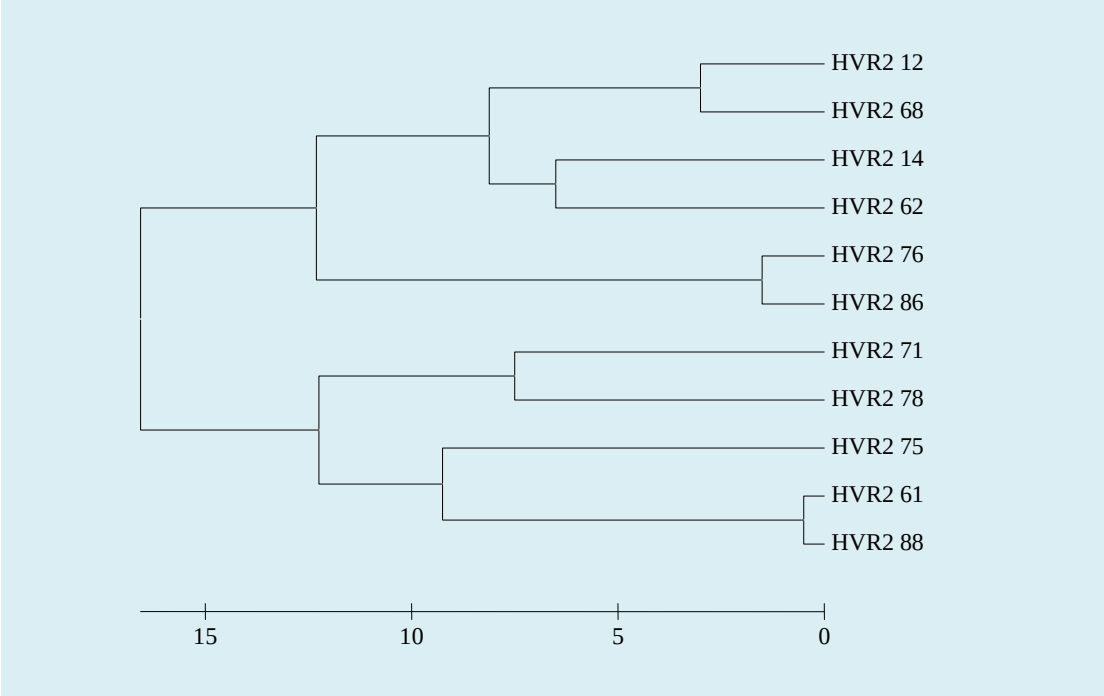
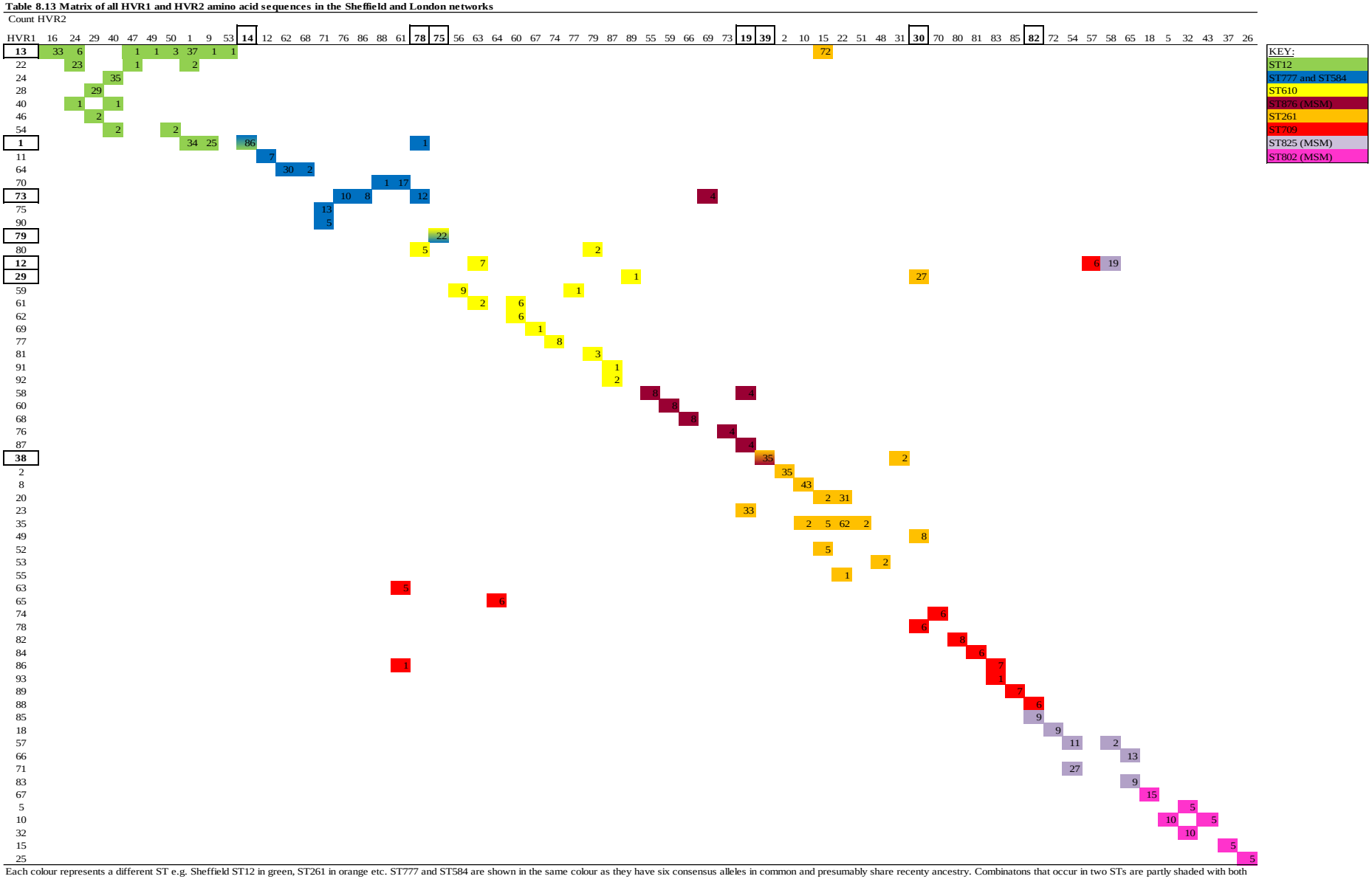


Figure 8.12b Dendrogram of ST584 HVR2 amino acid sequences based on the number of amino acid differences.

8.3.5 Sharing of HVRs between populations

Individual gonococcal clusters (based on NG-MAST genotype) should not only be looked at on their own in such analyses as they are part of a larger gonococcal population. By using NG-MAST STs to identify sexual networks, underlying ancestry may not be evident. A matrix (as done above for each individual network) was therefore made including all the alleles from the ST12 and ST261 networks and all seven London networks (Table 8.13). There were seven HVR1, seven HVR2 amino acid sequences and three (identical) combinations present in multiple distinct sexual networks. The other 57 HVR1 and 57 HVR2 were unique to a sexual network.

Table 8.13 Matrix of all HVR1 and HVR2 amino acid sequences in the Sheffield and London networks



Each colour represents a different ST e.g. Sheffield ST12 in green, ST261 in orange etc. ST777 and ST584 are shown in the same colour as they have six consensus alleles in common and presumably share recenty ancestry. Combinatons that occur in two STs are partly shaded with both colours. The boxed HVR id numbers are those that are present in more than one NG-MAST network.

8.4 Discussion

171 *opa* alleles were identified amongst the 106 isolates comprising the sexual networks studied in Sheffield and London suggesting a great amount of diversity. These were translated into 168 unique amino acid sequences. Despite the many unique sequences, there were only 98 unique combinations of HVR1 and HVR2, with the remaining 70 amino acid sequences having differences in other regions. Given that the two HVRs are expected to be under the most intense selective pressure, one might have expected more unique HVRs and combinations of HVR1 and HVR2. This seemingly high proportion of amino acid differences present in other regions of the gene suggests that future work should also look at the role the SVR more closely and the part that it plays (or not) in virulence or immune evasion. This observation also demonstrates the importance of considering the protein sequences in conjunction with the nucleotide sequences.

Within the networks examined, the presence of identical combinations of HVRs in different consensus alleles (at least one example in all STs except ST709) as well as non-consensus alleles is enough to suggest that within some gonococcal populations (as defined by NG-MAST STs) there may be certain HVR combinations that presumably infer beneficial qualities in terms of protein function and/or immune evasion. Certain HVRs were observed in more than one combination and the fact these include other consensus alleles as well as non consensus alleles clearly demonstrates that these combinations are also permissible and beneficial. It is unclear whether new combinations of HVRs present in consensus alleles will result in antigenically distinct epitopes but presumably, new combinations may result in

antigenically distinct proteins that are seen differently by the host immune system due to conformational changes in the tertiary protein structure. This would particularly be the case if the epitopes recognised by the human immune response were generally conformational involving amino acid residues from both HVR1 and HVR2. It is already known that the combination of HVR1 and HVR2 is important in binding to CEACAM receptors which are present on PMNs and epithelial cells (25, 51) and so would not be surprising if this were the case.

Looking at the changes identified between sexual contacts defined in Sheffield and London (n=31), the majority of *opa* allele changes were assigned to recombination. The three allele duplications would not have a diversifying effect on the *opa* repertoire, other than to potentially reduce it if it removes another Opa protein from the repertoire. Only 50% (n=14) of the 28 events that were expected to bring about change to the repertoire resulted in non-consensus HVR combinations. The rest simply re-generated a consensus HVR combination, although not necessarily the combination identified at that locus. If new combinations are not necessarily assumed to result in new epitopes, then only 32% of the genetic changes resulted in new epitopes being presented to the immune system.

Analysis of *opa* gene alleles showed that *opa11* (and in one case *opa8*, ST610) was far more variable than the other loci. HVR amino acid analysis showed that in fact, most of these unique alleles had consensus HVRs in both consensus and new combinations. There were not notably more novel HVRs identified at this locus as were expected. This suggests that although there are numerous genetic changes occurring at this locus, they have little diversifying effect on individual HVRs.

There is evidence that the combination of HVRs is important (25, 51) and there will undoubtedly exist permissible variations, therefore different combinations should be expected to seen. One might speculate that sharing HVRs that are almost identical would be permissible if they have equal or better receptor binding ability. Conversely, one could also argue that since each HVR is unlikely to be entirely independent of the other HVR in terms of tertiary protein structure, association of a HVR with dissimilar HVRs might result in distinct proteins that can evade host immune system whilst still having binding ability to host receptors. By cross referencing the matrices and the dendrograms drawn, it would seem that these scenarios are not mutually exclusive as there are numerous examples of both amongst the populations studied.

Since each network is not autonomous, all the alleles were combined and treated as one large population. Seven HVR1 and seven HVR2 amino acid sequences were present in more than one sexual network and three HVR combinations were present in more than one network showing that HVRs are sometimes shared, not only between loci within networks, but also between populations and is probably the result of horizontal exchange or unrecognised recent shared ancestry of some strains.

Meningococci, often carried asymptomatically, do not have highly overlapping HVRs in that HVRs are often not present in more than one combination, and this is believed to be due to the effect of the host immune system (29). This is suggested to arise because individuals in the population who have been exposed to a strain expressing a particular meningococcal Opa protein will have antibodies that may recognise epitopes on HVR1 and/or HVR2. Changing one HVR domain may reduce

antibody binding but, as antibodies can still recognise epitopes on the other HVR, such strains will still be at a slight transmission disadvantage in a population where individuals may have antibodies against the original meningococcal strain. Recombination that replaces the whole *opa* gene will generally introduce HVR1 and HVR2 regions that are not recognised by pre-existing antibodies. This is in contrast to gonococci where, due to the much lower population prevalence, the immune system of most individuals will be naive to the infecting strain, such that there are no issues with pre-existing antibodies. However, caution is required when comparing gonococci and meningococci in this way as, in the former, the risk of transmission is highly variable and for subgroups of a population pre-existing antibody may be present. Furthermore there are considerably more *opa* loci in gonococci than meningococci (11 compared to 4) so that the numbers of different HVR1 and HVR2 combinations that can be expressed at the surface are greater in gonococci.

In conclusion, analysis of HVR amino acid sequences shed much light on the effect of the vast genetic differences identified amongst *opa* alleles. Only a small proportion of the numerous genetic events identified between sexual contacts lead to diversification of the HVRs themselves due to the recombination events being re-assortments of existing HVR sequences. The Opa protein repertoire of gonococcal populations showed a preference for certain HVR combinations but sharing of HVRs in different combinations was not precluded and is presumed to be due to beneficial qualities in terms of immune evasion and/or efficient receptor binding. One might speculate that having 11 copies of a gene makes recombination amongst them easier in an already highly recombinogenic organism, generating a vast number of unique nucleotide sequences. However, the need to maintain protein function strongly filters

the numerous genetic events identified to maintain the preferred HVRs and combinations thereof. Although *opa*11 is far more variable than the other *opa* loci, it does not appear to significantly increase diversity in the Opa protein HVR repertoire. Future studies should look closer at the location of other significant differences, namely the SVR and its impact on virulence.

Chapter 9: Final discussion

The aim of this work was to increase the understanding of the extent of diversity among gonococcal *opa* genes and to quantify the relative contributions of mutation and recombination to the diversification of the *opa* allelic profiles of gonococcal strains. In order to achieve this, I developed a method to sequence all the *opa* genes that would permit the analysis of such events and define *opa* gene profiles and repertoires from various groups of isolates via a new nomenclature.

Great diversity was found amongst *opa* gene sequences and *opa* gene repertoires. The set of unrelated strains had none in common and there was very little sharing of *opa* alleles between the different sets of related isolates. In the few instances where it was observed, recent shared ancestry was almost always suspected as the cause rather than horizontal spread. Within the groups of closely related isolates from Sheffield and London, a predominant allele could almost always be detected at each locus such that a consensus allelic profile for a particular ST could be identified.

The combined results from both the Sheffield and London isolates indicated that the processes involved in the diversification of *opa* repertoires were dominated by recombination events between existing alleles (24/31) with importation accounting for 16% of changes and point mutation for 7%. The relative rarity of *de novo* NS mutations in HVRs does not imply that they are unimportant in the evolution of the *opa* repertoire. All of the observed differences between HVRs were ultimately generated by mutation but the appearance of new mutations occurs much less

frequently than the generation of novel mosaic *opa* alleles by the shuffling of variable regions by recombination.

The presence of identical alleles at multiple *opa* loci occurred commonly, without any indication of specific associations between loci, and has also been reported in meningococci (30, 149). In a study of gonococcal pilin antigenic variation, there was evidence that transformation events favour allelic replacement, rather than events which result in *pilE*/*pilS* mosaics, due to the homologous flanking sequences (90). This may be a possible explanation for observing identical *opa* alleles in different STs. However, there may well be differences in the DNA sequences involved in transformation events versus gene conversion events; for example length or degree of homology. The duplicated alleles in question here were consensus alleles observed at multiple loci within the same strain, which may as equally have been due to gene conversion (within the isolate) as to transformation with DNA from sibling bacteria.

The predominance of recombination events that generate mosaic *opa* alleles rather than duplications is not surprising as the alleles were defined by all of the *opa* gene sequences, apart from the region encoding the signal peptide. For an allele to be duplicated, recombination would have to occur within small regions at either end of the *opa* gene (i.e. proximal and distal to the variable regions), whereas mosaic alleles can be formed by recombination within the larger conserved regions within *opa* genes.

Analysis of the translated Opa protein sequences showed that there was an almost equal number of unique DNA and protein sequences but a limited number of HVR1 and HVR2 combinations. More specifically, analysis of the HVR1 and HVR2 amino acid sequences showed that different *opa* alleles frequently encode the same HVR1 and HVR2 sequences or new combinations thereof. Although only 5/31 (16%) changes analysed between isolates from sexual contacts and networks involved new combinations of resident/existing HVRs, many of the *opa* alleles identified also had distinct combinations (and therefore mosaics) of resident/existing HVRs. Therefore crossovers appeared to have occurred within the conserved sequence between the two HVRs, resulting in novel combinations of existing HVRs rather than the creation of novel mosaic (and possibly non-functional) HVRs. Crossovers within a HVR should be rare due to sequence divergence and may be counter-selected due to negative effects on HVR function. Only 29% (9/31) of changes in *opa* alleles resulted in a new HVR and were mainly due to mutation and importation events. The remaining 55% (17/31) of changes resulted in no change to the HVR1 and HVR2 combination at a locus or re-generated an existing combination present in that isolate and therefore crossovers in these instances occurred in the conserved regions before or after the HVRs.

The CEACAM family of receptors which bind to Opa proteins have been shown to require specific combinations of HVR1 and HVR2 (25) and, in gonococci, which can colonise both the male and female genital tract and the rectum (and, at least transiently, the throat); there may be HVRs or HVR1-HVR2 combinations with differential specificities for these different niches (60). The dynamic nature of the *opa* allelic profile may reflect selective forces that favour the duplication of those

opa genes that allow colonisation of a major niche, increasing the chance of expression of such Opa proteins at the cell surface, combined with subsequent selection against these *opa* alleles, and for novel *opa* alleles, imposed by the host immune response. For example, at least in London, individual strains (defined by NG-MAST) circulate either in heterosexual or MSM networks and it would be interesting to examine whether the Opa proteins from 'heterosexual' or 'MSM' strains have differential binding to female and rectal genital mucosa. More generally, gonococci alternate between mucosal sites (e.g. male to female genital tract, or genital tract and rectum) and the advantage of the combinatorial possibilities of a large multi-gene family and the ability to select for Opa proteins that efficiently colonise different mucosal epithelia may have driven the expansion of the gonococcal *opa* system to 11 genes, compared to the four genes in the closely related meningococcus and only one or two in human commensal *Neisseria* species which mainly colonise the nasopharynx. Shuffling of HVRs may therefore be favoured as these existing sequences are likely to bind to their receptors and new combinations may also escape pre-existing immunity if the major epitopes recognised by the host immune system are formed by amino acids contributed by both HVR1 and HVR2.

If transforming DNA is from a sibling gonococcus, it is not possible to distinguish whether an event was due to intra-genomic recombination (within the same bacterium) or transformation. If on the other hand the incoming DNA is not from a sibling gonococcus but from a distinct strain, this should be discernable by the introduction of a novel sequence which cannot be explained as a mosaic of resident alleles. A greater number of strains are likely to be circulating in London and so more such introductions of DNA from distinct strains were expected to be observed

(10). There are a few possible reasons that may explain the apparent lack of observation of such events. One may be due to the selection of isolates that are so closely related for the study, such that importation will undoubtedly play a minor role. Also, when mixed infections occur, some strains may be missed because only one is usually isolated (38). The position of the individual experiencing the mixed infection in the sexual network also contributes to the likelihood of mixed infections (134). Presumably, the higher the number of partners, particularly concurrent partners, the greater the likelihood of a mixed infection occurring.

Furthermore, importation of *opa* sequences that confer advantageous features to the organism may quickly become common, such that the allele would appear to be a consensus allele. Also, as changes to the *opa* repertoire occur, these may at times coincide, with changes and divergence in the *por* and *tbpB* loci (by recombination or mutation) such that the isolate may be assigned a different ST and therefore, observing the introduction of new alleles may not be as easy as expected.

Both *N. meningitidis* and *N. gonorrhoeae* have evolved to have multiple *opa* gene families yet have vastly different pathogenicities. Particular *opa* repertoires can also be associated with meningococcal clones. Meningococcal *opa* repertoires and genotypes can persist over long periods of time, often spanning decades (30, 149), whereas in gonococci the long-term persistence of clonal lineages is seldom observed (161). Meningococcal isolates of the same MLST ST or clonal complex that were recovered over several decades still show considerable similarities in their *opa* allelic profiles (30, 93), with retention of the same allele in one clonal complex over more than 50 years and, in two clonal complexes, of the same *opa* allelic profile

in isolates recovered 25 years apart (30). The sets of meningococcal isolates studied by Callaghan *et al*. are much more distantly related than the sets of gonococcal isolates used in this study, as the former are defined by having the same or similar allelic profiles by MLST, which indexes variation in house-keeping genes that evolve slowly, whereas the latter are defined by the identity of two antigen genes, which evolve much more rapidly. Although the sets of meningococcal isolates are much less closely related than the sets of gonococcal isolates, they appear not to have diversified to a much greater extent than some of the sets of gonococcal isolates. Gonococcal *opa* loci therefore appear to change at a higher rate than those in meningococci and, to at least some extent, this is likely to reflect the increased number of *opa* genes in gonococci, which provide more opportunities for recombination between existing alleles.

The gonococcus has been reported to be polyploid (225) and there is no doubt that if true, the implications for the potential for genetic exchange amongst the multi-gene families could be considerable. Imagine a single chromosome that has most genes in single copy but also has multi-gene families. In a polyploid scenario, single copy genes then become ‘multi-gene families’ but on separate chromosomes and the original multi-gene families become super multi-gene families. In a simple world, each of these chromosomes should be identical. However, when a distinct fragment of DNA is taken up by transformation and replaces the corresponding region of the resident chromosome, the chromosomes become different. Once cell division has occurred, daughter cells may therefore carry differences in their chromosomes.

The rates of change at *opa* loci are difficult to estimate with any precision but they clearly differ markedly between loci. Changes at *opa* loci occur more rapidly than changes at the antigen-encoding genes (*por* and *tbpB*) used to define NG-MAST STs, since multiple isolates of a single ST differed at up to five *opa* loci (excluding *opa11*, which differed in most isolates of the same ST). Changes at *opa11* occurred extremely frequently, with changes occurring in 17/21 (81%) pairs (or triplets) of isolates from recent sexual contacts, whereas no changes at *opa10* were found within any of the sets of closely related gonococci that were examined.

Both here and in a previous study (14), the *opa* locus nearest the *pilE* locus was found to be far more variable than the other *opa* loci. As mentioned before, *pilE* is the major pilin subunit expression site and undergoes high rates of antigenic variation (50) via gene conversion events between itself and numerous silent loci scattered around the chromosome. Analysis of the sequences indicated that the *pilE* and *opa11* loci have similarly high rates of variation, but did not provide evidence for a direct causal relationship between change at one locus and the other. Both loci were also found to vary in culture, providing unequivocal evidence for mosaic *opa* gene formation via recombination within a single population. However, variation occurred at *pilE* more than at *opa11* and not always in identical sub-cultures and therefore also does not reflect any direct relationship in the generation of variation.

There have been many models proposed explaining the mechanism of pilin antigenic variation, some of which complement each other, while others are more conflicting (9, 28, 88, 95). Needless to say, little about the mechanism is known for certain. It is tempting to speculate that the neighbouring *opa11* locus may somehow occasionally,

although opportunistically, be affected by these events leading to the increased variation observed. It is not clear what initiates any of these events, whether they are specific/non-specific, repetitive/conserved sequences. However, what is clear is that the pattern of variation in the *opa* and *pil* gene families have striking similarity in that one locus in each gene family is far more variable. The extent of variation at *pilS* loci has not been measured but they are probably not as variable as *pilE* due to the fact that they are silent and thus under little direct pressure from the immune system to do so. We do not know how rapidly the *pilS* loci change with respect to the *opa* loci or to *pilE,* and it would be helpful if they could be examined along with the *opa* and *pilE* genes. In contrast to the silent *pilS* loci, the *opa* loci are complete genes from which protein is usually expressed and suggests that there must be some process that specifically enhances recombination events in the region of *pilE* and *opa11*.

If the recombinational mechanism by which *pilS* sequences are introduced into the *pilE* locus somehow drives variation at *opa11*, this gene (*opa11)* should vary at a rate similar to that of the other *opa* loci if the adjacent *pilE* gene was not present at this location. The absence of *pilE* at this genomic location in some strains should be investigated more thoroughly to establish if the rate of *opa11* variation is much lower in strains that lack the flanking *pilE* locus.

A more radical suggestion is that *opa11* is preferentially expressed and under far more selection to change than the other *opa* loci. There is currently no evidence for this but there are scenarios under which this may be possible. For example, the strength of various promoters has been associated with increased *opa* phase variation

rates (12). If such a promoter were associated with *opa11*, then the increased rate of phase variation may result in the *opa11* gene being in the 'on' reading frame more often and hence under more selective pressure. Fur, the ferric uptake regulator protein, has a general regulatory function on gene expression. It has been shown to bind to several *opa* gene promoters but it is not known whether it has a positive or negative regulatory role and studies are underway to determine this (188). There may indeed be other factors (like Fur) that interact with the *opa11* promoter and enhance expression. Despite the numerous recombination events occurring at *opa11*, the diversity within the HVR1 and HVR2 protein sequences seems to be modest, further suggesting that there is tempering of Opa protein variation, most likely due to the need to maintain receptor binding ability.

The multiples of CRs also varied on sub-culture leading to the exclusion of this region when assigning allele numbers. The obvious significance of this is that it is not possible to tell which Opa proteins were being expressed *in vivo* once isolates have been cultured from a clinical specimen. In agreement with previous reports, these high rates of phase variation measured mean it is probably virtually impossible for a colony (*in vitro*) to be comprised of gonococci that all have identical phenotypes with respect to the Opa protein(s) on their cell surface, even if their *opa* repertoires may be identical (138). In order to include information on the Opa proteins produced during infection, the *opa* genes would have to be amplified directly without culture, either from the primary isolation plate or preferably from specimens obtained directly from the patient, and include the sequence of the regions encoding the poly A tract and CRs for analysis. Alternatively, reverse transcriptase PCR and sequencing could be used to look at the mRNA transcripts present in an

infection at any one time. Human challenge studies have shown that Opa proteins expressed during the course of infection may vary (104, 215) and, presumably, the infecting gonococcal population may therefore at times be a medley of variants, each expressing a different set of Opa proteins as receptor targets change or selection pressures imposed by the immune system change. Given the rate of change of the CRs, it is likely that the gonococci transmitted to a new host will include individuals expressing each of the Opa proteins, providing the potential for any of these to be selected depending on the ability of the individual Opa proteins to interact with the host mucosal epithelium or any prior immunity. It would be particularly interesting to address the impact of immune selection on diversification but this is not easily done as it would require human challenge studies (41, 42, 186) in which volunteers are repeatedly re-infected with the same strain followed by examination of the Opa proteins produced via one of the methods described above.

Clearly the gonococcus has evolved more than one way to vary its surface antigens. Some might consider it a specialist as there are at least two well studied surface proteins that vary antigenically (pilin and Opa) (147), although some mechanisms may be more efficient than others. There are three basic mechanisms described for antigenic variation amongst large families of surface antigens in general (21). The first is pre-transcriptional in which the coding sequence being expressed is switched. The gonococcal *pil* genes are a good example. By having only one pilin expression locus (except for one reported case of two *pilE* genes) (15, 144) and several silent loci which are recombined into the expression site, only one type of pilin is expressed at any one time (202). The second mechanism is transcriptional control whereby a silent gene is made active via a promoter and vice versa. For example, in

Salmonella, one of two alternative flagellar proteins is expressed depending on the orientation of an invertible segment carrying a promoter (87, 202). Also, variation in the promoter strength of the *opc* locus in meningococci occurs via changes in the length of a homopolymeric tract within the promoter of the gene (87, 183, 227).

The third mechanism is post-transcriptional control e.g. alterations of the reading frame of mRNA transcripts, as in the gonococcal *opa* genes. This means that all the genes are transcribed and seems quite a wasteful mechanism. However, analysis of four publically available neisserial genomes revealed 64 (in addition to *opa* and *pilC*) genes with the potential to be phase variable, underlining the importance of this phenomenon in this genus (107).

Several parallels can be drawn with the parasite *Plasmodium falciparum*, a major cause of malaria. As with gonococci, recurring infection by the pathogen indicates that sophisticated immune evasion mechanisms are in play (184, 186). These parasites spend the majority of their time within erythrocytes (157) and produce variant surface antigens within the *P. falciparum* erythrocyte membrane protein (PfEMP)1, that are transported to the surface of the erythrocyte where they are visible to the immune system (21, 157). This has been found to be necessary to avoid non-specific elimination by the spleen by allowing attachment to vascular endothelium which prevents them reaching the spleen (21, 157). Like Opa proteins, these proteins have an important biological function in terms of adhesion to host cells via specific domains which bind to diverse receptors and are also responsible for the efficient immune evasion (198). They are encoded by the large *var* multi-gene family of which there are approximately 50-60 copies dispersed across the

parasite's 14 linear chromosomes (21, 184). The expression of these genes is switched at high frequency (198).

These parasites either present a single PfEMP1 type (198), or a mixture that is dominated by a particular PfEMP1 type (58). This seems reasonable because if all the variants were expressed at all times, it would not be an effective antigenic variation system (21) and is equally true for the *opa* genes in gonococci. The rate of switching the *var* genes expressed is estimated to be as high as 2.4% per generation *in vitro* making it difficult to obtain large populations expressing only a single *var* gene (198), as does the phase variation of *opa* genes make it difficult to study which Opa proteins are expressed during infection.

How exactly this switching in *P. falciparum* occurs and is controlled is not known for certain, but several mechanisms and models have been suggested, including partial degradation of all but one mRNA transcript under the hypothesis that all the genes are transcribed (21, 58). This method, although wasteful, is not as wasteful as the post-transcriptional control of the gonococcal *opa* genes in which prematurely terminated proteins are generated (21). Additionally, G-quadruplex formation (when G rich sequences can fold into an intra-molecular four stranded structure) in the upstream DNA sequences of *var* genes has been implicated in transcriptional control of this gene family by repressing transcription themselves or by binding ligands (196). The CRs present in the signal peptide encoding region of gonococcal *opa* genes are also believed to be able to form structures (anti-parallel H structures) (11) that may result in changes in the way in which Opa proteins are produced, affecting the reading frame of the gene. Large scale genomic rearrangements do not appear to

be essential or necessary for antigenic switching (198) in *P. falciparum* and switches in transcription could not be associated with changes in promoter sequences and so it has been hypothesised that epigenetic factors control *var* gene expression which depend upon the structure of chromatin and the associated proteins (54).

In *P. falciparum*, the transition from haploid to diploid is believed to partially contribute to gene duplications, mutations and recombination. Duplications are frequently observed amongst *var* genes of some laboratory strains (174) and again mirror the *opa* gene allele duplications described here. Some duplicated genes are mutated or recombine with other genes in the *var* gene family. This high frequency of genetic changes highlights the fact that the gene products are essential for the parasites survival (174). Despite these changes, almost all can bind to a certain receptor (CD36) yet they also bind differentially to other receptors (174). This is similar to the fact that the CEACAM1 receptor binds most Opa proteins while other CEACAM receptors only bind certain Opa proteins (22, 74, 237).

As in gonococci where antigenic variation must be balanced with the ability to bind to necessary host receptors, the endothelial cell receptors available to the parasite will be limited and so variation in these antigens will need to be tempered (21). *In vivo*, the environment, immune pressure, host age, gender and immunity, have an impact on *var* gene selection as will some of these factors affect *opa* gene selection.

N. gonorrhoeae has a relatively small genome, yet the large number of adhesins identified emphasises the importance of host/pathogen interactions (156). Further

work is therefore essential to our understanding of the pathogenesis of infections caused by this organism.

The results presented here strongly support further study into the relevance of variation in other parts of the *opa* gene, particularly the SVR and its role in immune evasion. Recent studies have shown that SVR peptides can elicit antibodies in response to a broad range of Opa variants and may have important implications for vaccine development (44). Additionally, the relationship between *opa11* and *pilE* expression requires further clarification. No direct evidence was found in this work to support a causal relationship, but the elevated variation at *opa11* seems far to pronounced for there to be no contributory effect at all, or for some alternative mechanism, unrelated to the *pil* system, to be involved.

In conclusion, there is very extensive diversity among the *opa* genes of gonococci and unrelated strains are unlikely to share any *opa* alleles. The majority of events which led to the diversification of *opa* allelic profiles were due to recombination, primarily those which resulted in mosaics of the *opa* genes present in the repertoire and allele duplication. Mutations and importation of *opa* sequences from other strains occurred much less frequently than changes due to recombination. Rarely were new HVR sequences observed. HVRs were often re-arranged into new or existing combinations which strongly inferred selection for advantageous HVRs.

Bibliography

1. **Abu-Rajab, K., H. M. Palmer, A. Scoular, S. Church, J. Connell, G. Hart, and A. J. Winter.** 2009. To what extent does *Neisseria gonorrhoeae* multiantigen sequence typing of gonococcal isolates support information derived from patient interviews? Int. J. STD AIDS **20:**414-417.
2. **Aho, E. L.** 1989. Molecular biology of Class 5 outer membrane proteins of *Neisseria meningitidis*. Microb. Pathog. 7:249-253.
3. **Aho, E. L., G. L. Murphy, and J. G. Cannon.** 1987. Distribution of specific DNA sequences among pathogenic and commensal *Neisseria* species. Infect. Immun. **55:**1009-1013.
4. **Alexander, S.** 2009. The challenges of detecting gonorrhoea and chlamydia in rectal and pharyngeal sites: could we, should we, be doing more? Sex. Transm. Infect. **85:**159-160.
5. **Alexander, S., C. A. Ison, J. Parry, C. Llewellyn, S. Wayal, D. Richardson, A. Phillips, H. Smith, and M. Fisher.** 2008. Self-taken pharyngeal and rectal swabs are appropriate for the detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in asymptomatic men who have sex with men. Sex. Transm. Infect. **84:**488-492.
6. **Ashford, W. A., R. G. Golash, and V. G. Hemming.** 1976. Penicillinase-producing *Neisseria gonorrhoeae*. Lancet **2:**657-658.
7. **Bardin, T.** 2003. Gonococcal arthritis. Best Pract. Res. Clin. Rheumatol. **17:**201-208.
8. **Barritt, D. S., R. S. Schwalbe, D. G. Klapper, and J. G. Cannon.** 1987. Antigenic differences among six Protein II expressed by a single strain of *Neisseria gonorrhoeae*. Infect. Immun. **55:**2026-2031.
9. **Barten, R., and T. F. Meyer.** 2001. DNA circle formation in *Neisseria gonorrhoeae*: a possible intermediate in diverse genomic recombination processes. Mol. Gen. Genet. **264:**691-701.
10. **Bell, G., H. Ward, S. Day, A. C. Ghani, U. Goan, E. Claydon, and G. R. Kinghorn.** 1998. Partner notification for gonorrhoea: a comparative study with a provincial and a metropolitan UK clinic. Sex. Transm. Infect. **74:**409-414.
11. **Belland, R. J.** 1991. H-DNA formation by the coding repeat elements of neisserial *opa* genes. Mol. Microbiol. **5:**2351-2360.
12. **Belland, R. J., S. G. Morrison, J. H. Carlson, and D. M. Hogan.** 1997. Promoter strength influences phase variation of neisserial *opa* genes. Mol. Microbiol. **23:**123-135.
13. **Belland, R. J., S. G. Morrison, P. van der Ley, and J. Swanson.** 1989. Expression and phase variation of gonococcal P. II genes in *Escherichia coli* involves ribosomal frameshifting and slipped-strand mispairing. Mol. Microbiol. **3:**777-786.
14. **Bhat, K. S., C. P. Gibbs, O. Barrera, S. G. Morrison, F. Jähnig, A. Stern, E.-M. Kupsch, T. F. Meyer, and J. Swanson.** 1991. The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are encoded by a family of 11 complete genes. Mol. Microbiol. **5:**1889-1901.
15. **Bihlmaieer, A., U. Römling, T. F. Meyer, B. Tümmler, and C. P. Gibbs.** 1991. Physical and genetic map of the *Neisseria gonorrhoeae* strain MS11-N198 chromosome. Mol. Microbiol. **5:**2529-2539.

16. **Bilek, N., I. M. Martin, G. Bell, G. R. Kinghorn, C. A. Ison, and B. G. Spratt.** 2007. Concordance between *Neisseria gonorrhoeae* genotypes recovered from known sexual contacts. J. Clin. Microbiol. **45:**3564-3567.
17. **Billker, O., A. Popp, S. D. Gray-Owen, and T. F. Meyer.** 2000. The structural basis of CEACAM-receptor targeting by neisserial Opa proteins. Trends Microbiol. **8:**258-260.
18. **Biswas, G. D., T. Sox, E. Blackman, and P. F. Sparling.** 1977. Factors affecting genetic transformation of *Neisseria gonorrhoeae*. J. Bacteriol. **129:**983-992.
19. **Biswas, G. D., S. A. Thompson, and P. F. Sparling.** 1989. Gene transfer in *Neisseria gonorrhoeae*. Clin. Microbiol. Rev. **2(Suppl.):**S24-28.
20. **Black, W. J., R. S. Schwalbe, I. Nachamkin, and J. G. Cannon.** 1984. Characterization of *Neisseria gonorrhoeae* protein II phase variation by use of monoclonal antibodies. Infect. Immun. **45:**453-457.
21. **Borst, P., W. Bitter, R. McCulloch, F. Van Leeuwen, and G. Rudenko.** 1995. Antigenic variation in malaria. Cell **82:**1-4.
22. **Bos, M., F. Grunert, and R. Belland.** 1997. Differential recognition of members of the carcinoembryonic antigen family by Opa variants of *Neisseria gonorrhoeae*. Infect. Immun. **65:**2353-2361.
23. **Bos, M. P., D. Hogan, and R. J. Belland.** 1999. Homologue scanning mutagenesis reveals CD66 receptor residues required for neisserial Opa protein binding. J. Exp. Med. **190:**331-340.
24. **Bos, M. P., D. Hogan, and R. J. Belland.** 1997. Selection of Opa+ *Neisseria gonorrhoeae* by limited availability of normal human serum. Infect. Immun. **65:**645-650.
25. **Bos, M. P., D. Kao, D. M. Hogan, C. C. R. Grant, and R. J. Belland.** 2002. Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains. Infect. Immun. **70:**1715-1723.
26. **Boulton, I. C., and S. D. Gray-Owen.** 2002. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. Nat. Immunol. **3:**229-236.
27. **Brooks, G. F., L. Olinger, C. J. Lammel, K. S. Bhat, C. A. Calvello, M. L. Palmer, J. S. Knapp, and R. S. Stephens.** 1991. Prevalence of gene sequences coding for hypervariable regions of Opa (protein II) in *Neisseria gonorrhoeae*. Mol. Microbiol. **5:**3063-3072.
28. **Cahoon, L. A., and H. S. Seifert.** 2009. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. Science **325:**764-767.
29. **Callaghan, M. J., C. O. Buckee, K. A. Jolley, P. Kriz, M. C. J. Maiden, and S. Gupta.** 2008. The effect of immune selection on the structure of the meningococcal Opa protein repertoire. PLoS Pathog. **4:**doi:10.1371/journal.ppat.1000020.
30. **Callaghan, M. J., K. A. Jolley, and M. C. J. Maiden.** 2006. Opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. Infect. Immun. **74:**5085-5094.
31. **Carifo, K., and B. W. Catlin.** 1973. *Neisseria gonorrhoeae* Auxotyping: differentiation of clinical isolates based on growth responses on chemically defined media. Appl. Microbiol. **26:**223-230.

32. **Catlin, B. W.** 1973. Nutritional profiles of *Neisseria gonorrhoeae*, *Neisseria meningitidis*, and *Neisseria lactamica* in chemically defined media and the use of growth requirements for gonococcal typing. J. Infect. Dis. **128:**178-194.
33. C**hen, M. I., A. C. Ghani, and J. Edmunds.** 2008. Mind the gap: the role of time between sex with two consecutive partners on the transmission dynamics of gonorrhea. Sex. Transm. Dis. **35:**435-444.
34. **Chen, T., R. Belland, J. Wilson, and J. Swanson.** 1995. Adherence of pilus- Opa+ gonococci to epithelial cells in vitro involves heparan sulfate. J. Exp. Med. **182:**511-517.
35. **Chen, T., and E. C. Gotschlich.** 1996. CGM1a antigen of neutrophils, a receptor of gonococcal opacity proteins. Proc. Natl. Acad. Sci. U.S.A. **93:**14851-14856.
36. C**hen, T., F. Grunert, A. Medina-Marino, and E. C. Gotschlich.** 1997. Several carcinoembryonic antigens (CD66) serve as receptors for gonococcal opacity proteins. J. Exp. Med. **185:**1557-1564.
37. **Chen, T., J. Swanson, J. Wilson, and R. J. Belland.** 1995. Heparin protects Opa+ *Neisseria gonorrhoeae* from the bactericidal action of normal human serum. Infect. Immun. **63:**1790-1795.
38. **Choudhury, B., C. L. Risley, A. C. Ghani, C. J. Bishop, H. Ward, K. A. Fenton, C. A. Ison, and B. G. Spratt.** 2006. Identification of individuals with gonorrhoea within sexual networks: a population-based study. Lancet **368:**139-146.
39. **Chung, G. T., J. S. Yoo, H. B. Oh, Y. S. Lee, S. H. Cha, S. J. Kim, and C. K. Yoo.** 2008. Complete genome sequence of *Neisseria gonorrhoeae* NCCP11945. J. Bacteriol. **190:**6035-6036.
40. **Clinical Effectiveness Group.** 2005. National guideline on the diagnosis and treatment of gonorrhoea in adults 2005. British Association for Sexual Health and HIV.
41. **Cohen, M. S., and J. G. Cannon.** 1999. Human experimentation with *Neisseria gonorrhoeae*: progress and goals. J. Infect. Dis. **179(Suppl 2):**S375-S379.
42. **Cohen, M. S., J. G. Cannon, A. E. Jerse, L. M. Charniga, S. F. Isbey, and L. G. Whicker.** 1994. Human experimentation with *Neisseria gonorrhoeae*: rationale, methods, and implications for the biology of infection and vaccine development. J. Infect. Dis. **169:**532-537.
43. **Cole, J. G., N. B. Fulcher, and A. E. Jerse.** 2010. Opacity proteins increase *Neisseria gonorrhoeae* fitness in the female genital tract due to a factor under ovarian control. Infect. Immun. **78:**1629-1641.
44. **Cole, J. G., and A. E. Jerse.** 2009. Functional characterization of antibodies against *Neisseria gonorrhoeae* opacity protein loops. PLoS ONE **4:**doi:10.1371/journal.pone.0008108.
45. **Connell, T. D., W. J. Black, T. H. Kawula, D. S. Barritt, J. A. Dempsey, K. Kverneland, Jr., A. Stephenson, B. S. Schepart, G. L. Murphy, and J. G. Cannon.** 1988. Recombination among Protein II genes of *Neisseria gonorrhoeae* generates new coding sequences and increases structural variability in the Protein II family. Mol. Microbiol. **2:**227-236.
46. **Connell, T. D., D. Shaffer, and J. G. Cannon.** 1990. Characterization of the repertoire of hypervariable regions in the Protein II *(opa)* gene family of *Neisseria gonorrhoeae*. Mol. Microbiol. **4:**439-449.

47. **Correia, F. F., S. Inouye, and M. Inouye.** 1986. A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. J. Bacteriol. **167:**1009-1015.
48. **Correia, F. F., S. Inouye, and M. Inouye.** 1988. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. J. Biol. Chem. **263:**12194-12198.
49. **Criss, A. K., B. Z. Katz, and H. S. Seifert.** 2009. Resistance of *Neisseria gonorrhoeae* to non-oxidative killing by adherent human polymorphonuclear leucocytes. Cell. Microbiol. **11:**1074-1087.
50. **Criss, A. K., K. A. Kline, and H. S. Seifert.** 2005. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. Mol. Microbiol. **58:**510-519.
51. **de Jonge, M. I., H. J. Hamstra, L. van Alphen, J. Dankert, and P. van der Ley.** 2003. Mapping the binding domains on meningococcal Opa proteins for CEACAM1 and CEA receptors. Mol. Microbiol. **50:**1005-1015.
52. **Dehio, C., S. D. Gray-Owen, and T. F. Meyer.** 1998. The role of neisserial Opa proteins in interactions with host cells. Trends Microbiol. **6:**489-495.
53. **Dehio, M., O. G. Gómez-Duarte, C. Dehio, and T. F. Meyer.** 1998. Vitronectin-dependent invasion of epithelial cells by *Neisseria gonorrhoeae* involves alpha(v) integrin receptors. FEBS Lett. **424:**84-88.
54. **Deitsch, K. W., A. del Pinal, and T. E. Wellems.** 1999. Intra-cluster recombination and *var* transcription switches in the antigenic variation of *Plasmodium falciparum*. Mol. Biochem. Parasitol. **101:**107-116.
55. **Dempsey, J. A. F., W. Litaker, A. Madhure, T. L. Snodgrass, and J. G. Cannon.** 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.
56. **Dillard, J. P., and H. S. Seifert.** 2001. A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in dissemintated infection isolates. Mol. Microbiol. **41:**263-277.
57. **Djuretic, T., M. Catchpole, J. S. Bingham, A. Robinson, G. Hughes, and G. Kinghorn.** 2001. Genitourinary medicine services in the United Kingdom are failing to meet current demand. Int. J. STD AIDS **12:**571-572.
58. **Duffy, M. F., J. C. Reeder, and G. V. Brown.** 2003. Regulation of antigenic variation in *Plasmodium falciparum*: censoring freedom of expression? Trends Parasitol. **19:**121-124.
59. **Easmon, C. S., C. A. Ison, C. M. Bellinger, and J. W. Harris.** 1982. Emergence of resistance after spectinomycin treatment for gonorrhoea due to beta-lactamase producing strain of *Neisseria gonorrhoeae*. Br. Med. J. **284:**1604-1605.
60. **Edwards, J. L., and M. A. Apicella.** 2004. The molecular mechanisms used by *Neisseria gonorrhoeae* to initiate infection differ between men and women. Clin. Microbiol. Rev. **17:**965-981.
61. **Elkins, C., and R. F. Rest.** 1990. Monoclonal antibodies to outer membrane protein PII block interactions of *Neisseria gonorrhoeae* with human neutrophils. Infect. Immun. **58:**1078-1084.
62. **Elkins, C., C. E. Thomas, H. S. Seifert, and P. F. Sparling.** 1991. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. J. Bacteriol. **173:**3911-3913.

63. **Feil, E., J. Zhou, J. Maynard Smith, and B. G. Spratt.** 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.
64. **Fischer, S. H., and R. F. Rest.** 1988. Gonococci possessing only certain P.II outer membrane proteins interact with human neutrophils. Infect. Immun. **56:**1574-1579.
65. **Fleming, D. T., and J. N. Wasserheit.** 1999. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex. Transm. Infect. **75:**3-17.
66. **Frosch, M., and T. F. Meyer.** 1992. Transformation-mediated exchange of virulence determinants by co-cultivation of pathogenic *Neisseriae*. FEMS Microbiol. Lett. **100:**345-349.
67. **Galimand, M., G. Gerbaud, and P. Courvalin.** 2000. Spectinomycin resistance in *Neisseria* spp. due to mutations in 16S rRNA. Antimicrobial Agents Chemother. **44:**1365-1366.
68. **Gerbase A. C., R. J., Heymann D. H., Berkley S. F., Piot P.** 1998. Global prevalence and incidence estimates of selected curable STDs. Sex. Transm. Infect. **74(Suppl. 1):**S12-S16.
69. **Gibbs, C. P., and T. F. Meyer.** 1996. Genome plasticity in *Neisseria gonorrhoeae*. FEMS Microbiol. Lett. **142:**173-179.
70. **Goodman, S. D., and J. J. Scocca.** 1988. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. U.S.A. **85:**6982-6986.
71. **Grant, C. C. R., M. P. Bos, and R. J. Belland.** 1999. Proteoglycan receptor binding by *Neisseria gonorrhoeae* MS11 is determined by the HV-1 region of OpaA. Mol. Microbiol. **32:**233-242.
72. **Grassmé, H., E. Gulbins, B. Brenner, K. Ferlinz, K. Sandhoff, K. Harzer, F. Lang, and T. F. Meyer.** 1997. Acidic sphingomyelinase mediates entry of *N. gonorrhoeae* into nonphagocytic cells. Cell **91:**605-615.
73. **Gray-Owen, S. D., C. Dehio, A. Haude, F. Grunert, and T. F. Meyer.** 1997. CD66 carcinoembryonic antigens mediate interactions between Opa-expressing *Neisseria gonorrhoeae* and human polymorphonuclear phagocytes. **16:**3435-3445.
74. **Gray-Owen, S. D., D. R. Lorenzen, A. Haude, T. F. Meyer, and C. Dehio.** 1997. Differential Opa specificities for CD66 receptors influence tissue interactions and cellular response to *Neisseria gonorrhoeae*. Mol. Microbiol. **26:**971-980.
75. **Griffiths, N. J., C. J. Bradley, R. S. Heyderman, and M. Virji.** 2007. IFN-γ amplifies NFκB-dependent *Neisseria meningitidis* invasion of epithelial cells via specific upregulation of CEA-related cell adhesion molecule 1. Cell. Microbiol. **9:**2968-2983.
76. **Haas, R., and T. F. Meyer.** 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell **44:**107-115.
77. **Hamilton, H. L., and J. P. Dillard.** 2006. Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. Mol. Microbiol. **59:**376-385.

78. **Hamilton, H. L., N. M. Dominguez, K. J. Schwartz, K. T. Hackett, and J. P. Dillard.** 2005. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. Mol. Microbiol. **55:**1704-1721.
79. **Hamrick, T. S., J. A. F. Dempsey, M. S. Cohen, and J. G. Cannon.** 2001. Antigenic variation of gonococcal pilin expression *in vivo*: analysis of the strain FA1090 pilin repertoire and identification of the pilS gene copies recombining with pilE during experimental human infection. Microbiology **147:**839-849.
80. **Harvey, H. A., M. P. Jennings, C. A. Campbell, R. Williams, and M. A. Apicella.** 2001. Receptor-mediated endocytosis of *Neisseria gonorrhoeae* into primary human urethral epithelial cells: the role of the asialoglycoprotein receptor. Mol. Microbiol. **42:**659-672.
81. **Hauck, C. R., and T. F. Meyer.** 2003. 'Small' talk: Opa proteins as mediators of *Neisseria*-host-cell communication. Curr. Opin. Microbiol. **6:**43-49.
82. **Heckels, J. E.** 1984. Molecular studies on the pathogenesis of gonorrhoea. J. Med. Microbiol. **18:**293-307.
83. **Heckels, J. E.** 1981. Structural comparison of *Neisseria gonorrhoeae* outer membrane proteins. J. Bacteriol. **145:**736-742.
84. **Heckels, J. E., B. Blackett, J. S. Everson, and M. E. Ward.** 1976. The influence of surface charge on the attachment of *Neisseria gonorrhoeae* to human cells. J. Gen. Microbiol. **96:**359-364.
85. **Hedges, S. R., M. S. Mayo, J. Mestecky, E. W. Hook III, and M. W. Russell.** 1999. Limited local and systemic antibody responses to *Neisseria gonorrhoeae* during uncomplicated genital infections. Infect. Immun. **67:**3937-3946.
86. **Hedges, S. R., D. Sibley, M. S. Mayo, E. W. Hook III, and M. W. Russell.** 1998. Cytokine and antibody responses in women infected with *Neisseria gonorrhoeae*: effects of concomitant infections. J. Infect. Dis. **178:**742-751.
87. **Henderson, I. R., P. Owen, and J. P. Nataro.** 1999. Molecular Switches - the ON and OFF of bacterial phase variation. Mol. Microbiol. **33:**919-932.
88. **Hill, S. A., and C. C. Grant.** 2002. Recombinational error and deletion formation in *Neisseria gonorrhoeae*: a role for RecJ in the production of $pilE_L$ deletions. Mol. Genet. Genomics **266:**962-972.
89. **Hill, S. A.** 1999. Cell to cell transmission of donor DNA overcomes differential incorporation of non-homologous and homologous markers in *Neisseria gonorrhoeae*. Gene **240:**175-182.
90. **Hill, S. A.** 1996. Limited variation and maintenance of tight genetic linkage characterize heteroallelic *pilE* recombination following DNA transformation of *Neisseria gonorrhoeae*. Mol. Microbiol. **20:**507-518.
91. **Hill, S. A.** 2000. Opa expression correlates with elevated transformation rates in *Neisseria gonorrhoeae*. J. Bacteriol. **182:**171-178.
92. **Hill, S. A., S. G. Morrison, and J. Swanson.** 1990. The role of direct oligonucleotide repeats in gonococcal pilin gene variation. Mol. Microbiol. **4:**1341-1352.
93. **Hobbs, M., B. Malorny, P. Prasad, G. Morelli, B. Kusecek, J. Heckels, J. Cannon, and M. Achtman.** 1998. Recombinational reassortment among *opa* genes from ET-37 complex *Neisseria meningitidis* isolates of diverse geographical origins. Microbiology **144:**157-166.

94. **Hobbs, M. M., A. Seiler, M. Achtman, and J. G. Cannon.** 1994. Microevolution within a clonal population of pathogenic bacteria: recombination, gene duplication and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. Mol. Microbiol. **12:**171-180.
95. **Howell-Adams, B., and H. S. Seifert.** 2000. Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. Mol. Microbiol. **37:**1146-1158.
96. **Howie, F., H. Young, and A. McMillan.** 2004. The diversity of the *opa* gene in gonococcal isolates from men who have sex with men. Sex. Transm. Infect. **80:**286-288.
97. **Ison, C. A.** 1998. Genotyping of *Neisseria gonorrhoeae*. Curr. Opin. Infect. Dis. **11:**43-46.
98. **Ison, C. A.** 2006. *Neisseria gonorrhoeae*, p. 221-230. *In* S. H. Gillespie and P. M. Hawkey (ed.), Principles and Practice of Clinical Bacteriology, 2nd ed. John Wiley & Sons, Ltd., Chichester, UK. doi:10.1002/9780470017968.ch15.
99. **Ison, C. A., G. Onwuchekwa, and C. S. F. Easmon.** 1990. Epidemiology of *Neisseria gonorrhoeae* isolated in London, p. 75-80. *In* M. Achtman, P. Kohl, C. Marchal, G. Morelli, A. Seiler, and B. Thiesen (ed.), Neisseriae 1990. Walter de Gruyter, Berlin.
100. **Ison, C. A., L. Whitaker, and A. Renton.** 1992. Concordance of auxotype/ serovar classes of *Neisseria gonorrhoeae* between sexual contacts. Epidemiol. Infect. **109:**265-271.
101. **Ison, C. A., P. J. Woodford, H. Madders, and E. Claydon.** 1998. Drift in susceptibility of *Neisseria gonorrhoeae* to ciprofloxacin and emergence of therapeutic failure. Antimicrobial Agents Chemother. **42:**2919-2922.
102. **James, J. F., and J. Swanson.** 1978. Studies on gonococcus infection. XIII. Occurrence of color/opacity colonial variants in clinical cultures. Infect. Immun. **19:**332-340.
103. **James, J. F., E. Zurlinden, C. J. Lammel, and G. F. Brooks.** 1982. Relation of Protein I and colony opacity to serum killing of *Neisseria gonorrhoeae*. J. Infect. Dis. **145:**37-44.
104. **Jerse, A. E., M. S. Cohen, P. M. Drown, L. G. Whicker, S. F. Isbey, H. S. Seifert, and J. G. Cannon.** 1994. Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. J. Exp. Med. **179:**911-920.
105. **Johannsen, D. B., D. M. Johnston, H. O. Koymen, M. S. Cohen, and J. G. Cannon.** 1999. A *Neisseria gonorrhoeae* immunoglobulin A1 protease mutant is infectious in the human challenge model of urethral infection. Infect. Immun. **67:**3009-3013.
106. **Jonsson, A.-B., G. Nyberg, and S. Normark.** 1991. Phase variation of gonococcal pili by frameshift mutation in *pilC*, a novel gene for pilus assembly. EMBO J. **10:**477-488.
107. **Jordan, P., L. Snyder, and N. Saunders.** 2005. Strain-specific differences in *Neisseria gonorrhoeae* associated with the phase variable gene repertoire. BMC Microbiol. **5:**21. doi:10.1186/1471-2180-5-21.
108. **Kallstrom, H., D. Blackmer Gill, B. Albiger, M. K. Liszewski, J. P. Atkinson, and A.-B. Jonsson.** 2001. Attachment of *Neisseria gonorrhoeae* to the cellular pilus receptor CD46: identification of domains important for bacterial adherence. Cell. Microbiol. **3:**133-143.

109. **Kallstrom, H., M. K. Liszewski, J. P. Atkinson, and A. B. Jonsson.** 1997. Membrane cofactor protein (MCP or CD46) is a cellular pilus receptor for pathogenic *Neisseria*. Mol. Microbiol. **25:**639-647.
110. **Kawula, T. H., E. L. Aho, D. S. Barritt, D. G. Klapper, and J. G. Cannon.** 1988. Reversible phase variation of expression of *Neisseria meningitidis* Class 5 outer membrane proteins and their relationship to gonococcal Proteins II. Infect. Immun. **56:**380-386.
111. **Kellogg, D. S., Jr., I. R. Cohen, L. C. Norins, A. L. Schroeter, and G. Reising.** 1968. *Neisseria gonorrhoeae* II. Colonial variation and pathogenicity during 35 months *in vitro*. J. Bacteriol. **96:**596-605.
112. **Kirchner, M., D. Heuer, and T. F. Meyer.** 2005. CD46-independent binding of neisserial type IV pili and the major pilus adhesin, PilC, to human epithelial cells. Infect. Immun. **73:**3072-3082.
113. **Kirchner, M., and T. F. Meyer.** 2005. The PilC adhesin of the *Neisseria* type IV pilus-binding specificities and new insights into the nature of the host cell receptor. Mol. Microbiol. **56:**945-957.
114. **Kline, K. A., E. V. Sechman, E. P. Skaar, and H. S. Seifert.** 2003. Recombination, repair and replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human-specific bacterial pathogens. Mol. Microbiol. **50:**3-13.
115. **Knapp, J. S.** 1988. Historical perspectives and identification of *Neisseria* and related species. Clin. Microbiol. Rev. **1:**415-431.
116. **Knapp, J. S., and K. K. Holmes.** 1975. Disseminated gonococcal infections caused by *Neisseria gonorrhoeae* with unique nutritional requirements. J. Infect. Dis. **132:**204-208.
117. **Knapp, J. S., M. R. Tam, R. C. Nowinski, K. K. Holmes, and E. G. Sandstrom.** 1984. Serological classification of *Neisseria gonorrhoeae* with use of monoclonal antibodies to gonococcal outer membrane Protein I. J. Infect. Dis. **150:**44-48.
118. **Knapp, J. S., C. Thornsberry, G. K. Schoolnik, P. J. Wiesner, K. K. Holmes, and Cooperative Study Group.** 1978. Phenotypic and epidemiologic correlates of auxotype in *Neisseria gonorrhoeae.* J. Infect. Dis. **138:**160- 165.
119. **Koomey, M.** 1998. Competence for natural transformation in *Neisseria gonorrhoeae*: a model system for studies of horizontal gene transfer. APMIS **106(Suppl. 84):**56-61.
120. **Koomey, M.** 2001. Implications of molecular contacts and signaling initiated by *Neisseria gonorrhoeae*. Curr. Opin. Microbiol. **4:**53-57.
121. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:**150-163.
122. **Kupsch, E.-M., B. Knepper, T. Kuroki, I. Heuer, and T. F. Meyer.** 1993. Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by *Neisseria gonorrhoeae* for human leukocytes and epithelial cells. EMBO J. **12:**641-650.
123. **Lambden, P. R., and J. E. Heckels.** 1979. Outer membrane protein composition and colonial morphology of *Neisseria gonorrhoeae* strain P9. FEMS Microbiol. Lett. **5:**263-265.
124. **Lambden, P. R., J. E. Heckels, L. T. James, and P. J. Watt.** 1979. Variations in surface protein composition associated with virulence

properties in opacity types of *Neisseria gonorrhoeae*. J. Gen. Microbiol. **114:**305-312.

125. **Lammel, C. J., R. L. Sweet, P. A. Rice, J. S. Knapp, G. K. Schoolnik, D. C. Helibron, and G. F. Brooks.** 1985. Antibody-antigen specificity in the immune response to infection with *Neisseria gonorrhoeae*. J. Infect. Dis. **152:**990-1001.
126. **Linz, B., M. Schenker, P. Zhu, and M. Achtman.** 2000. Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. Mol. Microbiol. **36:**1049-1058.
127. **Liu, S. V., N. J. Saunders, A. Jeffries, and R. F. Rest.** 2002. Genome analysis and strain comparison of correia repeats and correia repeat-enclosed elements in pathogenic *Neisseria*. J. Bacteriol. **184:**6163-6173.
128. L**ong, C. D., R. N. Madraswala, and H. S. Seifert.** 1998. Comparisons between colony phase variation of *Neisseria gonorrhoeae* FA1090 and pilus, pilin, and S-pilin expression. Infect. Immun. **66:**1918-1927.
129. **Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 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.
130. **Makepeace, B. L., P. J. Watt, J. E. Heckels, and M. Christodoulides.** 2001. Interactions of *Neisseria gonorrhoeae* with mature human macrophage opacity proteins influence production of proinflammatory cytokines. Infect. Immun. **69:**1909-1913.
131. **Makino, S.-I., J. P. M. van Putten, and T. F. Meyer.** 1991. Phase variation of the opacity outer membrane protein controls invasion by *Neisseria gonorrhoeae* into human epithelial cells. EMBO J. **10:**1307-1315.
132. **Malorny, B., G. Morelli, B. Kusecek, J. Kolberg, and M. Achtman.** 1998. Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins. J. Bacteriol. **180:**1323-1330.
133. **Marri, P. R., M. Paniscus, N. J. Weyand, M. A. A. Rendón, C. M. Calton, D. R. Hernández, D. L. Higashi, E. Sodergren, G. M. Weinstock, S. D. Rounsley, and M. So.** 2010. Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. PLoS ONE **5:**doi:10.1371/journal.pone.0011835.
134. **Martin, I. M. C., A. Ghani, G. Bell, G. Kinghorn, and C. A. Ison.** 2003. Persistence of two genotypes of *Neisseria gonorrhoeae* during transmission. J. Clin. Microbiol. **41:**5609-5614.
135. **Martin, I. M. C., and C. A. Ison.** 2003. Detection of mixed infection of *Neisseria gonorrhoeae.* Sex. Transm. Infect. **79:**56-58.
136. **Martin, I. M. C., C. A. Ison, D. M. Aanensen, K. A. Fenton, and B. G. Spratt.** 2004. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. J. Infect. Dis. **189:**1497-1505.
137. **Massari, P., S. Ram, H. Macleod, and L. M. Wetzler.** 2003. The role of porins in neisserial pathogenesis and immunity. Trends Microbiol. **11:**87-93.
138. **Mayer, L. W.** 1982. Rates of *in vitro* changes of gonococcal colony opacity phenotypes. Infect. Immun. **37:**481-485.
139. **Maynard Smith, J., N. H. Smith, M. O'Rouke, and B. G. Spratt.** 1993. How clonal are bacteria. Proc. Natl. Acad. Sci. U.S.A. **90:**4384-4388.

140. **McKnew, D. L., F. Lynn, J. M. Zenilman, and M. C. Bash.** 2002. Porin variation among clinical isolates of *Neisseria gonorrhoeae* over a 10-year period, as determined by *por* variable region typing. J. Infect. Dis. **187:**1213-1222.
141. **Mehr, I. J., and H. S. Seifert.** 1998. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. Mol. Microbiol. **30:**697-710.
142. **Merz, A. J., and M. So.** 2000. Interactions of pathogenic *Neisseriae* with epithelial cell membranes. Annu. Rev. Cell Dev. Biol **16:**423-457.
143. **Mestecky, J., Z. Moldoveanu, and M. W. Russell.** 2005. Immunologic uniqueness of the genital tract: challenge for vaccine development. Am. J. Reprod. Immunol. **53:**208-214.
144. **Meyer, T. F., E. Billyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrheae*: chromosomal organization and DNA sequence. Proc. Natl. Acad. Sci. U.S.A **81:**6110-6114.
145. **Meyer, T. F., C. P. Gibbs, and R. Haas.** 1990. Variation and control of protein expression in *Neisseria*. Annu. Rev. Microbiol. **44:**451-477.
146. **Meyer, T. F., R. Haas, A. Stern, H. Fiedler, M. Frosch, F. Jähnig, K. Muralidharan, and S. Veit.** 1986. Surface proteins of pathogenic *Neisseriae*: antigenic variation and conserved epitopes. Ann. Sclavo Collana Monogr. **3:**407-414.
147. **Meyer, T. F., and J. P. M. Putten.** 1989. Genetic mechanisms and biological implications of phase variation in pathogenic *Neisseriae*. Clin. Microbiol. Rev. **2(Suppl.):**S139-S145.
148. **Moore, J., S. E. S. Bailey, Z. Benmechernene, C. Tzitzilonis, N. J. E. Griffiths, M. Virji, and J. P. Derrick.** 2005. Recognition of saccharides by the OpcA, OpaD, and OpaB outer membrane proteins from *Neisseria meningitidis*. J. Biol. Chem. **280:**31489-31497.
149. **Morelli, G., B. Malorny, K. Muller, A. Seiler, J.-F. Wang, J. del Valle, and M. Achtman.** 1997. Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. Mol. Microbiol. **25:**1047-1064.
150. **Morello, J. A., S. A. Lerner, and M. Bohnhoff.** 1976. Characteristics of atypical *Neisseria gonorrhoeae* from disseminated and localized infections. Infect. Immun. **13:**1510-1516.
151. **Morris, A. K., H. M. Palmer, and H. Young.** 2008. Opa-typing can identify epidemiologically distinct subgroups within *Neisseria gonorrhoeae* multi-antigen sequence type (NG-MAST) clusters. Epidemiol. Infect. **136:**417-420.
152. **Morse, S. A.** 1996. *Neisseria gonorrhoeae*. *In* S. Baron (ed.), Medical Microbiology 4th ed. University of Texas Medical Branch at Galveston, Galveston, Texas, USA.
153. **Muenzner, P., C. Dehio, T. Fujiwara, M. Achtman, T. F. Meyer, and S. D. Gray-Owen.** 2000. Carcinoembryonic antigen family receptor specificity of *Neisseria meningitidis* Opa variants influences adherence to and invasion of proinflammatory cytokine-activated endothelial cells. Infect. Immun. **68:**3601-3607.
154. **Murphy, G. L., T. D. Connell, D. S. Barritt, M. Koomey, and J. G. Cannon.** 1989. Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell **56:**539-547.

155. **Naids, F. L., B. Belisle, N. Lee, and R. F. Rest.** 1991. Interactions of *Neisseria gonorrhoeae* with human neutrophils: studies with purified PII (Opa) outer membrane proteins and synthetic Opa peptides. Infect. Immun. **59:**4628-4635.
156. **Nassif, X.** 2000. Gonococcal lipooligosaccharide: an adhesin for bacterial dissemination. Trends Microbiol. **8:**539-540.
157. **Newbold, C. I.** 1999. Antigenic variation in *Plasmodium falciparum*: mechanisms and consequences. Curr. Opin. Microbiol. **2:**420-425.
158. **Ng, P. P. L., R. K. W. Chan, and A. E. Ling.** 1998. Gonorrhoea treatment failure and ciprofloxacin resistance. Int. J. STD AIDS **9:**323-325.
159. **Nicoll, A., G. Hughes, M. Donnelly, S. Livingstone, D. De Angelis, K. Fenton, B. Evans, O. N. Gill, and M. Catchpole.** 2001. Assessing the impact of national anti-HIV sexual health campaigns: trends in the transmission of HIV and other sexually transmitted infections in England. Sex. Transm. Infect. **77:**242-247.
160. **O'Rourke, M., C. A. Ison, A. M. Renton, and B. G. Spratt.** 1995. Opa-typing: a high-resolution tool for studying the epidemiology of gonorrhoea. Mol. Microbiol. **17:**865-875.
161. **O'Rourke, M., and B. G. Spratt.** 1994. Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: extensive genetic diversity within isolates of the same electrophoretic type. Microbiology **140:**1285-1290.
162. **O'Rourke, M., and E. Stevens.** 1993. Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. J. Gen. Microbiol. **139:**2603-2611.
163. **Palmer, H. M., and C. Arnold.** 2001. Genotyping *Neisseria gonorrhoeae* using fluorescent amplified fragment length polymorphism analysis. J. Clin. Microbiol. **39:**2325-2329.
164. **Palmer, H. M., H. Mallinson, R. L. Wood, and A. J. Herring.** 2003. Evaluation of the specificities of five DNA amplification methods for the detection of *Neisseria gonorrhoeae*. J. Clin. Microbiol. **41:**835-837.
165. **Pantelic, M., Y.-J. Kim, S. Bolland, I. Chen, J. Shively, and T. Chen.** 2005. *Neisseria gonorrhoeae* kills carcinoembryonic antigen-related cellular adhesion molecule 1 (CD66a)-expressing human B cells and inhibits antibody production. Infect. Immun. **73:**4171-4179.
166. **Parsons, N. J., P. V. Patel, E. L. Tan, J. R. C. Andrade, C. A. Nairn, M. Goldner, J. A. Cole, and H. Smith.** 1988. Cytidine 5'-monophospho-N-acetyl neuraminic acid and a low molecular weight factor from human blood cells induce lipopolysaccharide alteration in gonococci when conferring resistance to killing by human serum. Microb. Pathog. **5:**303-309.
167. **Patrone, J. B., and D. C. Stein.** 2007. Effect of gonococcal lipooligosaccharide variation on human monocytic cytokine profile. BMC Microbiol. **7:**doi:10.1186/1471-2180-7-7.
168. **Perrin, A., S. Bonacorsi, E. Carbonnelle, D. Talibi, P. Dessen, X. Nassif, and C. Tinsley.** 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.
169. **Phillips, I.** 1976. Beta-lactamase-producing, penicillin-resistant gonococcus. Lancet **2:**656-657.

170. **Piot, P., and M. Q. Islam. 1994.** Sexually transmitted diseases in the 1990s. Global epidemiology and challenges for control. Sex. Transm. Dis. **21(Suppl.):**S7-S13.
171. **Poh, C. L., H. P. Khng, C. K. Lim, and G. K. Loh.** 1992. Molecular typing of *Neisseria gonorrhoeae* by restriction fragment length polymorphisms. Genitourin. Med. **68:**106-110.
172. **Poh, C. L., and Q. C. Lau.** 1993. Subtyping of *Neisseria gonorrhoeae* auxotype-serovar groups by pulsed-field gel electrophoresis. J. Med. Microbiol. **38:**366-370.
173. **The Gonococcal Resistance to Antimicrobials Surveillance Programme.** 2008. GRASP 2008 report: trends in antimicrobial resistant gonorrhoea. Health Protection Agency.
174. **Rasti, N., M. Wahlgren, and Q. Chen.** 2004. Molecular aspects of malaria pathogenesis. FEMS Immunol. Med. Microbiol. **41:**9-26.
175. **Risley, C. L., H. Ward, B. Choudhury, C. J. Bishop, K. A. Fenton, B. G. Spratt, C. A. Ison, and A. C. Ghani.** 2007. Geographical and demographic clustering of gonorrhoea in London. Sex. Transm. Infect. **83:**481-487.
176. **Roberts, M. C.** 1989. Plasmids of *Neisseria gonorrhoeae* and other *Neisseria* species. Clin. Microbiol. Rev. **2(Suppl.):**S18-S23.
177. **Ross, J. D. C.** 1996. Systemic gonococcal infection. Genitourin. Med. **72:**404-407.
178. **Rozen, S., and H. J. Skaletsky.** 1998. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
179. **Rudel, T., I. Scheuerpflug, and T. F. Meyer.** 1995. Neisseria PilC protein identified as type-4 pilus tip-located adhesin. Nature **373:**357-359.
180. **Russell, M. W., S. R. Hedges, H. Y. Wu, E. W. I. Hook, and J. Mestecky.** 1999. Mucosal immunity in the genital tract: prospects for vaccines against sexually transmitted diseases-a review. Am. J. Reprod. Immunol. **42:**58-63.
181. **Russell, M. W., and J. Mestecky.** 2002. Humoral immune responses to microbial infections in the genital tract. Microbes Infect. **4:**667-677.
182. **Sarafian, S. K., and J. S. Knapp.** 1989. Molecular epidemiology of gonorrhoea. Clin. Microbiol. Rev. **2(Suppl.):**S49- S55.
183. **Sarkari, J., N. Pandit, E. R. Moxon, and M. Achtman.** 1994. Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. Mol. Microbiol. **13:**207-217.
184. **Scherf, A., J. J. Lopez-Rubio, and L. C. Riviere.** 2008. Antigenic variation in *Plasmodium falciparum*. Annu. Rev. Microbiol. **62:**445-470.
185. **Schmidt, K. A., C. D. Deal, M. Kwan, E. Thattassery, and H. Schneider.** 2000. *Neisseria gonorrhoeae* MS11mkC opacity protein expression *in vitro* and during human volunteer infectivity studies. Sex. Transm. Dis. **27:**278-283.
186. **Schmidt, K. A. P., H. P. Schneider, J. A. M. Lindstrom, J. W. M. Boslego, R. A. P. Warren, L. P. Van De Verg, C. D. P. Deal, J. B. M. McClain, and J. M. M. D. Griffiss.** 2001. Experimental gonococcal urethritis and reinfection with homologous gonococci in male volunteers. Sex. Transm. Dis. **28:**555-564.
187. **Schwalbe, R. S., and J. G. Cannon.** 1986. Genetic transformation of genes for Protein II in *Neisseria gonorrhoeae*. J. Bacteriol. **167:**186-190.

188. **Sebastian, S., S. Agarwal, J. R. Murphy, and C. A. Genco.** 2002. The gonococcal Fur regulon: identification of additional genes involved in major catabolic, recombination, and secretory pathways. J. Bacteriol. **184:**3965-3974.
189. **Segal, E., E. Billyard, M. So, S. Storzbach, and T. F. Meyer.** 1985. Role of chromosomal rearrangement in *N. gonorrhoeae* pilus phase variation. Cell **40:**293-300.
190. **Seifert, H. S.** 1996. Questions about gonococcal pilus phase- and antigenic variation. Mol. Microbiol. **21:**433-440.
191. **Seifert, H. S., C. J. Wright, A. E. Jerse, M. S. Cohen, and J. G. Cannon.** 1994. Multiple gonococcal pilin antigenic variants are produced during experimental human infections. J. Clin. Invest. **93:**2744-2749.
192. **Selander, R. K., and J. M. Musser.** 1990. Population genetics of bacterial pathogenesis, p. 11-36. *In* B. H. Iglewski and V. L. Clark (ed.), Molecular Basis of Bacterial Infections. Academic Press, San Diego.
193. **Simon, D., and R. F. Rest.** 1992. *Escherichia coli* expressing a *Neisseria gonorrhoeae* opacity-associated outer membrane protein invade human cervical and endometrial epithelial cell lines. Proc. Natl. Acad. Sci. U.S.A. **89:**5512-5516.
194. **Simons, M. P., W. M. Nauseef, T. S. Griffith, and M. A. Apicella.** 2006. *Neisseria gonorrhoeae* delays the onset of apoptosis in polymorphonuclear leukocytes. Cell. Microbiol. **8:**1780-1790.
195. **Slevogt, H., S. Zabel, B. Opitz, A. Hocke, J. Eitel, P. D. N'Guessan, L. Lucka, K. Riesbeck, W. Zimmermann, J. Zweigner, B. Temmesfeld-Wollbrueck, N. Suttorp, and B. B. Singer.** 2008. CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. **9:**1270-1278.
196. **Smargiasso, N., V. Gabelica, C. Damblon, F. Rosu, E. De Pauw, M.-P. Teulade-Fichou, J. Rowe, and A. Claessens.** 2009. Putative DNA G-quadruplex formation within the promoters of *Plasmodium falciparum var* genes. BMC Genomics **10:**362.
197. **Smith, H., N. J. Parsons, and J. A. Cole.** 1995. Sialylation of neisserial lipopolysaccharide: a major influence on pathogenicity. Microb. Pathog. **19:**365-377.
198. **Smith, J. D., C. E. Chitnis, A. G. Craig, D. J. Roberts, D. E. Hudson-Taylor, D. S. Peterson, R. Pinches, C. I. Newbold, and L. H. Miller.** 1995. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell **82:**101-110.
199. **Snyder, L. A. S., S. A. Butcher, and N. J. Saunders.** 2001. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp. Microbiology **147:**2321-2332.
200. S**nyder, L. A. S., J. K. Davies, C. S. Ryan, and N. J. Saunders.** 2005. Comparative overview of the genomic and genetic differences between the pathogenic *Neisseria* strains and species. Plasmid **54:**191-218.
201. **Snyder, L. A. S., and N. J. Saunders.** 2006. The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as 'virulence genes'. BMC Microbiol. **7:**doi:10.1186/1471-2164-7-128.

202. **Sparling, F. P., J. G. Cannon, and M. So.** 1986. Phase and antigenic variation of pili and outer membrane Protein II of *Neisseria gonorrhoeae*. J. Infect. Dis. **153:**196-201.
203. **Spratt, B. G. 1999.** Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. Curr. Opin. Microbiol. **2:**312-316.
204. **Spratt, B. G., and M. C. J. Maiden.** 1999. Bacterial population genetics, evolution and epidemiology. Phil. Trans. R. Soc. Lond. B, Biol. Sci. **354:**701-710.
205. **Spratt, B. G., N. H. Smith, J. Zhou, M. O'Rourke, and E. Feil.** 1995. The Population genetics of pathogenic *Neisseria*, p. 143-160. *In* S. Baumberg, J. P. W. Young, E. M. H. Wellington, and J. R. Saunders (ed.), The population genetics of bacteria. Presented at the SGM Symposium No. 52.
206. S**tern, A., M. Brown, P. Nickel, and T. F. Meyer.** 1986. Opacity genes in *Neisseria gonorrhoeae*: control of phase variation and antigenic variation. Cell **47:**61-71.
207. **Stern, A., and T. F. Meyer.** 1987. Common mechanisms controlling phase and antigenic variation in pathogenic *Neisseriae*. Mol. Microbiol. **1:**5-12.
208. **Stern, A., P. Nickel, T. F. Meyer, and M. So.** 1984. Opacity determinants of *Neisseria gonorrhoeae*: gene expression and chromosomal linkage to the gonococcal pilus gene. Cell **37:**447-456.
209. **The UK Collaborative Group for HIV and STI Surveillance.** 2005. Mapping the issues. HIV and other sexually transmitted infections in the United Kingdom: 2005. Health Protection Agency. London, UK.
210. **Swanson, J.** 1980. ^{125}I-labelled peptide mapping of some heat-modifiable proteins of the gonococcal outer membrane. Infect. Immun. **28:**54-64.
211. **Swanson, J.** 1982. Colony opacity and Protein II compositions of gonococci. Infect. Immun. **37:**359-368.
212. **Swanson, J.** 1973. Studies on gonococcus infection IV. Pili: their role in attachment of gonococci to tissue culture cells. J. Exp. Med. **137:**571-589.
213. **Swanson, J.** 1978. Studies on gonococcus infection XII. Colony color and opacity variants of gonococci. Infect. Immun. **19:**320-331.
214. **Swanson, J.** 1978. Studies on gonococcus infection. XIV. Cell wall protein differences among color/opacity colony variants of *Neisseria gonorrhoeae*. Infect. Immun. **21:**292-302.
215. **Swanson, J., O. Barrera, J. Sola, and J. Boslego.** 1988. Expression of outer membrane protein II by gonococci in experimental gonorrhea. J. Exp. Med. **168:**2121-2129.
216. **Swanson, J., S. G. Morrison, O. Barrera, and S. A. Hill.** 1990. Piliation changes in transformation-defective gonococci. J. Exp. Med. **171**:2131-2139.
217. **Swanson, J., K. Robbins, O. Barrera, D. Corwin, J. Boslego, J. Ciak, M. Blake, and J. Koomey.** 1987. Gonococcal pilin variants in experimental gonorrhea. J. Exp. Med. **165:**1344-1357.
218. **Tam, M. R., T. M. Buchanan, E. G. Sandstrom, K. K. Holmes, J. S. Knapp, A. W. Siadak, and R. C. Nowinski.** 1982. Serological classification of *Neisseria gonorrhoeae* with monoclonal antibodies. Infect. Immun. **36:**1042-1053.
219. **Tapsall, J.** 1999. Antimicrobial resistance in *Neisseria gonorrhoeae*. World Health Organisation.

220. **Tettelin, H. e., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter.** 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287:**1809-1815.

221. **Thomas, J. C., and M. J. Tucker.** 1996. The development and use of the concept of a sexually transmitted disease core. J. Infect. Dis. **174(Suppl. 2):**S134-143.

222. **Thomson, C. J.** 1999. The global epidemiology of resistance to ciprofloxacin and the changing nature of antibiotic resistance: a 10 year perspective. J. Antimicrob. Chemother. **43:**31-40.

223. **Tinsley, C. R., and X. Nassif.** 1996. Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. Proc. Natl. Acad. Sci. U.S.A. **93:**11109-11114.

224. **Tobiason, D. M., and H. S. Seifert.** 2001. Inverse relationship between pilus-mediated gonococcal adherence and surface expression of the pilus receptor, CD46. Microbiology **147:**2333-2340.

225. **Tobiason, D. M., and H. S. Seifert.** 2006. The obligate human pathogen, *Neisseria gonorrhoeae*, is polyploid PLoS Biol. **4:**doi:10.1371/journal.pbio.0040185.

226. **Toleman, M., E. Aho, and M. Virji.** 2001. Expression of pathogen-like Opa adhesins in commensal *Neisseria*: genetic and functional analysis. Cell. Microbiol. **3:**33-44.

227. v**an der Ende, A., C. Hopman, S. Zaat, B. Essink, B. Berkhout, and J. Dankert.** 1995. Variable expression of Class I outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and - 35 regions of the promoter. J. Bacteriol. **177:**2475-2480.

228. **van der Ley, P.** 1988. Three copies of a single Protein II-encoding sequence in the genome of *Neisseria gonorrhoeae* JS3: evidence for gene conversion and gene duplication. Mol. Microbiol. **2:**797-806.

229. **Van Dyck, E., M. Ieven, S. Pattyn, L. Van Damme, and M. Laga.** 2001. Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by enzyme immunoassay, culture, and three nucleic acid amplification tests. J. Clin. Microbiol. **39:**1751-1756.

230. **Van Looveren, M., C. A. Ison, M. Ieven, P. Vandamme, I. M. Martin, K. Vermeulen, A. Renton, and H. Goossens.** 1999. Evaluation of the discriminatory power of typing methods for *Neisseria gonorrhoeae*. J. Clin. Microbiol. **37:**2183-2188.

231. **van Putten, J. P., and B. D. Robertson.** 1995. Molecular mechanisms and implications for infection of lipopolysaccharide variation in *Neisseria*. Mol. Microbiol. **16:**847-853.

232. **Vasquez, J., L. de la Fuente, S. Berrón, M. O'Rourke, N. H. Smith, J. Zhou, and B. G. Spratt.** 1993. Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. Curr. Biol. **3:**567-572.

233. **Vázquez, J. A., S. Berrón, M. O'Rouke, Carpenter, E. Feil, N. H. Smith, and B. G. Spratt.** 1995. Interspecies recombination in nature: a meningococcus that has acquired a gonococcal PIB porin. Mol. Microbiol. **15:**1001-1007.
234. **Virji, M., D. Evans, A. Hadfield, F. Grunert, A. M. Teixeira, and S. M. Watt.** 1999. Critical determinants of host receptor targeting by *Neisseria meningitidis* and *Neisseria gonorrhoeae*: identification of Opa adhesiotopes on the N-domain of CD66 molecules. Mol. Microbiol. **34:**538-551.
235. **Virji, M., and J. E. Heckels.** 1986. The Effect of Protein II and pili on the interaction of *Neisseria gonorrhoeae* with human polymorphonuclear leucocytes. J. Gen. Microbiol. **132:**503-512.
236. **Virji, M., K. Makepeace, D. J. P. Ferguson, and S. M. Watt.** 1996. Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic *Neisseriae*. Mol. Microbiol. **22:**941-950.
237. **Virji, M., S. M. Watt, S. Barker, K. Makepeace, and R. Doyonnas.** 1996. The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. Mol. Microbiol. **22:**929-939.
238. **Viscidi, R. P., and J. C. Demma.** 2003. Genetic diversity of *Neisseria gonorrhoeae* housekeeping genes. J. Clin. Microbiol. **41:**197-204.
239. **Viscidi, R. P., J. C. Demma, J. Gu, and J. M. Zenilman.** 2000. Comparison of sequencing of the *por* gene and typing of the *opa* gene for discrimination of *Neisseria gonorrhoeae* strains from sexual contacts. J. Clin. Microbiol. **38:**4430-4438.
240. **Ward, H., C. A. Ison, S. E. Day, I. Martin, A. C. Ghani, G. P. Garnett, G. Bell, G. Kinghorn, and J. N. Weber.** 2000. A prospective social and molecular investigation of gonococcal transmission. Lancet **356:**1812-1817.
241. **Wetzler, L. M., K. Barry, M. S. Blake, and E. C. Gotschlich.** 1992. Gonococcal lipooligosaccharide sialylation prevents complement-dependent killing by immune sera. Infect. Immun. **60:**39-43.
242. **Whiley, D. M., J. W. Tapsall, and T. P. Sloots.** 2006. Nucleic acid amplification testing for *Neisseria gonorrhoeae*: an ongoing challenge. J. Mol. Diagn. **8:**3-15.
243. **White, P. J., H. Ward, J. A. Cassell, C. H. Mercer, and G. P. Garnett.** 2005. Vicious and virtuous circles in the dynamics of infectious disease and the provision of health care: gonorrhea in Britain as an example. J. Infect. Dis. **192:**824-836.
244. **Wiertz, E., A. Delvig, E. Donders, H. Brugghe, L. van Unen, H. Timmermans, M. Achtman, P. Hoogerhout, and J. Poolman.** 1996. T-cell responses to outer membrane proteins of *Neisseria meningitidis*: comparative study of the Opa, Opc, and PorA proteins. Infect. Immun. **64:**298-304.
245. **Zak, K., J. L. Diaz, D. Jackson, and J. E. Heckels.** 1984. Antigenic variation during infection with *Neisseria gonorrhoeae*: detection of antibodies to surface proteins in sera of patients with gonorrhoea. J. Infect. Dis. **149:**166-174.
246. **Zhang, Q., D. DeRyckere, P. Lauer, and M. Koomey.** 1992. Gene conversion in *Neisseria gonorrhoeae*: evidence for its role in pilus antigenic variation. Proc. Natl. Acad. Sci. U.S.A. **89:**5366-5370.

APPENDIX 1

Primer pairs used to amplify each *opa* locus and annealing temperatures

***opa* locus**	**Primer name and sequence**	**Annealing temperature**	**Extension time**
opa1	Opa1up 5'-CTGAAATACTTTACCTTCCTGTCCAT-3'		
	Opa1dn 5'-TTGTGAACAAAGAAAGTCATCG-3'	60°C	3 min 20 sec
opa2	Opa2up 5'-GCCGATGACTTTCTTTGTTCACA-3'		
	Opa2dn 5'-GTGTGCCGAATAAACCAGTGTC-3'	61°C	3 min
opa3	Opa3up 5'-CAGGCACACTTCTTCATCTTGA-3'		
	Opa3dn 5'-ATTTCTTGGATGTTGGGCATAC-3'	62°C	2 min 40 sec
opa4	Opa4up 5'-ATCACGGAAGCCACATCGTC-3'		
	Opa4dn 5'-CGGTGATTTTGTTTTGGAAGAA-3'	61°C	3 min
opa5	Opa5up 5'-CAAACAGTACGGCAACCACA-3'		
	Opa5dn 5'-AATGTCTTGGTTGGGATTGG-3'	62°C	1 min 20 sec
opa6	Opa6up 5'-GCCCATATAGCCAAGAATGGTA-3'		
	Opa6dn 5'-CGAGCAGTTCGACTATCTCGAC-3'	61°C	5 min 30 sec
opa7	Opa7up 5'-TCCTCTGTTTTGAAACCCTGAC-3'		
	Opa7dn 5'-ATCAAAGCCATCATCCATCC-3'	62°C	3 min
opa8	Opa8up 5'-AAGTGTCCGACAAACACCATC-3'		
	Opa8dn 5'-TTTGTTACGATGTCCGCTTTC-3'	61°C	2 min 10 sec
opa9	Opa9up 5'-GTTCAACGCCGAAGAGAAAA-3'		
	Opa9dn 5'-TTTGCAGGAACTCAAACTCG-3'	61°C	3 min
opa10	Opa10up 5'-TCCCAGCCAGTTAAACAGCTT-3'		
	Opa10dn 5'-CACCTATCCCGACTGCAACTAC-3'	64°C	1 min 45 sec
opa11	Opa11up 5'-AAGCCTTTGACGACGCATT-3'		
	Opa11dn 5'-CATTTCACGAGCTGAACATCA-3'	62°C	2 min 25 sec

APPENDIX 2

The following manuscripts are presented in their published form:

Bilek, N., I. M. Martin, G. Bell, G. R. Kinghorn, C. A. Ison, and B. G. Spratt. 2007. Concordance between *Neisseria gonorrhoeae* genotypes recovered from known sexual contacts. J. Clin. Microbiol. **45:**3564-3567.

Bilek, N., C. A. Ison, and B. G. Spratt. 2009. Relative contributions of recombination and mutation to the diversification of the *opa* gene repertoire of *Neisseria gonorrhoeae*. J. Bacteriol. **191:**1878-1890.

JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2007, p. 3564–3567 Vol. 45, No. 11
0095-1137/07/$08.00+0 doi:10.1128/JCM.01453-07
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Concordance between *Neisseria gonorrhoeae* Genotypes Recovered from Known Sexual Contacts[▽]

Nicole Bilek,[1] Iona M. Martin,[2] Gill Bell,[3] George R. Kinghorn,[3] Catherine A. Ison,[2] and Brian G. Spratt[1]*

Department of Infectious Disease Epidemiology, Imperial College London, St. Mary's Hospital, Norfolk Place, London W2 1PG,[1] Sexually Transmitted Bacteria Reference Laboratory, Health Protection Agency Centre for Infections, Health Protection Agency, 61 Colindale Avenue, London NW9 5HT,[2] and Department of Genitourinary Medicine, Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF,[3] United Kingdom

Received 19 July 2007/Returned for modification 29 August 2007/Accepted 4 September 2007

***Neisseria gonorrhoeae* multiantigen sequence typing (NG-MAST) is a highly discriminatory molecular typing procedure that provides precise and unambiguous strain characterization. Since molecular typing can complement contact tracing for reconstructing gonorrhea sexual networks, the concordance between the NG-MAST genotypes of pairs of *N. gonorrhoeae* isolates from recent sexual contacts was examined. Among 72 pairs of gonococci from recent sexual contacts, the genotypes of each pair were concordant in 65 cases (90.3%). In two further pairs, the isolates from sexual contacts differed by only a single nonsynonymous substitution in the porin gene, and in both of these pairs, the isolates were the same by *opa* typing. The other five nonconcordant pairs of isolates were clearly different strains. *opa* typing data were available for 51 of the pairs of isolates from sexual contacts, and concordant *opa* types were obtained in 38 cases (74.5%). NG-MAST should therefore be better than *opa* typing at identifying recent sexual contacts and has the important advantage over *opa* typing of being a more precise method of strain characterization.**

Typing of *Neisseria gonorrhoeae* isolates has provided a valuable adjunct to contact tracing for reconstructing sexual networks (3, 14) and for identifying individuals predicted to be in the same sexual network (2, 6). Phenotypic typing methods for gonococci—for example, the combination of auxotype and serovar—lack sufficient discrimination for this purpose, but a number of more-discriminatory molecular methods have been developed. Of these, pulsed-field gel electrophoresis and *opa* typing are considered to be the most highly discriminatory but have the disadvantage that they require the comparison of complex DNA fragment patterns on agarose or acrylamide gels (10, 12). More recently there has been a move toward the use of digital data for molecular typing, since this is more suitable for the development of molecular-typing Internet databases and the unambiguous comparison of new isolates with those already deposited in the database. In gonococci, the sequences of the porin (*por*) gene and fragments of both *por* and the transferrin-binding protein subunit B gene (*tbpB*) have been evaluated for molecular typing (9, 11, 13). These genes show high levels of sequence diversity among strains; they are believed to evolve rapidly, because their gene products are surface exposed and are targets of the host immune response to infection. The combination of the sequences at *por* and *tbpB* have been used to develop *N. gonorrhoeae* multiantigen sequence typing (NG-MAST). In this procedure each unique sequence at each locus is given a different allele number, so that a strain is defined unambiguously by two digits, corresponding to the allele numbers at *por* and *tbpB*. Each different two-digit number is assigned to a distinct sequence type (ST) that is used to describe the strain (9).

Rapid diversification of the loci used to characterize gonococci provides very high levels of discrimination between strains, but too high rates could lead to slight differences in the genotypes of isolates recovered from recent sexual contacts. We therefore used NG-MAST to examine the concordance between the genotypes of isolates of *N. gonorrhoeae* recovered from recent sexual contacts, in order to assess its use in supporting traditional contact-tracing methods, and compared the concordance with that obtained using *opa* typing.

MATERIALS AND METHODS

Gonococcal isolates were recovered from patients attending the genitourinary clinic at the Royal Hallamshire Hospital in Sheffield, United Kingdom, between March 1995 and December 1998. An additional set of isolates came from patients attending the same clinic between May 1999 and April 2000. The great majority of patients in Sheffield with gonorrhea are diagnosed and managed by this clinic, and most of those diagnosed elsewhere are referred to the clinic for management. Patients are interviewed at their first visit to the clinic, and behavioral data and information on sexual contacts during the last 3 months are recorded. Patients are encouraged to ask their sexual contacts to visit the clinic (and if necessary, clinic staff) and, with permission from the index case patient, to inform the contacts of their potential exposure to gonorrhea. Components of the sexual network were reconstructed using the contact-tracing data, and the recorded dates at which specimens were obtained from named sexual contacts were used to identify pairs of gonococcal isolates from recent, mutually named sexual contacts.

N. gonorrhoeae was grown on single-strength Difco GC medium base (Becton Dickinson, Oxford, United Kingdom) supplemented with 1% Vitox (Oxoid, Basingstoke, United Kingdom) and incubated at 37°C under 5% CO_2. Chromosomal DNA was prepared and NG-MAST carried out as described by Martin et al. (9). Previously identified alleles and STs were obtained by interrogating the NG-MAST website (www.ng-mast.net); previously unrecognized alleles and allele combinations were submitted to the website in order to obtain the new allele and ST numbers. *opa* typing was performed as described by O'Rourke et al. (10)

* Corresponding author. Mailing address: Department of Infectious Disease Epidemiology, Imperial College London, St. Mary's Hospital, Norfolk Place, London W2 1PG, United Kingdom. Phone: 0044 (0) 20 7594 3398. Fax: 0044 (0) 20 7402 3927. E-mail: b.spratt@imperial.ac.uk.

▽ Published ahead of print on 12 September 2007.

using TaqI. Isolates were considered to have the same *opa* type if the DNA fragment patterns were indistinguishable. Genotypic diversity was measured using Simpson's index of diversity (5) with its 95% confidence limit (4).

RESULTS

Identification of recent sexual contacts. Exhaustive contact tracing of individuals with gonorrhea in Sheffield allowed the reconstruction of components of the sexual network. These data were used to identify named sexual contacts where the *N. gonorrhoeae* isolates were available for molecular typing. We then evaluated the concordance between the genotypes of the gonococcal isolates recovered from pairs of individuals in Sheffield who were most clearly identified as recent sexual contacts. Two criteria were used to identify recent sexual contacts. Each of the contacts must have named the other as a sexual contact (mutual naming), and the gonococcal isolates from the contacts must have been from specimens taken at the clinic within 1 month of each other. The average time between recovery of isolates from all 72 mutually named sexual contacts was 6.1 days (range, 0 to 29 days), and 72.2% of the pairs of isolates were recovered within a week of each other.

Analysis of gonococci from recent sexual contacts. Between March 1995 and December 1998, there were 56 pairs of sexual contacts who met the above criteria, and 51 of the 56 pairs of gonococcal isolates from these contacts (91.1%) were indistinguishable by NG-MAST. Two of the five nonconcordant pairs had the same *tpbB* allele and differed at only a single nonsynonymous nucleotide site in *por*, which altered an amino acid in either loop 5 or loop 7 of the Por protein. Both of these pairs had indistinguishable *opa* types. For 51 of the 56 pairs of isolates from sexual contacts, there were *opa*-typing data; 38 of 51 pairs (74.5%) had indistinguishable *opa* types.

During this period, two major strains were identified in Sheffield using *opa* typing (7). All 167 isolates recovered between December 1995 and November 1996 were characterized by NG-MAST and confirmed the presence of the two major strains, which corresponded to ST12 and ST261 (16.2% and 26.3% of the isolates, respectively). Several of the pairs of sexual contacts were infected with ST12 or ST261, and it is possible that for some sexual contacts, infection with these prevalent strains had not been passed from one contact to the other, but the same strain had been independently acquired from additional sexual contacts. We therefore excluded pairs of sexual contacts who were both infected with either ST12 or ST261 and reexamined the proportion of sexual contacts infected with the same strain. Of the remaining 32 pairs of sexual contacts, 27 (84.4%) were infected with the same strain. Two of the nonconcordant pairs in this subset were those that differed at a single nonsynonymous nucleotide site, and if these pairs are both considered to be infections with the same strain, the concordance increases from 84.4% to 90.6%.

Isolates were also available from contact tracing carried out in Sheffield between May 1999 and April 2000. In this period, ST12 and ST261 were absent or rare (only 1 of the 119 isolates recovered during this period was ST12, and none was ST261), and the gonococcal population was significantly more diverse than in the earlier sampling period. Simpson's index of diversity was 0.90 (confidence interval, 0.86, 0.93) for the earlier period and 0.96 (confidence interval, 0.95, 0.98) for the later period. Of the 16 pairs of isolates from mutually named recent sexual contacts (according to the criteria described above) recovered in this later period, 14 were concordant in genotype by NG-MAST. If the data from these two periods are combined, there was concordance between the NG-MAST genotypes for 65/72 (90.3%) pairs of isolates from recent sexual contacts (or 93.1% concordance if isolates with single nonsynonymous substitutions in *por* are considered to be the same strain).

Analysis of the sexual-network components that include nonconcordant strains. Excluding the two pairs of isolates that differed at a single site within *por*, five pairs of isolates were nonconcordant. Four pairs were infected with completely different strains, having multiple differences in both *tbpB* and *por*. The fifth pair had the same *tbpB* allele and three nonsynonymous differences in *por*, changing amino acids in loops 5 and 6. The isolates in the latter pair had different *opa* types and were also considered to be completely different strains.

The components of the sexual network that included these five nonconcordant pairs were examined. Four of the pairs were from two sexual-network components. One involved a male (S109) infected with ST224 who named four female contacts, all within a month (Fig. 1A). Isolates were available from three of these contacts, and all were ST261. The different isolates from the male and two of the female contacts contributed two of the five nonconcordant pairs of strains. The isolate from the third female contact (S107) was not included, because there was not mutual naming. Isolates from two additional male contacts of this female were both also ST261. It is assumed that during the 1-month period in which these isolates were recovered, the male at the center of the network (S109) had been infected with ST224 and also ST261, the former infection possibly obtained from the additional contact (isolate not available) or additional undisclosed contacts.

The other component that contributed two nonconcordant pairs (Fig. 1B) included six identified individuals. Isolates were available from four, and all were ST250, except for the isolate from a female sexual contact (NS036). The isolates from this female and her two recent male contacts represented the two nonconcordant pairs. The fifth nonconcordant pair was from a small-network component (Fig. 1C) of one male who reported three female sexual contacts; one female was infected with the same strain as the male (ST51), one with a different strain (ST276), and no isolate was available from the third. Although we consider ST51 and ST276 to be different strains, they do share the same *tbpB* allele. However, the *por* alleles differ at three nucleotides, spread over a 135-bp region, indicating that for these isolates to be minor variants of the same strain, either three independent mutations in *por* or a single recombinational replacement would have had to occur between sexual contacts. We therefore prefer the cautious view that they are different strains.

DISCUSSION

For individuals who become infected with gonorrhea and infect a regular sexual contact, with no additional sexual contacts, there should usually be concordance between the genotypes of the isolates from the two individuals. Lack of concordance could occur either if there was incorrect or incomplete

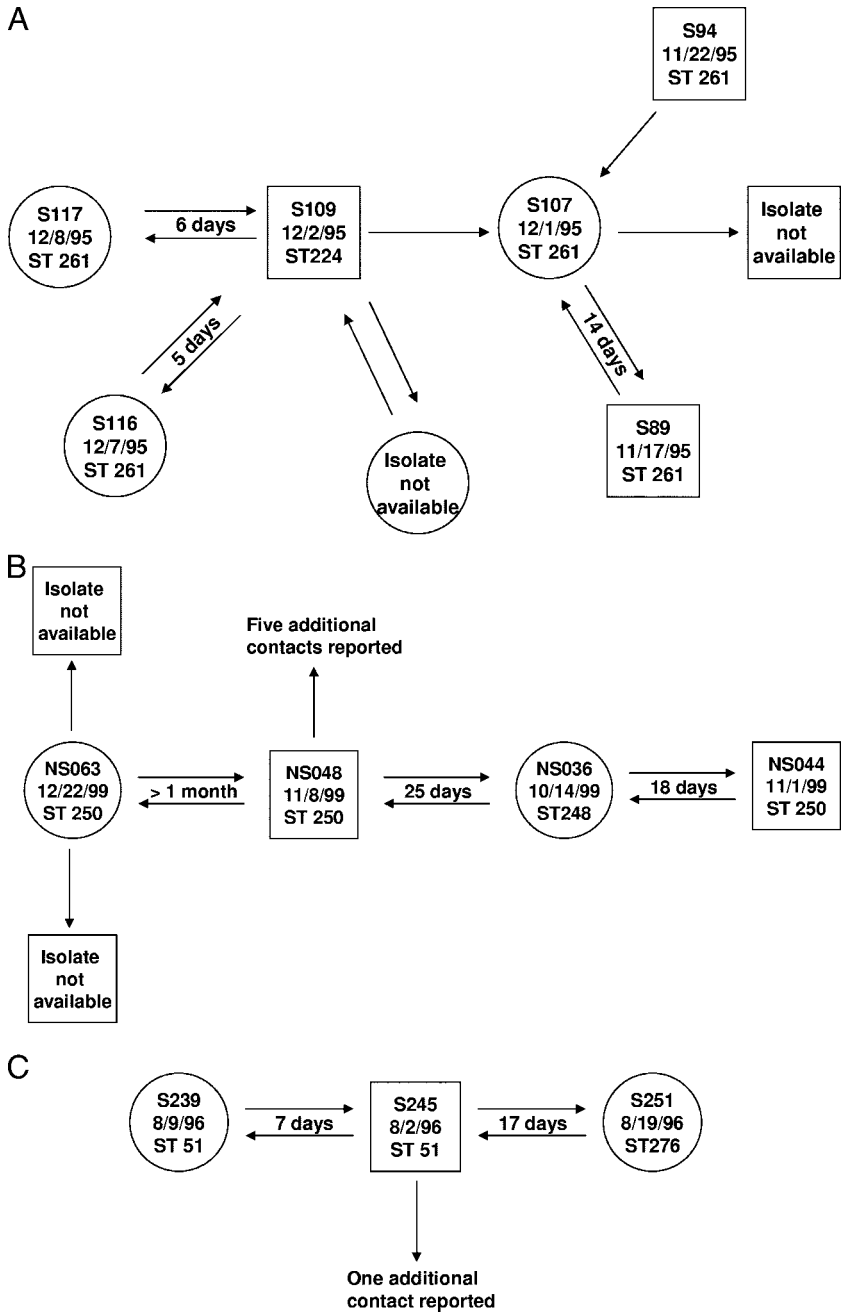


FIG. 1. Sexual network components including recent contacts with nonconcordant gonococcal genotypes. Panels A, B, and C show data for three network components. Circles (females) or squares (males) represent identified sexual contacts. Where an isolate was available, the information within the circle or square shows the patient number, the date of the clinic visit, and the ST. Additional contacts that are not in a circle or square were reported but unidentified. Each arrow indicates that an individual identified the other as a contact. Two arrows between individuals denotes mutual naming. The period between the visits of the contacts to the clinic are indicated.

reporting of sexual contacts, if genetic variation indexed by the typing procedure evolved extremely rapidly, or if the initially infected individual had a mixed infection. In many cases, the situation is more complex, and individuals with gonorrhea may have had multiple recent sexual contacts. In such cases, different isolates may be recovered from mutually named recent sexual contacts as a result of additional infections from other sexual contacts.

Constructing gonorrhea sexual networks is difficult, because contact tracing is time-consuming and usually incomplete (1), and molecular typing can be combined with contact tracing to build a more complete picture of the sexual network (3, 14). A major requirement for using molecular typing for this purpose is that the typing method should be highly discriminatory, so that many different strains can be shown to be circulating in the community, and yet isolates from recent sexual contacts should

usually be indistinguishable. NG-MAST is one of the most convenient and discriminatory of the available molecular-typing methods for *N. gonorrhoeae*, but except for an analysis of 10 pairs of isolates from sexual contacts that were preselected as being indistinguishable by phenotypic and molecular methods (9), as far as we are aware, there is no information on the degree of concordance between the genotypes of isolates from known sexual contacts obtained using this method.

In this study, the degree of concordance determined by using NG-MAST and sexual-contact tracing data was 90.3 to 93.1%; the higher value included the two pairs of isolates that differed at a single nonsynonymous site within *por* as the same strain. Concordance was higher for NG-MAST than for *opa* typing, presumably because variation in the *opa* gene repertoire occurs more rapidly than variation in *por* and *tbpB*. Two examples were found of changes in an ST between sexual contacts due to single nonsynonymous nucleotide substitutions in *por*, which had presumably occurred in one of the sexual contacts as a result of selection imposed by the host immune response, consistent with the fact that in both cases these substitutions alter an amino acid within a predicted surface-exposed loop in the porin protein.

The other examples of nonconcordance were due to completely different strains in sexual contacts, presumably due to additional infections from other sexual contacts or to mixed infection. Under such circumstances, nonconcordance is inevitable with any molecular typing method. Previous studies have shown that mixed gonococcal infections occur not infrequently but are difficult to detect after culturing, presumably due to overgrowth of one of the strains (8). Nonconcordance has been observed using other molecular-typing procedures. For example, 2/17 pairs of isolates from sexual contacts in Baltimore were nonconcordant by *opa* typing, and 1 of these 17 pairs was nonconcordant by using *por* sequences (13). The degree of concordance with NG-MAST in this larger study was therefore similar to that observed for these other molecular-typing procedures. NG-MAST therefore provides a molecular-typing procedure that is well suited as an adjunct to contact tracing for reconstructing sexual networks and for this purpose appears to be slightly better than *opa* typing, since concordance was higher. Perhaps more importantly, NG-MAST is easier to perform than *opa* typing, and comparisons between isolates are much simpler, because sequence data are precise and unambiguous.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust. B.G.S. is a Wellcome Trust Principal Research Fellow.

We thank Paul Zadik for help with the study.

REFERENCES

1. **Bell, G., H. Ward, S. Day, A. C. Ghani, U. Goan, E. Claydon, and G. R. Kinghorn.** 1998. Partner notification for gonorrhoea: a comparative study with a provincial and a metropolitan UK clinic. Sex. Transm. Infect. **74:**409–414.
2. **Choudhury, B., C. L. Risley, A. C. Ghani, C. J. Bishop, H. Ward, K. A. Fenton, C. A. Ison, and B. G. Spratt.** 2006. Identification of individuals with gonorrhoea within sexual networks: a population-based study. Lancet **368:** 139–146.
3. **Ghani, A. C., C. A. Ison, H. Ward, G. P. Garnett, G. Bell, G. R. Kinghorn, J. Weber, and S. Day.** 1996. Sexual contact networks in the transmission of sexually transmitted diseases. An analysis of gonorrhea cases in Sheffield, UK. Sex. Transm. Dis. **23:**498–503.
4. **Grundmann, H., S. Hori, and G. Tanner.** 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J. Clin. Microbiol. **39:**4190–4192.
5. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465–2466.
6. **Kolader, M. E., N. H. Dukers, A. K. van der Bij, M. Dierdorp, J. S. Fennema, R. A. Coutinho, and S. M. Bruisten.** 2006. Molecular epidemiology of *Neisseria gonorrhoeae* in Amsterdam, The Netherlands, shows distinct heterosexual and homosexual networks. J. Clin. Microbiol. **44:**2689–2697.
7. **Martin, I. M., A. Ghani, G. Bell, G. Kinghorn, and C. A. Ison.** 2003. Persistence of two genotypes of *Neisseria gonorrhoeae* during transmission. J. Clin. Microbiol. **41:**5609–5614.
8. **Martin, I. M., and C. A. Ison.** 2003. Detection of mixed infection of *Neisseria gonorrhoeae*. Sex. Transm. Infect. **79:**56–58.
9. **Martin, I. M., C. A. Ison, D. M. Aanensen, K. A. Fenton, and B. G. Spratt.** 2004. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. J. Infect. Dis. **189:**1497–1505.
10. **O'Rourke, M., C. A. Ison, A. M. Renton, and B. G. Spratt.** 1995. Opa-typing: a high-resolution tool for studying the epidemiology of gonorrhoea. Mol. Microbiol. **17:**865–875.
11. **Unemo, M., P. Olcén, J. Albert, and H. Fredlund.** 2003. Comparison of serologic and genetic *porB*-based typing of *Neisseria gonorrhoeae*: consequences for future characterization. J. Clin. Microbiol. **41:**4141–4147.
12. **van Looveren, M., C. A. Ison, M. Ieven, P. Vandamme, I. M. Martin, K. Vermeulen, A. Renton, and H. Goossens.** 1999. Evaluation of the discriminatory power of typing methods for *Neisseria gonorrhoeae*. J. Clin. Microbiol. **37:**2183–2188.
13. **Viscidi, R. P., J. C. Demma, J. Gu, and J. Zenilman.** 2000. Comparison of sequencing of the *por* gene and typing of the *opa* gene for discrimination of *Neisseria gonorrhoeae* strains from sexual contacts. J. Clin. Microbiol. **38:** 4430–4438.
14. **Ward, H., C. A. Ison, S. E. Day, I. Martin, A. C. Ghani, G. P. Garnett, G. Bell, G. Kinghorn, and J. N. Weber.** 2000. A prospective social and molecular investigation of gonococcal transmission. Lancet **356:**1812–1817.

JOURNAL OF BACTERIOLOGY, Mar. 2009, p. 1878–1890 Vol. 191, No. 6
0021-9193/09/$08.00+0 doi:10.1128/JB.01518-08
Copyright © 2009, American Society for Microbiology. All Rights Reserved.

Relative Contributions of Recombination and Mutation to the Diversification of the *opa* Gene Repertoire of *Neisseria gonorrhoeae*[▽]

Nicole Bilek,[1] Catherine A. Ison,[2] and Brian G. Spratt[1]*

Department of Infectious Disease Epidemiology, Imperial College London, St. Mary's Hospital, Norfolk Place, London W2 1PG, United Kingdom,[1] and Sexually Transmitted Bacteria Reference Laboratory, Centre for Infections, Health Protection Agency, Colindale, London NW9 5HT, United Kingdom[2]

Received 27 October 2008/Accepted 16 December 2008

To understand the rates and mechanisms of *Neisseria gonorrhoeae opa* gene variation, the 11 *opa* genes were amplified independently so that an *opa* allelic profile could be defined for any isolate from the sequences at each locus. The *opa* allelic profiles from 14 unrelated isolates were all different, with no *opa* alleles shared between isolates. Examination of very closely related isolates from sexual contacts and sexual networks showed that these typically shared most *opa* alleles, and the mechanisms by which recent changes occurred at individual *opa* loci could be determined. The great majority of changes were due to recombination among existing alleles that duplicated an *opa* allele present at another locus or resulted in a mosaic of existing *opa* alleles. Single nucleotide changes or insertion/deletion of a single codon also occurred, but few of these events were assigned to mutation, the majority being assigned to localized recombination. Introduction of novel *opa* genes from coinfecting strains was rare, and all but one were observed in the same sexual network. Changes at *opa* loci occurred at a greater rate than those at the porin locus, and the *opa11* locus changed more rapidly than other *opa* loci, almost always differing even between recent sexual contacts. Examination of the neighboring *pilE* gene showed that changes at *opa11* and *pilE* often occurred together, although this linkage may not be a causal one.

The Opa proteins of human pathogenic *Neisseria* species are adhesins that play an important role in the colonization of genital and nasopharyngeal surfaces and in subsequent invasion of the mucosal epithelium. They can be classed into two broad categories based on the human receptors they bind to, namely, the carcinoembryonic antigen cell adhesion molecules and heparan sulfate proteoglycans (17). Opa proteins differ in the ability to interact with their receptors and in the ability to promote invasion of human cells (7, 22, 35). Human challenge studies have shown that Opa expression is important in the early stages of establishing a gonococcal infection and that the Opa proteins expressed may vary during an infection (33). In human volunteers infected with the same strain of *Neisseria gonorrhoeae* (20), different Opa proteins are expressed in different individuals, suggesting that there are multiple Opa proteins within a gonococcal strain that can promote successful interactions with the genital mucosa.

Isolates of *Neisseria meningitidis* have three or four *opa* genes (1), while most commensal species have only one gene, with the exception of *Neisseria lactamica*, which has been reported to have two (34). In *N. gonorrhoeae*, which is believed to have arisen relatively recently as a lineage of one of the above nasopharyngeal neisseriae that can colonize and be transmitted from the genital tract, there appears to have been an expansion of the *opa* gene family and 11 *opa* genes have been identified (4, 5). All *Neisseria* Opa proteins have a similar structural framework, with highly conserved regions interspersed with two hypervariable (HV) regions, HV1 and HV2, which are believed to be surface exposed and to be the sites of interaction with host receptors (8, 14). The HV regions are also the targets of the host immune response to infection (24). Two additional surface-exposed loops are present, one being semivariable and the other largely conserved (24).

In both gonococci and meningococci, expression of the Opa proteins is controlled posttranscriptionally by reversible phase variation (31, 32). Strand slippage and mispairing events are believed to occur during DNA replication within a series of tandem pentameric repeats (CTCTT), termed the coding repeats (CRs), present in the region of the gene encoding the signal peptide (2, 28, 31). Insertions or deletions of a pentamer repeat cause changes in the translational reading frame and determine which of the Opa proteins are presented on the cell surface. The expression of Opa proteins may not be random, as the rate of transition from *opa* nonexpression to expression appears to correlate with the strength of the different *opa* gene promoters (3).

Opa proteins are immunogenic (37), and due to the strong selection pressures imposed by the human immune system, the *opa* genes evolve rapidly and consequently are highly variable. Since the receptor-binding sites involve the surface-exposed HV regions, nonsynonymous (NS) substitutions in these regions, selected by the need to evade the host immune system, are presumably strongly filtered by the need to maintain receptor-binding activity. The mechanisms by which changes in *opa* genes occur have been studied in *N. meningitidis* by ana-

* Corresponding author. Mailing address: Department of Infectious Disease Epidemiology, St. Mary's Hospital, Imperial College London, Norfolk Place, London W2 1PG, United Kingdom. Phone: 44 20 7594 3398. Fax: 44 20 7402 3927. E-mail: b.spratt@imperial.ac.uk.

▽ Published ahead of print on 29 December 2008.

lyzing isolates belonging to the same clonal complex, defined by multilocus enzyme electrophoresis (18, 19) or multilocus sequence typing (MLST; 10), and include NS mutations within HV regions, recombination events that include *opa* gene duplication and the generation of mosaic alleles, and import of novel *opa* sequences from other strains.

While *opa* gene diversity and the general mechanisms involved in *opa* gene variation have been described in meningococci, there is little information on the rates of change at *opa* loci or the relative contributions of the different mechanisms of *opa* variation. The diversity of the larger gonococcal *opa* gene family has been much less well characterized and has been restricted to analysis of the *opa* sequences from strain MS11 (4) and the HV regions from FA1090 (12) and analysis of additional strains with a set of HV1 and HV2 DNA hybridization probes (9).

In this work, the alleles present at each of the 11 *opa* loci have been determined for sets of closely related isolates of the same *N. gonorrhoeae* multiantigen sequence typing sequence type (NG-MAST ST; 26) from known recent sexual contacts in Sheffield and from small sexual networks in London. We show that most changes at gonococcal *opa* loci involve the duplication of an existing allele or the creation of a mosaic between two existing alleles, with few new alleles being generated by point mutation or importation from coinfecting strains. We also show that changes at one *opa* gene occur much more rapidly than those at the others and are often linked to changes at the flanking pilin expression (*pilE*) locus.

MATERIALS AND METHODS

Strains and isolates. Methods were developed with a panel of 14 strains from our collection of gonococci that have been characterized by NG-MAST, selecting those that were distinct STs and which had neither the *porB* nor the *tbpB* alleles in common.

Isolates from cases of gonorrhea in Sheffield between April 1995 and March 1997 were from the study of Martin et al. (25, 36). Additional isolates from Sheffield collected up to November 1997 were also used. Many of these isolates have been characterized by both *opa* typing (25, 36) and NG-MAST (6), and detailed sexual contact tracing data are available. Isolates from mutually named recent sexual contacts were from this collection.

Isolates from sexual networks in London were from the study of Choudhury et al. (11, 30), which characterized the majority of gonococcal isolates recovered in London over a 6-month period. Networks were identified as groups of individuals sharing the same strain (by NG-MAST). Networks composed of fewer than 10 individuals, where all isolates were recovered within a period of 3 months, were chosen. Four were composed exclusively of heterosexuals, three consisted almost exclusively of men who have sex with men (MSM; two unknowns), and one was mainly heterosexual but included one bisexual and one MSM.

Bacterial culture and DNA preparation. To prepare DNA, isolates were removed from storage at −80°C, plated onto single-strength Difco GC medium base (Becton Dickinson, Oxford, United Kingdom) supplemented with 1% Vitox (Oxoid, Basingstoke, United Kingdom), and incubated overnight at 37°C under 5% CO_2–95% air. Single colonies were picked and subcultured, and DNA lysates were made from them by boiling the resuspended bacterial growth in phosphate-buffered saline.

Identification of *opa* genes in the genome of strain FA1090. The complete genome sequence of *N. gonorrhoeae* strain FA1090 (GenBank accession number AE0049969) is available at http://www.genome.ou.edu/gono.html but has to date not been fully annotated. The position of each *opa* gene in the genome sequence was located by identifying highly conserved sequences which are unique to *opa* genes, such as the variable-number CRs (CTCTT) in the region encoding the signal sequence. The genes are dispersed around the chromosome, the distance between any two neighboring *opa* genes ranging between ~5 kb and ~850 kb.

Amplification and sequencing of each independent *opa* gene. The genome sequence data were used to identify the genes flanking each *opa* locus. Flanking genes that were common to *N. gonorrhoeae* and *N. meningitidis* were used, and a comparison of these sequences was used to identify the most conserved regions in each of the selected flanking genes, from which primers could be designed to amplify each *opa* gene independently by PCR. All *opa* loci were amplified with the Extensor Hi-Fidelity PCR Master Mix, Buffer 1, Reddymix Version (Abgene, Epsom, United Kingdom), according to the manufacturer's instructions with the following primer pairs and annealing temperatures: *opa1*, Opa1up (5′-CTGAA ATACTTTACCTTCCTGTCCAT-3′) and Opa1dn (5′-TTGTGAACAAAGAA AGTCATCG-3′), annealed at 60°C; *opa2*, Opa2up (5′-GCCGATGACTTTCT TTGTTCACA-3′) and Opa2dn (5′-GTGTGCCGAATAAACCAGTGTC-3′), annealed at 61°C; *opa3*, Opa3up (5′-CAGGCACACTTCTTCATCTTGA-3′) and Opa3dn (5′-ATTTCTTGGATGTTGGGCATAC-3′), annealed at 62°C; *opa4*, Opa4up (5′-ATCACGGAAGCCACATCGTC-3′) and Opa4dn (5′-CGG TGATTTTGTTTTGGAAGAA-3′), annealed at 61°C; *opa5*, Opa5up (5′-CAA ACAGTACGGCAACCACA-3′) and Opa5dn (5′-AATGTCTTGGTTGGGAT TGG-3′), annealed at 62°C; *opa6*, Opa6up (5′-GCCCATATAGCCAAGAATG GTA-3′) and Opa6dn (5′-CGAGCAGTTCGACTATCTCGAC-3′), annealed at 61°C; *opa7*, Opa7up (5′-TCCTCTGTTTTGAAACCCTGAC-3′) and Opa7dn (5′-ATCAAAGCCATCATCCATCC-3′), annealed at 62°C; *opa8*, Opa8up (5′-AAGTGTCCGACAAACACCATC-3′) and Opa8dn (5′-TTTGTTACGATGTC CGCTTTC-3′), annealed at 61°C; *opa9*, Opa9up (5′-GTTCAACGCCGAAGA GAAAA-3′) and Opa9dn (5′-TTTGCAGGAACTCAAACTCG-3′), annealed at 61°C; *opa10*, Opa10up (5′-TCCCAGCCAGTTAAACAGCTT-3′) and Opa10dn (5′-CACCTATCCCGACTGCAACTAC-3′), annealed at 64°C; *opa11*, Opa11up (5′-AAGCCTTTGACGACGCATT-3′) and Opa11dn (5′-CATTTCACGAGCT GAACATCA-3′), annealed at 62°C.

The size of each PCR product varied, depending on the locus, ranging between approximately 2.1 and 7.5 kb. A 2-μl volume of each product was electrophoresed with a quantitative size standard (Lambda DNA/HindIII marker 2; Fermentas Life Sciences, Burlington, Canada) in a 0.8% agarose gel containing SafeView Nucleic Acid Stain (according to the manufacturer's instructions [NBS Biologicals, Huntingdon, Cambridgeshire, United Kingdom]) at 15 V/cm for 40 min in 1× Tris-borate-EDTA buffer and visualized under UV light.

The remainder of each PCR product was cleaned of excess primers and nucleotides with equal volumes of SureClean (Bioline, Randolph, MA) according to the manufacturer's instructions and resuspended in a volume of sterile water to achieve a final concentration of approximately 50 ng/μl.

Each *opa* gene was then sequenced with BigDye terminator Cycle Sequencing kit version 1.1 (Applied Biosystems). Each reaction mixture contained 2 μl of PCR product, 1 μl of primer (1 pmol/μl), and 2 μl of Big Dye and was cycled as follows: initial denaturation for 3 min at 96°C, followed by 30 cycles of 30 s at 96°C, 10 s at 50°C, and 2 min at 60°C; the reaction mixture was then cooled to 4°C at a rate of 0.1°C/s. The primers used to sequence each *opa* gene included those designed for *opa* typing (29) {opa-up [5′-GCGATTATTTCAGAAACAT CCG-3′] and opa-dn [5′-GCTTCGTGGGTTTTGAAGCG-3′] and two internal primers, int1 [5′-GATTATGCCCGTTACAG-3′] and either int2-GAT or int2-TGC [5′-CC(GAT/TGC)ATAGGGTTTGAA-3′]} designed from central conserved regions that sequence outward toward the start and end of the genes. Sequence off the former primer determined which of the two alternative second primers was used. Two additional primers that were the reverse complement of the *opa* typing primers were used to obtain sequence from the start and end of each *opa* gene and its immediate flanking regions (Ropa-up [5′-CGGATGTTT CTGAAATAATCGC-3′] and Ropa-dn [5′-CGCTTCAAAACCCACGAAGC-3′]). This combination of primers provides sequence data off both strands over almost the entire length of the *opa* gene, with only short regions at either end of *opa* that were only sequenced on one strand but with two different primers.

All products were then cleaned of excess primers, BigDye, and salts and separated on an ABI 3700 DNA analyzer (Applied Biosystems). All sequence data were viewed, edited, trimmed, and assembled with MEGA version 3.1 (21). Identical sequences were identified with the nonredundant database tool at http://linux.mlst.net/nrdb/nrdb.htm. Alignments were also done with MEGA.

Assigning *opa* changes to recombination or mutation. Change of one *opa* allele to another can occur by recombination or point mutation. Several classes of recombinational event can be distinguished, the simplest being a change of an allele to one present at another locus in the same isolate, resulting in allele duplication. Recombination can also result in the replacement of only part of the allele with the corresponding region of an allele present at another locus in the same isolate, resulting in a simple mosaic allele. In other cases, more complex mosaics were observed, where the new allele was composed of parts of two or more alleles present in the same isolate but which would have required more than one localized recombinational event. The above examples of recombinational changes could occur by gene conversion within a single bacterial cell or, following uptake of *opa* sequences, via natural transformation from sibling cells (these alternatives cannot be distinguished) and are all classified as recombina-

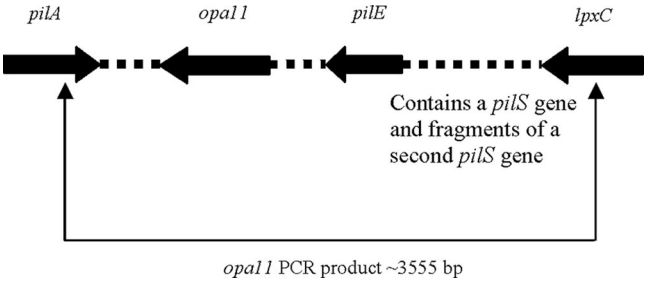


FIG. 1. Region of chromosome amplified for sequencing of the *opa11* gene. In *N. gonorrhoeae* strain FA1090, the 3.5-kb amplified fragment (vertical arrows) includes the *pilE* expression locus and a silent pilin gene, *pilS*. The *pilA* and *lpxC* genes encode a FtsY homolog and a deacetylase involved in lipid A biosynthesis, respectively.

tion among existing alleles. In contrast, a new allele is a sequence that is not already present at another *opa* locus in the isolate and cannot be formed as a mosaic of the existing alleles in the isolate (or by a simple point mutation) and therefore presumably represents the introduction of a novel *opa* allele from a coinfecting strain.

Point mutations in *opa* genes were recognized as alleles that differed from the consensus allele at a single nucleotide site or in some cases by a single codon insertion or deletion. Where these substitutions were unique among the *opa* genes of a set of isolates, rather than a polymorphism present in other alleles, they were assigned as the result of de novo mutation. All of these unique mutations were NS substitutions within HV1 or HV2. Where the single nucleotide changes were not unique, these included both synonymous and NS changes in conserved regions, excepting one within an HV region. As these latter nucleotide changes were present in other *opa* alleles within a set of closely related isolates (i.e., they were polymorphisms), they could have arisen by highly localized recombination that replaced a region of a consensus allele with a small region that differed only at the single nucleotide site, rather than by point mutation. As the appearance of unique point mutations was rare compared to changes by recombination (see Results), we considered that most of the changes at sites that were polymorphic were more likely to be due to localized recombination rather than point mutation and assigned them as such.

Sequencing of the *pilE* locus. The PCR fragment from strain FA1090 that contains the *opa11* gene also includes the upstream pilin expression locus, *pilE*, one silent pilin gene, *pilS*, and a partial pilin gene sequence composed of interrupted portions of pilin genes (16) (Fig. 1). The *pilS* genes lack the 5′ conserved region required for expression (15), and a sequencing primer could therefore be designed in this 5′ region of the *pilE* gene (pilE1, 5′-ATCGAGCTGATGATT GT-3′) to sequence (off one strand) the variable cassettes of the expression locus (but not those of *pilS*). The *pilE* sequences were trimmed to begin with the conserved sequence TTGGCGGCA and to end after the last nucleotide of the stop codon, with allele numbers assigned to each unique sequence. Frameshift mutations occur frequently in *pilE* and generate premature stop codons, and sequences were trimmed to the stop codon identified based on alignments of *pilE* genes that were full length and in frame, rather than to any premature stop codons generated by frameshifts. When sequence could not be obtained off the pilE1 primer, an alternative primer was used, based on a conserved sequence at the 3′ end of both the *pilE* and *pilS* genes (pilE6, 5′-AGCACCTGCCGTCAA C-3′). In the case of strain FA1090, this primer yields a mixed sequence as priming occurs from both *pilE* and *pilS*. However, in isolates where the pilE1 primer did not give sequence, the pilE6 primer usually gave unambiguous sequence rather than mixed sequence. This latter combination suggests that in these isolates a *pilE* gene is absent at this position and that the unambiguous sequence is from a *pilS* gene. This view is supported by the smaller size of the PCR product of these isolates (data not shown). The absence of *pilE* was confirmed by sequencing through the region from *opa11* to the end of the PCR product with the Ropa-up primer and the opa11dn PCR primer described above.

MLST. The internal fragments of the seven loci used in the *Neisseria* MLST scheme (23) were amplified by PCR, with the primers used for *N. meningitidis* being replaced with those with the slightly different sequences found in *N. gonorrhoeae* FA1090. These gonococcus-optimized primers were also used for the sequencing reactions.

RESULTS

Amplification and sequencing of the 11 gonococcal *opa* genes. The complete genome sequence of *N. gonorrhoeae* strain FA1090 is available although not fully annotated. The 11 *opa* genes were located and numbered 1 to 11 in the order they appear around the chromosome, beginning at *dnaA*. The flanking sequences were analyzed to identify conserved genes within which PCR primers could be designed to allow the amplification of each *opa* gene independently. This approach allows amplification of the corresponding *opa* loci in other isolates, even if their order on the chromosome has changed due to large-scale rearrangements.

The number of CRs sometimes varied in different DNA preparations of the same isolate (data not shown) and also on subculture (see below). To keep these events from complicating the assignment of alleles, the *opa* sequences were trimmed (in the +1 reading frame) after the CRs at the start of the highly conserved sequence GCAGC and after the last nucleotide of the stop codon, resulting in trimmed sequences between 696 bp and 765 bp in length. In a very small number of alleles, the first G in this conserved sequence was absent and the trimmed sequence began instead with a T. Trimmed sequences were compared, and a different allele number was assigned to each unique sequence, irrespective of the *opa* locus, which allows the identification of identical *opa* alleles at different loci. We defined the *opa* allelic profile of an isolate by a series of 11 integers corresponding to the alleles present at each locus, whereas the *opa* gene repertoire refers to the collection of *opa* alleles irrespective of the locus.

***opa* allelic profiles from a set of diverse gonococcal isolates.** To assess diversity among unrelated gonococci, the *opa* allelic profiles were determined for 14 diverse strains, including FA1090. Each strain had a different ST by NG-MAST (26), and none of the alleles at *porB* or *tbpB* were shared between strains. The *opa* allelic profiles of these strains were all different, and no *opa* alleles were shared between strains (Table 1). In 10/14 (71%) strains, at least one *opa* allele was present at more than one locus; two strains had the same *opa* allele at three loci, and another had the same *opa* allele at four loci.

In strain S58, the *opa9* locus was truncated approximately 250 bp downstream of the initiation codon. In another two strains, S242 and LG0617, the PCR products generated for *opa11* and *opa2*, respectively, were considerably smaller than expected. Since all of the *opa* sequencing primers failed to give sequence from these PCR products, it is most probable that no *opa* gene was present at these loci.

***opa* allelic profiles of isolates with the same NG-MAST ST.** The lack of any shared *opa* alleles in the 14 diverse strains suggests a great diversity of *opa* sequences in the gonococcal population and rapid *opa* gene variation. To understand the relative contributions of different mechanisms in generating *opa* gene diversity, and the rates of change of *opa* loci, we examined gonococci that are extremely closely related. Such strains have very recent shared ancestry and are expected to have identical or similar *opa* allelic profiles, and any *opa* changes that are observed will be recent and can be classified as due to point mutation or recombinational events.

Two sets of closely related isolates that had been characterized by *opa* typing (29) were available from a study of gonor-

TABLE 1. *opa* allelic profiles of 14 diverse strains of *N. gonorrhoeae*

Isolate	ST	Allele at locus[a]												
		por	*tbpb*	*opa1*	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
FA1090	773	514	219	8	15	24	27	35	43	52	59	24	73	15
LG0540	610	39	2	204	151	201	196	158	145	193	169	181	176	169
S58	12	8	5	1	10	16	25	29	36	45	54	†	67	29
S89	261	197	11	2	11	17	17	30	37	46	96	62	68	123
LG1930	825	558	21	199	148	199	195	156	142	190	142	179	173	142
LG0159	584	7	27	207	154	171	198	164	147	183	163	184	178	231
L6	302	238	29	7	13	23	13	13	42	51	58	65	71	13
LG0923	709	479	171	205	153	202	197	160	146	194	188	182	177	225
S242	232	98	50	4	12	19	267	32	39	48	56	290	70	#
LG0906	876	59	166	168	150	200	150	157	144	192	185	180	175	168
G181	662	451	192	255	258	262	266	266	275	280	284	289	294	300
LG0617	1063	474	281	257	#	264	270	274	257	282	286	292	296	302
LG1461	1131	759	291	256	259	263	269	273	276	281	285	291	295	301
T0679/05	2707	1605	632	265	260	265	271	265	277	283	288	293	297	303

[a] Colored alleles indicate the same allele at mutiple *loci*. #, PCR product smaller than expected and all primers designed from *opa* genes fail to give sequence. †, 5′ region of *opa* gene present but gene truncated.

rhea in Sheffield between April 1995 and March 1997 (25, 36), along with additional *opa*-typed isolates from the same study collected up to November 1997. Two prevalent strains were circulating during this 32-month period, previously identified by *opa* typing as *opa* type 1 and *opa* type 4, together making up 41% of the isolates studied (25). Subsequent work on these isolates has shown that *opa* type 1 and *opa* type 4 almost always correspond to ST12 and ST261, respectively, by NG-MAST (6). The consensus *opa* allelic profiles for these two strains were determined by examining the profiles of 21 ST12/*opa* type 1 isolates and 27 ST261/*opa* type 4 isolates recovered over the sampling period (Table 2). Although these two strains were cocirculating in Sheffield, no alleles were shared between any isolates of the two different STs.

Excepting *opa11*, 57 to 100% and 63 to 100% of the alleles at individual loci were conserved in the ST12 and ST261 isolates, allowing consensus alleles at each locus, and a consensus *opa* allelic profile, to be identified. At *opa11*, almost every isolate had a different allele, some of which corresponded to alleles present at other loci, suggesting that variation at this locus is far more rapid than at the other loci. Although a consensus *opa* allelic profile could be identified for both STs, 48% of ST12/*opa* type 1 and 81% of ST261/*opa* type 4 isolates departed from the consensus profile in at least one locus other than *opa11*. As found with the diverse strains, identical *opa* alleles were observed at more than one locus in both ST12 and ST261 isolates. For example, allele 17 was present at both *opa3* and *opa4* in most ST261 isolates and in one isolate was present also at *opa8*.

Assuming that each unique difference from the consensus profile (excluding *opa11*, where no consensus was identified) is generated only once, eight changes to the ST12/*opa* type 1 consensus profile were identified. Three changes were due to a consensus allele at one locus being duplicated at another, whereas the other five generated a new allele. Excluding *opa11*, 19 changes to the ST261/*opa* type 4 consensus profile were observed; only 1 appeared to be due to the duplication of a consensus allele, the others being new alleles.

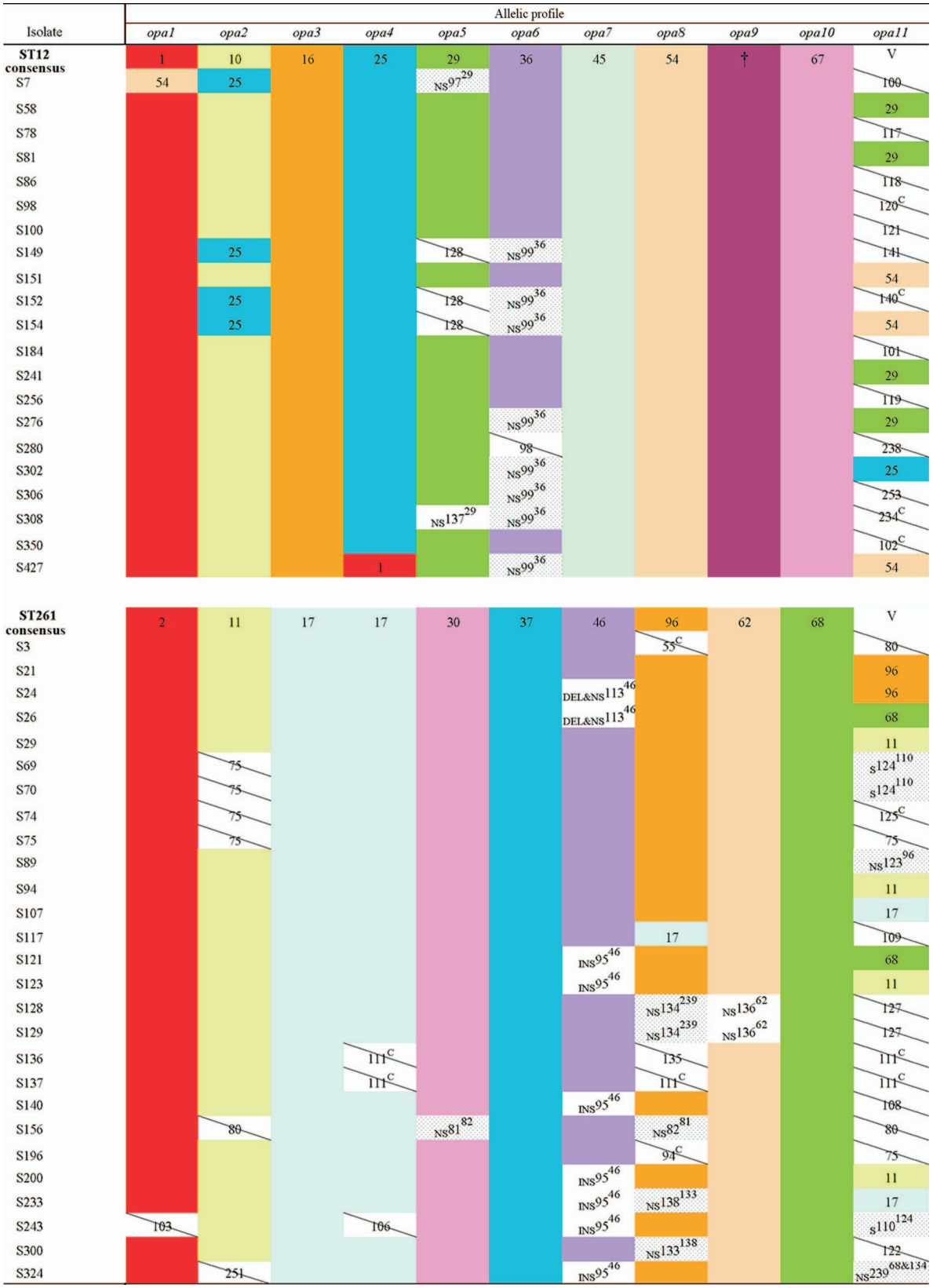
The differences in *opa* allelic profiles of isolates within the ST12/*opa* type 1 and ST261/*opa* type 4 clusters were unexpected, as the fragment patterns obtained by *opa* typing had previously been shown to be indistinguishable. In silico *opa* typing was carried out (data not shown) by using the 11 *opa* sequences to determine the fragment patterns expected to be generated by *opa* typing of the 21 ST12/*opa* type 1 and 27 ST261/*opa* type 4 isolates. In most cases, the observed changes in the *opa* allelic profile did not alter the fragment patterns obtained by in silico *opa* typing. The few observed differences in the in silico patterns resulted in changes in fragment size of only a few nucleotides that would not have been detectable by *opa* typing.

***opa* allelic profiles of isolates from known sexual contacts in Sheffield.** Considerable *opa* variation was found among the closely related isolates with the same ST/*opa* type circulating in the same city over a 2-year period. Isolates that were even more closely related were therefore examined by focusing on all of the ST12 and ST261 isolates from Sheffield that were obtained within 1 month of each other from mutually named sexual contacts. In this case, the selection of strains was blind to any available *opa* typing results to avoid selection bias for or against differences in their *opa* repertoires. Three triplets and 18 pairs of isolates (45 isolates in total) that fulfilled these criteria were identified and characterized. Table 3 shows the *opa* profiles of these isolates in relation to the consensus *opa* profile determined for ST12 and ST261.

Despite the use of isolates from recent sexual contacts, 14 of the 18 pairs, as well as all three triplets, differed at *opa11*, supporting the previous observation that the rate of variation at this locus is much higher than at other loci. At the other 10 *opa* loci, several of the pairs and triplets differed, as expected, from the ST12 or ST261 consensus *opa* allelic profile, but a difference in the profiles between isolates from recent sexual contacts was only seen in one triplet and three pairs (shown within a box in Table 3). The mechanisms by which these changes in *opa* allelic profile occurred are discussed in a later section.

***opa* allelic profiles of isolates obtained from individuals in sexual networks in London.** The large ST12 and ST261 sexual networks in Sheffield were heterosexual, and relatively few strains were circulating within this city during the 32-month sampling period. To include both MSM and heterosexual networks from an epidemiological setting in which many more strains are circulating, we analyzed gonococcal isolates from London. In a previous study, most isolates from cases of gonorrhea in London between June and November 2004 were analyzed by NG-MAST, and groups of individuals who shared

TABLE 2. *opa* allelic profiles of ST12/*opa* type 1 and ST261/*opa* type 4 isolates[a]



[a] Details are as in Table 1. Colored alleles indicate the presence of a consensus allele at a locus or the same allele at another locus, uncolored alleles are nonconsensus alleles. The colors for ST12 and ST261 isolates do not correspond to the same alleles. V, variable. For alleles differing at a single site, the allele number is shown followed by a superscript that denotes the allele from which it differs by a single synonymous (S) or NS substitution or an insertion (INS) or deletion (DEL), all of which were 3 bp (one codon). Stippling indicates that the difference from the allele shown by the superscript could be the result of a recombination event or a point mutation. A diagonal line through the allele number indicates that the allele is a simple hybrid/mosaic of *opa* alleles present within the strain. A superscript C denotes a more complex mosaic which may involve multiple recombination events.

TABLE 3. *opa* allelic profiles of ST12 and ST261 isolates from recent sexual contacts[a]

Pair or triplet	Isolate	Allelic profile											
		opa1	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*	*pilE*
ST12 consensus		1	10	16	25	29	36	45	54	†	67	V	V
Triplet1	**S149**		25			128	NS99[36]					141	14
	S152		25			128	NS99[36]					140[C]	23
	S154		25			128	NS99[36]					54	14
Triplet 2	S306						NS99[36]					253	21
	S308					NS137[29]	NS99[36]					234[C]	26
	S317					NS129[29]	NS99[36]					233	25
Triplet 3	**S445**						130[C]					114	14
	S463						130[C]					115[C]	22
	S466						130[C]					67	40
Pair 1	S78											117	27
	S86											118	28
Pair 2	S98											120[C]	14
	S100											121	30
Pair 3	S241											29	29
	S256											119	14
Pair 4	S253		25				132	NS112[45]				54	P
	S257		25				132	NS112[45]				25	14
Pair 5	S493						131[C]					116[C]	24
	S514						131[C]					116[C]	14
Pair 6	S517						NS99[36]					1	54
	S522						NS99[36]					25	P
ST261 consensus		2	11	17	17	30	37	46	96	62	68	V	V
Pair 7	S21											96	33
	S29											11	46
Pair 8	S24							DEL&NS113[46]				96	31
	S26							DEL&NS113[46]				68	32
Pair 9	S69		75									S124[110]	Abs
	S70		75									S124[110]	42
Pair 10	S74		75									125[C]	Abs
	S75		75									75	Abs
Pair 11	S89											NS123[96]	Abs
	S107											17	Abs
Pair 12	S121							INS95[46]				68	36
	S123							INS95[46]				11	Partial
Pair 13	S128								NS134[239]	NS136[62]		127	Abs
	S129								NS134[239]	NS136[62]		127	37
Pair 14	S136				111[C]				135			111[C]	Abs
	S137				111[C]				111[C]			111[C]	Abs
Pair 15	S156		80			NS81[82]			NS82[81]			80	Abs
	S162		80			NS81[82]			NS82[81]			83[C]	Abs
Pair 16	S192								94[C]			236	43
	S196								94[C]			75	Partial
Pair 17	S266								NS133[138]			103	34
	S271								139			252[C]	44
Pair 18	S254							INS95[46]	NS133[138]			108	Abs
	S258								NS133[138]			NS237[96]	45

[a] As in Tables 1 and 2, except that bold isolate numbers for the ST12 triplets indicate the isolate from the individual who named the other two individuals in the triplet as recent sexual contacts. Boxed alleles indicate difference in *opa* gene profiles (or *pilE* alleles) of isolates from recent sexual contacts. P, *pilE* present, but allele could not be assigned; see text. Abs, *pilE* absent at this location in the genome; see text. Partial, 5′ region of *pilE* absent; see text.

TABLE 4. *opa* allelic profiles of isolates from MSM in London recently infected with the same strain[a]

MSM network and sequence or isolate	Isolation date	Allelic profile										
		opa1	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST802 Consensus		203	149	167	13	13	143	49	167	63	174	167
LG0983	08/31/2004											
W0690	10/11/2004											
LG1497	10/18/2004											
W0735	10/19/2004											
LG1682	11/03/2004											
ST825 Consensus		199	148	199	195	156	142	190	166	179	173	V
LG1930	09/09/2004								142			142
W0494	09/16/2004											148
W0577	09/28/2004											165[C]
LG1304	09/30/2004											165[C]
LG1471	10/13/2004											166
LG1526	10/18/2004											166
LG1604	10/27/2004											165[C]
LG1775	11/11/2004		235									235
LG1846	11/25/2004								NS220[166]			148
ST876 Consensus		168	150	200	150	157	144	192	185	180	175	168
W0231	07/14/2004		210									
LG0893	07/20/2004											
LG0903	07/27/2004											221
LG0906	07/28/2004											

[a] As in Tables 1 to 3. Double-underlined allele numbers indicate alleles which are seen in another ST.

the same ST and had similar behavioral and demographic characteristics were considered to be within the same sexual network (11, 30). We examined the *opa* allelic profiles of isolates from individuals within three small heterosexual networks, three small MSM networks, and one of mixed sexual orientation. To focus on isolates that are likely to be the most closely related, we only used sexual networks in which all isolates were recovered within a period of 3 months. The *opa* allelic profiles of isolates from each of the London networks are shown in Tables 4 and 5.

As observed in other strains, the same consensus allele was present at more than one locus in one of the three heterosexual networks (Table 5) and all three MSM networks (Table 4). Atypically, there was no consensus allele at *opa8* in the ST610 heterosexual network. Furthermore, in two MSM networks (ST802 and ST876) and one heterosexual network (ST777), a consensus allele could be identified at *opa11*; the time frame in which these isolates were collected (a maximum of 3 months) was similar to or greater than that in which the isolates from sexual contacts in Sheffield were obtained (collected within a month of each other), which suggests that the apparent stability of *opa11* in these London networks is not due to a shorter time between recoveries of the isolates.

Isolates of the two dominant strains in Sheffield were not seen to share any *opa* alleles, and similarly, the strains in the three MSM networks in London did not share any alleles. However, alleles were shared between the heterosexual and mixed sexual orientation networks (Table 5). Five of the six shared alleles were among isolates of the ST777 and ST584 sexual networks. The consensus alleles at 9 of the 11 loci of ST777 are also consensus alleles in ST584, and in five cases, these shared alleles are at the same *opa* locus. Similarly, six consensus alleles in the ST584 network were in common with ST777. The substantial similarities in the *opa* profiles of ST584 and ST777 isolates are almost certainly due to recent shared ancestry, rather than horizontal spread of *opa* alleles. Common ancestry is not evident by NG-MAST, as the strains have diverged sufficiently to have different alleles at both *porB* and *tbpB*, but is strongly supported by the identical allelic profiles obtained by MLST (MLST ST6720), which indexes variation that accumulates more slowly than that used in NG-MAST.

The other shared allele (183) was found at *opa7* in one isolate of the ST610 network and at the same locus in all isolates of the ST584/777 networks. ST610 and ST584/777 differ at two loci by MLST (MLSTs ST6957 and ST6720, respectively), and the sharing of allele 183 could possibly have been due to relatively distant common ancestry rather than the recent introduction of this allele from another strain. The fact that this allele was only present in the most recently recovered isolate of ST610 (all but one of the earlier isolates have allele 193 at this locus) argues in favor of importation.

Although the isolates of the MSM network, ST802, had no alleles in common with any of the other London isolates studied here (neither MSM, heterosexual, nor mixed sexual orientation), examination of our database shows that they do share three alleles (present at the same loci) with strains of ST64

TABLE 5. *opa* allelic profiles of isolates from individuals in London recently infected with the same strain[a]

Network and sequence or isolate (gender)[b]	Isolation date	Allelic profile										
		opa1	*opa2*	*opa3*	*opa4*	*opa5*	*opa6*	*opa7*	*opa8*	*opa9*	*opa10*	*opa11*
ST610[c]		204	151	201	196	158	145	193	V	181	176	V
Consensus												
LG0104 (M)	06/10/2004	191						**218**	**186**			222
W0182 (F)	07/07/2004		152[C]			159			187			170[C]
W0236 (F)	07/14/2004								s169[145]			s169[145]
LG0540 (M)	07/22/2004								s169[145]			s169[145]
LG1081 (M)	09/06/2004	**209**	s**211**[223]			159		183	187			s**223**[211]
ST709[c]		205	153	202	197	160	146	194	188	182	177	V
Consensus												
LG0643 (F)	07/29/2004											X
LG0718 (M)	08/09/2004											224
LG0923 (M)	08/26/2004											225
LG1067 (M)	08/26/2004		248[C]							247		188
LG1062 (F)	09/06/2004											242
LG1569 (M)	10/29/2004											NS240[188]
ST777[c]		206	147	171	198	161	147	183	154	183	178	171
Consensus												
W0287 (M)	07/29/2004		154		215							
LG0954 (F)	08/26/2004											
LG1082 (F)	09/08/2004		213				s216[147]					
W0457 (M)	09/10/2004											s227[171]
W0656 (M)	10/05/2004											228[C]
ST584[d]		207	154	171	198	164	147	183	163	184	178	V
Consensus												
LG0161 (M)	06/15/2004											230[C]
LG0159 (F)	06/15/2004											231[C]
LG0201 (M)	06/21/2004					162			189[C]			229
LG0214 (M)	06/21/2004											231[C]
LG0456 (F)	07/13/2004		NS155[154]	NS214[171]					243			241
LG0592 (M)	07/29/2004					162			189[C]			172[C]
LG0594 (M)	07/29/2004					162			189[C]			172[C]
LG0788 (M)	08/16/2004											154

[a] As in Tables 1 to 4. Double-underlined allele numbers indicate alleles which are seen in another ST. Red bold allele numbers contain sequence from an unknown source, presumably as a result of a mixed infection. X, an allele number could not be assigned due to a short stretch of sequence in the 5′ region which could not be determined.

[b] M, male; F, female.

[c] Heterosexual.

[d] Mixed sexual orientation.

circulating among MSM in Sheffield and Lincoln (data not shown). The strains from MSM in London, Sheffield, and Lincoln were identical by MLST (MLST ST1580) and had the same *tbpB* allele, and their sharing of *opa* alleles was almost certainly due to the recent common ancestry of these strains rather than horizontal spread of the alleles.

Assigning allelic change in *opa* genes to recombination or point mutation. Isolates that are extremely closely related would typically have the same allele at each of the *opa* loci, and in those isolates where the allele at a single locus has changed, we can be confident that this has occurred by a single molecular event and can attempt to determine the nature of this event. Changes away from a consensus allele could be due to one of a number of mechanisms that can be broadly divided into those involving recombination and those due to point mutation. The criteria used to assign the mechanisms of allelic changes are given in Materials and Methods.

To evaluate the contribution of mutation and recombination to changes in the *opa* profile, we focused on the isolates from recent sexual contacts in Sheffield and the sexual networks in London to ensure that most allelic changes were due to a single molecular event rather than multiple sequential changes. The problem of multiple sequential changes is particularly problematic with the *opa11* locus, within which changes occurred much more rapidly than at the other *opa* loci. No attempt was therefore made to understand the mechanisms by which allelic changes at *opa11* occurred, except in the sexual networks where a consensus allele could be identified at *opa11*. Isolates of ST12 and ST261 shown in Table 2 were not included, as these were selected for being the same by both NG-MAST and

opa typing, and their inclusion could underestimate the contribution of recombination, which would be more likely to produce changes in *opa* type than point mutations.

The mechanisms of allelic change among isolates from sexual contacts in Sheffield were first assigned. The changes in triplet 2 (ST12 isolates) occurred at *opa5*, where one isolate had the consensus allele 29 and the other two each had a different allele (Table 3). Alleles 129 and 137 both differed from the consensus allele by a single NS substitution in HV2. However, the substitution in allele 137 was unique and was thus assigned to a mutation, whereas that in 129 was also present in the allele at *opa6* (allele 99), and we assign this change to a small localized recombinational event rather than a point mutation.

The difference in pair 14 (ST261 isolates) occurred at *opa8*, and neither had the consensus allele. Allele 111 was also present at both *opa4* and *opa11* and is assumed to be the ancestral allele at *opa8* from which allele 135 arose. Allele 135 is a simple mosaic of allele 111 and allele 46 (the consensus allele at *opa7*), and the allelic change in pair 14 is therefore assigned to recombination.

The two isolates in pair 17 also differed at *opa8*, and again neither had the consensus allele. Allele 133 is considered to be the ancestral allele in the pair, as it is present at the same locus in three other ST261 isolates, including pair 18 (and one ST261/*opa* type 4 isolate [Table 2]), whereas allele 139 was a unique allele. Allele 139 can be derived as a simple mosaic of allele 133 and allele 17, the consensus allele at both *opa3* and *opa4*, and is thus assumed to have arisen by recombination.

Pair 18 differed at *opa7*. One isolate had the consensus allele (allele 45), and the other had allele 95, which differs from the consensus by a codon insertion at the end of HV1. Since the codon insertion was not found in other alleles, allele 95 almost certainly originally arose from the consensus allele by mutation. However, allele 95 was present at *opa7* in ST261 isolates that were recovered up to 8 months before the isolates of pair 18, and the codon insertion is unlikely to have occurred as a de novo mutation in this pair. It is more likely that allele 95 was introduced during coinfection with another isolate of the prevalent ST261 strain that had the codon insertion, and the allelic change within this pair is therefore assigned to recombination.

Therefore, of the five allelic changes in the isolates from recent sexual contacts in Sheffield, one was assigned to point mutation and four were assigned to recombination, three of the latter involving recombination among alleles within the same isolate and the one involving introduction of an allele from another coinfecting isolate of the same strain.

The allelic changes in the London sexual network isolates (Tables 4 and 5) were analyzed in the same way. The majority of changes were assigned to recombination among existing alleles (see below), and only those networks where changes were assigned to importation of alleles or mutation are considered here. Heterosexual network ST610 (Table 5) included six isolates recovered over a period of 3 months, although one was excluded as a complete *opa* profile could not be obtained. A consensus allele was identified at all loci except *opa8* and *opa11*, and all isolates had the same allele at five loci. As *opa8* and *opa11* did not have consensus alleles, no attempt was made to explain the origin of the differences in the alleles at these loci. Three *opa* alleles (209, 211, and 218, shown in red) in the ST610 isolates were not found in any other isolates that were examined and could not have been generated as mosaics of other alleles in these isolates. These novel alleles are candidates for *opa* alleles acquired from another strain during a mixed infection. Allele 183, which differs from the consensus at *opa7*, was also novel to the ST610 network and could not be formed as a mosaic of other alleles, although it was present as a consensus allele at the same locus in both the ST777 heterosexual network and the mixed sexual orientation network (ST584). As discussed above, the origin of allele 183 was unclear but it was probably imported from a coinfecting strain. Thus, in this network, four allelic changes appeared to be due to the introduction by recombination of alleles from other strains and three other changes appeared to be due to recombination among existing alleles.

The isolates in the mixed sexual orientation network, ST584 (Table 5), were recovered over a 2-month period. Six were from heterosexuals, one was from a bisexual, and one was from an MSM. One change to the consensus (at *opa5*) was due to the duplication of a consensus allele at another locus. Two were simple mosaics of alleles present in this network, and one was a complex mosaic. Allele 214 at *opa3* differed from the consensus by a single unique NS substitution in HV2 and was assigned as a point mutation. Allele 155 at *opa2* also had a single NS substitution compared to the consensus allele, in a conserved region that borders HV2, but this substitution is seen in other alleles of this network and was assigned to recombination rather than point mutation.

Summing the data for the isolates from all seven London networks, there were 20 recent allelic changes due to recombination among existing alleles, 4 due to the apparent importation of alleles (all in the same network), and 1 due to point mutation. Combining these values with those from the sexual contacts in Sheffield, there were 24 changes assigned to recombination (3 of which being allele duplications, the remainder being due to the formation of mosaic alleles), 5 to importation of alleles from other strains, and only 2 to point mutation.

Relationship between variation at *opa11* and at the neighboring *pilE* gene. Variation occurs much more rapidly at *opa11* than at other *opa* loci. The pilin expression locus, *pilE*, is located adjacent to *opa11* in strain FA1090. These two genes, along with a silent pilin gene, *pilS*, and a partial pilin gene consisting of portions of interrupted pilin genes (16), are clustered together over a 3-kb region (Fig. 1). Gene conversion events that replace *pilE* sequences with those from *pilS* loci occur frequently in gonococci, and the variable regions of *pilE* were sequenced for the pairs and triplets of isolates obtained from recent sexual contacts in Sheffield to determine whether variation at one locus is linked to that at the other.

In two isolates, the exact sequence of the pilin genes flanking *opa11* could not be determined (designated P in Table 3) due to double peaks in regions of the *pilE* gene, possibly due to a change at *pilE* during growth of the cultures for DNA preparation. Eleven isolates did not have the *pilE* gene downstream of *opa11* (designated absent). In a further two isolates, there appeared to have been a partial deletion involving the 5′ region of *pilE* since no sequence could be obtained with the primer located in this region, while the sequence with the pilE6 primer (present in the 3′ end of both *pilE* and *pilS*) gave double peaks, suggesting that the 3′ end of *pilE* was intact. This was

consistent with the size of the PCR products, being slightly smaller than expected but not as small as the product obtained from isolates where *pilE* is completely absent. Pairs (or triplets) with a *pilE* allele in one isolate and an apparent absence of *pilE* (or one which carries a partial deletion or where an allele could not be assigned) at the other were classed as having different *pilE* genes (Table 3).

Changes between sexual contacts typically occurred at both loci, and only one pair of isolates (pair 14) had no changes at either *pilE* or *opa11*. In almost all of the pairs or triplets, where differences in *pilE* were observed, they were accompanied by differences in *opa11*, suggesting that the two loci have similar rates of variation. Three pairs (5, 9, and 13) had changes at *pilE* but no change at *opa11*, suggesting that changes at the former locus are not linked to changes at the latter; there were no clear examples of changes at *opa11* without changes at *pilE*. However, three pairs (10, 11, and 15) had differences at *opa11* while not appearing to have a *pilE* gene next to *opa11*, indicating a lack of a direct linkage between the presence of the flanking *pilE* locus and the rapid changes at *opa11*.

Stability of *opa* allelic profiles, *pilE* sequences, and CRs during in vitro subculturing. The stability of the *opa* profile was examined during serial subculture. Twenty isolates, including at least one isolate of each ST in Tables 2 to 5, were subcultured 10 times, always from single colonies. A fresh DNA lysate was prepared from each subculture, and the *opa* profiles obtained on the first and last subcultures were compared. At loci where differences in the *opa* profile were detected, the alleles for the intervening subcultures were determined to identify the subculture at which the change occurred. No differences in the *opa* allelic profiles were detected between the 1st and 10th subcultures in 18/20 isolates. The differences in the other two isolates both occurred at *opa11*, one on the seventh and the other on the ninth subculture (Table 6). Both of the new *opa* alleles were mosaics of existing alleles in their profiles and therefore involved recombination events.

We also looked at differences in the adjacent *pilE* gene during subculture. In the two isolates where *opa11* changed, *pilE* also changed, at the same subculture in one isolate but a different subculture in the other (Table 6). An additional six isolates had differences in *pilE*, while *opa11* remained unchanged, suggesting that *pilE* changes more rapidly than *opa11* in vitro. Subcultures in which neither *opa11* nor *pilE* changed are not shown in Table 6.

The numbers of CRs were also compared to estimate how rapidly these change during subculture (data not shown). Ignoring the two *opa11* loci where allelic changes were observed, 9 (45%) of the 20 isolates showed variation in the number of CRs during subculture. In some isolates, this occurred at more than one *opa* locus. A total of 18 (8%) of the 218 *opa* loci that could be compared changed the number of CRs, 50% (9/18) of which were expansions and 33% (6/18) of which were contractions by a multiple of one repeat. The three other loci expanded or contracted by a greater multiple (two or three repeats).

TABLE 6. Variation in *opall* and *pilE* in isolates from subculturing[a]

Isolate, subculture	ST	Allele at:	
		opa11	*pilE*
S350, sub1	12	102	1
S350, sub2		102	1
S350, sub3		102	1
S350, sub4		102	1
S350, sub5		102	1
S350, sub6		102	1
S350, sub7		102	1
S350, sub8		102	1
S350, sub9		249	13
S350, sub10	12	249	13
S280, sub1	12	244	2
S280, sub10		244	14
S427, sub1	12	245	3
S427, sub10		245	15
S324, sub1	261	96	4
S324, sub10		96	16
LG1846, sub1	825	148	5
LG1846, sub10		148	Abs
LG1775, sub1	825	235	6
LG1775, sub10		235	17
LG1304, sub1	825	165	7
LG1304, sub10		165	18
LG0456, sub1	584	241	12
LG0456, sub2		241	19
LG0456, sub3		F[b]	19
LG0456, sub4		241	19
LG0456, sub5		241	19
LG0456, sub6		241	19
LG0456, sub7		250	19
LG0456, sub8		250	19
LG0456, sub9		250	19
LG0456, sub10	584	250	19

[a] As in Tables 1 to 5. Boxed cells show where changes in *opa* or *pilE* alleles were observed.
[b] F, sequencing reactions failed.

DISCUSSION

The main aims of the present study were to understand the extent of diversity among gonococcal *opa* genes and to quantify the relative contributions of mutation and recombination to the diversification of the *opa* allelic profiles of gonococcal strains. As expected, there was very extensive diversity of *opa* sequences and *opa* gene repertoires in gonococci, such that distantly related strains, such as the diverse set of 14 gonococci, had no *opa* alleles in common. Similarly, there was an almost total lack of shared alleles among the other gonococcal strains (NG-MAST STs) that were examined. Where there were identical alleles in different strains (e.g., between ST584 and ST777 or the strains from MSM in London, Sheffield, and Lincoln), it appeared to be due to the recent common ancestry of the strains rather than horizontal spread of *opa* alleles.

Allele sharing due to horizontal transfer requires mixed gonococcal infections and might be expected among isolates recovered from the same location during the same time period, but excepting the ST610 network, it was not found among isolates from the MSM and heterosexual networks in London or among isolates of the two major strains that had been cocirculating among heterosexuals in Sheffield for at least 32 months. For reasons that are unclear, the small ST610 heterosexual cluster was atypical as it included two isolates that appeared to have acquired multiple novel alleles from other strains.

Estimating the relative contributions of point mutation and recombination requires the study of sets of isolates that are very closely related, identifying individual isolates where the normal *opa* profile has changed and, ideally, where the ancestral and derived *opa* alleles can be deduced. The examination of isolates of ST12 and ST261 from Sheffield showed that in both STs the *opa* allelic profiles of the isolates were very similar, and a consensus allele could be identified at 10/11 loci, but there were a considerable number of changes from the consensus, and in many cases the ancestral and derived alleles were not clear. We therefore used isolates that are even more closely related where we could usually identify the direction of allelic change and could with more confidence assign these changes to point mutation or recombination.

Of the 31 examples of recent changes at an *opa* locus, 24 (77%) were assigned to recombination among alleles in the same isolate, of which 3 resulted in allele duplication and the others resulted in the generation of mosaic alleles. A further five changes (16%) appeared to be due to the importation by recombination of alleles from other strains, while point mutation accounted for only two (7%) of the changes. These values clearly demonstrate that the processes involved in the diversification of *opa* repertoires are dominated by recombination events among existing alleles. The introduction of alleles from other strains appeared to be rare, and the value of 16% may be an overestimate as four of the five examples were found within two isolates of the ST610 network.

The predominance of recombination events that generate mosaic *opa* alleles, rather than allele duplication, is not surprising as we defined alleles by using all of the sequence of the *opa* gene except the region encoding the signal peptide. Thus, for an allele to be duplicated, recombination (or gene conversion) would have to occur within small regions at each end of the *opa* gene whereas the formation of mosaic alleles can occur by recombination within the larger conserved regions within *opa* genes. The relative rarity of de novo NS mutations in HV regions does not imply that these are unimportant in the evolution of the *opa* gene repertoire. All of the observed nucleotide differences between HV regions ultimately have been generated by mutation, but the appearance of new mutations occurs much less frequently than the generation of novel mosaic *opa* alleles by the shuffling of the HV regions of different *opa* genes by recombination. Shuffling of HV regions may be favored, as the existing sequences are likely to bind to their receptors and new combinations may escape existing immunity if the major epitopes recognized by the host immune system are formed by amino acids contributed by both HV1 and HV2.

The rates of change at *opa* loci are difficult to estimate with any precision, but they clearly differ markedly between loci. Changes at *opa11* occurred extremely frequently, with changes occurring in 17/21 (81%) pairs (or triplets) of isolates from recent sexual contacts, whereas no changes at *opa10* were found within any of the sets of closely related gonococci that were examined. Changes at *opa* loci occur more rapidly than changes at the antigen-encoding genes (*por* and *tbpB*) used to define the NG-MAST ST, since multiple isolates of a single ST differed at up to five *opa* loci (excluding *opa11*, which differed in most isolates on the same ST).

Meningococcal isolates of the same MLST ST or clonal complex that were recovered over several decades still show considerable similarities in their *opa* allelic profiles (10, 19), with retention of the same allele in one clonal complex over more than 50 years and, in two clonal complexes, of the same *opa* allelic profile in isolates recovered 25 years apart (10). The sets of meningococcal isolates studied by Callaghan et al. are much more distantly related than the sets of gonococcal isolates used in this study, as the former are defined by having the same or similar allelic profiles by MLST, which indexes variation in housekeeping genes that evolve slowly, whereas the latter are defined by identity of antigen genes, which evolve very much more rapidly. Although the meningococcal sets of isolates are much less closely related than the sets of gonococcal isolates, they appear not to have diversified to a much greater extent than some of the sets of gonococcal isolates. Gonococcal *opa* loci therefore appear to change at a higher rate than in meningococci. To at least some extent, this is likely to reflect the increased number of *opa* genes in gonococci, which provide more opportunities for recombination between existing alleles.

The presence of identical alleles at multiple *opa* loci was a common phenomenon, occurring in more than 70% of the diverse strains (and in many of the strains in Tables 2 to 5), and 10% of the recent changes in *opa* profiles involved allele duplication. Duplicated alleles have also been observed in meningococci (10, 27). The presence of duplicated alleles could be simply an inevitable result of gene conversion activity, but it is tempting to believe that such events are favored under some circumstances and occur as a result of selection. The CAECAM family of receptors which bind to Opa proteins have been shown to require specific combinations of HV1 and HV2 (8), and in gonococci, which can colonize both male and female genital tracts and the rectum (and, at least transiently, the throat), there may be *opa* genes with differential specificities for these different niches. The dynamic nature of the *opa* allelic profile may reflect selective forces that favor the duplication of those *opa* genes that allow the colonization of a major niche, increasing the chance of expression of such Opa proteins at the cell surface, combined with subsequent selection against these *opa* alleles, and for novel *opa* alleles, imposed by the host immune response.

Bhat et al. (4) previously reported one gonococcal *opa* locus (*opaE* in strain MS11) to be more variable than others, and this was located next to *pilE* and thus corresponds to *opa11* in our nomenclature. The *pilE* locus is the expression site for the major pilin subunit and is known to have one of the highest measured rates of antigenic variation (13), which occurs via gene conversion events between *pilE* and numerous silent partial pilin genes (*pilS* loci). Examination of the sequences of the adjacent *pilE* gene for each of the pairs and triplets from sexual contacts in Sheffield showed that a change at one locus was almost always accompanied by a change in the other. This implied that the two loci have similarly high rates of variation but does not provide evidence of a direct causal relationship.

The absence of *pilE* in some isolates was surprising and unlikely to be due to loss during subculture or storage since, in the majority of cases, *pilE* was absent in both isolates from sexual contact pairs (Table 3). Some gonococcal strains have a second copy of *pilE* elsewhere on the chromosome (e.g., strain MS11; reference 5), and those strains that lack a *pilE* gene adjacent to *opa11* may have the gene at this alternative loca-

tion. If the recombinational mechanism by which *pilS* sequences are introduced into the *pilE* locus somehow drives variation at *opa11*, this gene should vary at a rate similar to that of the other *opa* loci if the adjacent *pilE* gene were not present at this location. The absence of *pilE* at this genomic location in some strains could be used to establish if the rate of *opa11* variation is much lower in strains that lack the flanking *pilE* locus. The *opa* and *pilE* genes were also found to vary during subculture. The two changes that occurred at *opa11* generated mosaic alleles and are therefore clear evidence that mosaics can arise by recombination between the alleles within an isolate. Eight changes were observed at *pilE*, indicating that *pilE* changes more rapidly that *opa11* in vitro. In the two cases where *opa11* changed, there was also a change at *pilE*, although in one case this occurred at a different subculture. These changes are considered to simply reflect the greater rate of *pilE* changes compared to *opa11* changes rather than any mechanistic relationship between the changes at the two loci.

The multiples of the CRs also varied on subculture, with 45% of the subcultured strains showing changes in the multiple of CRs in at least one of their *opa* loci. This is consistent with the high frequency of CR phase variation in vitro (1×10^{-3} to 3×10^{-3} per cell per generation) reported by Belland et al. (2). Consequently, it is not possible to know which of the *opa* genes are being expressed in vivo once isolates have been cultured from a clinical specimen.

In conclusion, there is very extensive diversity among the *opa* genes of gonococci, and unrelated strains are unlikely to share any *opa* alleles. The *opa* allelic profiles change very rapidly such that isolates that have not accumulated any changes in the antigen genes *por* and *tbpB* often have undergone changes at multiple *opa* loci. The majority of events which led to the diversification of *opa* allelic profiles were due to recombination, primarily those which resulted in mosaics of the *opa* genes present in the repertoire and allele duplication. Mutations and import of *opa* sequences from other strains occurred much less frequently than changes due to recombination. To understand the origins of *opa* variation, we have focused on nucleotide sequence changes within *opa* genes, but the biologically significant events are those that change the amino acid sequences of the HV regions or which bring together new combinations of HV regions through recombination. An analysis of the role of mutation and recombination in generating novel Opa proteins and novel HV1-HV2 combinations will be discussed elsewhere.

ACKNOWLEDGMENTS

We are grateful to Iona Martin, Gill Bell, and George Kinghorn for strains and sexual contact data from studies of gonorrhea in Sheffield and to Bhudipa Choudhury, Cynthia Bishop, Claire Risley, and Azra Ghani for strains and epidemiological data from gonorrhea in London. We thank Daniel Godoy for providing gonococcus-optimized MLST primers.

This work was funded by Wellcome Trust grant 030662. B.G.S. is a Wellcome Trust Principal Research Fellow.

REFERENCES

1. **Aho, E. L.** 1989. Molecular biology of class 5 outer membrane proteins of *Neisseria meningitidis*. Microb. Pathog. **7:**249–253.
2. **Belland, R. J., S. G. Morrison, P. van der Ley, and J. Swanson.** 1989. Expression and phase variation of gonococcal P.II genes in *Escherichia coli* involves ribosomal frameshifting and slipped-strand mispairing. Mol. Microbiol. **3:**777–786.
3. **Belland, R. J., S. G. Morrison, J. H. Carlson, and D. M. Hogan.** 1997. Promoter strength influences phase variation of neisserial *opa* genes. Mol. Microbiol. **23:**123–135.
4. **Bhat, K. S., C. P. Gibbs, O. Barrera, S. G. Morrison, F. Jähnig, A. Stern, E.-M. Kupsch, T. F. Meyer, and J. Swanson.** 1991. The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are encoded by a family of 11 complete genes. Mol. Microbiol. **5:**1889–1901.
5. **Bihimaier, A., U. Römling, T. F. Meyer, B. Tümmler, and C. P. Gibbs.** 1991. Physical and genetic map of the *Neisseria gonorrhoeae* strain MS11-N198 chromosome. Mol. Microbiol. **5:**2529–2539.
6. **Bilek, N., I. M. Martin, G. Bell, G. R. Kinghorn, C. A. Ison, and B. G. Spratt.** 2007. Concordance between *Neisseria gonorrhoeae* genotypes recovered from known sexual contacts. J. Clin. Microbiol. **45:**3564–3567.
7. **Bos, M., F. Grunert, and R. Belland.** 1997. Differential recognition of members of the carcinoembryonic antigen family by Opa variants of *Neisseria gonorrhoeae*. Infect. Immun. **65:**2353–2361.
8. **Bos, M. P., D. Kao, D. M. Hogan, C. C. R. Grant, and R. J. Belland.** 2002. Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains. Infect. Immun. **70:**1715–1723.
9. **Brooks, G. F., L. Olinger, C. J. Lammel, K. S. Bhat, C. A. Calvello, M. L. Palmer, J. S. Knapp, and R. S. Stephens.** 1991. Prevalence of gene sequences coding for hypervariable regions of Opa (protein II) in *Neisseria gonorrhoeae*. Mol. Microbiol. **5:**3063–3072.
10. **Callaghan, M. J., K. A. Jolley, and M. C. J. Maiden.** 2006. Opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. Infect. Immun. **74:**5085–5094.
11. **Choudhury, B., C. L. Risley, A. C. Ghani, C. J. Bishop, H. Ward, K. A. Fenton, C. A. Ison, and B. G. Spratt.** 2006. Identification of individuals with gonorrhoea within sexual networks: a population-based study. Lancet **368:** 139–146.
12. **Connell, T. D., D. Shaffer, and J. G. Cannon.** 1990. Characterization of the repertoire of hypervariable regions in the protein II (*opa*) gene family of *Neisseria gonorrhoeae*. Mol. Microbiol. **4:**439–449.
13. **Criss, A. K., K. A. Kline, and H. S. Seifert.** 2005. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. Mol. Microbiol. **58:**510–519.
14. **de Jonge, M. I., H. J. Hamstra, L. van Alphen, J. Dankert, and P. van der Ley.** 2003. Mapping the binding domains on meningococcal Opa proteins for CEACAM1 and CEA receptors. Mol. Microbiol. **50:**1005–1015.
15. **Haas, R., and T. F. Meyer.** 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: Evidence for gene conversion. Cell **44:**107–115.
16. **Hamrick, T. S., J. A. F. Dempsey, M. S. Cohen, and J. G. Cannon.** 2001. Antigenic variation of gonococcal pilin expression *in vivo*: analysis of the strain FA1090 pilin repertoire and identification of the *pilS* gene copies recombining with *pilE* during experimental human infection. Microbiology **147:**839–849.
17. **Hauck, C. R., and T. F. Meyer.** 2003. 'Small' talk: Opa proteins as mediators of *Neisseria*-host-cell communication. Curr. Opin. Microbiol. **6:**43–49.
18. **Hobbs, M. M., A. Seiler, M. Achtman, and J. G. Cannon.** 1994. Microevolution within a clonal population of pathogenic bacteria: recombination, gene duplication and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. Mol. Microbiol. **12:**171–180.
19. **Hobbs, M., B. Malorny, P. Prasad, G. Morelli, B. Kusecek, J. Heckels, J. Cannon, and M. Achtman.** 1998. Recombinational reassortment among *opa* genes from ET-37 complex *Neisseria meningitidis* isolates of diverse geographical origins. Microbiology **144:**157–166.
20. **Jerse, A. E., M. S. Cohen, P. M. Drown, L. G. Whicker, S. F. Isbey, H. S. Seifert, and J. G. Cannon.** 1994. Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. J. Exp. Med. **179:**911–920.
21. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. **5:**150–163.
22. **Kupsch, E.-M., B. Knepper, T. Kuroki, I. Heuer, and T. F. Meyer.** 1993. Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by *Neisseria gonorrhoeae* for human leukocytes and epithelial cells. EMBO J. **12:**641–650.
23. **Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
24. **Malorny, B., G. Morelli, B. Kusecek, J. Kolberg, and M. Achtman.** 1998. Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins. J. Bacteriol. **180:**1323–1330.
25. **Martin, I. M. C., A. Ghani, G. Bell, G. Kinghorn, and C. A. Ison.** 2003. Persistence of two genotypes of *Neisseria gonorrhoeae* during transmission. J. Clin. Microbiol. **41:**5609–5614.
26. **Martin, I. M. C., C. A. Ison, D. M. Aanensen, K. A. Fenton, and B. G. Spratt.** 2004. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. J. Infect. Dis. **189:**1497–1505.

27. **Morelli, G., B. Malorny, K. Muller, A. Seiler, J.-F. Wang, J. del Valle, and M. Achtman.** 1997. Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. Mol. Microbiol. **25:**1047–1064.
28. **Murphy, G. L., T. D. Connell, D. S. Barritt, M. Koomey, and J. G. Cannon.** 1989. Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell **56:**539–547.
29. **O'Rourke, M., C. A. Ison, A. M. Renton, and B. G. Spratt.** 1995. Opa-typing: a high-resolution tool for studying the epidemiology of gonorrhoea. Mol. Microbiol. **17:**865–875.
30. **Risley, C. L., H. Ward, B. Choudhury, C. J. Bishop, K. A. Fenton, B. G. Spratt, C. A. Ison, and A. C. Ghani.** 2007. Geographical and demographic clustering of gonorrhoea in London. Sex. Transm. Infect. **83:**481–487.
31. **Stern, A., M. Brown, P. Nickel, and T. F. Meyer.** 1986. Opacity genes in *Neisseria gonorrhoeae*: control of phase variation and antigenic variation. Cell **47:**61–71.
32. **Stern, A., and T. F. Meyer.** 1987. Common mechanisms controlling phase and antigenic variation in pathogenic neisseriae. Mol. Microbiol. **1:**5–12.
33. **Swanson, J., O. Barrera, J. Sola, and J. Boslego.** 1988. Expression of outer membrane protein II by gonococci in experimental gonorrhea. J. Exp. Med. **168:**2121–2129.
34. **Toleman, M., E. Aho, and M. Virji.** 2001. Expression of pathogen-like Opa adhesins in commensal *Neisseria*: genetic and functional analysis. Cell. Microbiol. **3:**33–44.
35. **Virji, M., K. Makepeace, D. J. P. Ferguson, and S. M. Watt.** 1996. Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic neisseriae. Mol. Microbiol. **22:**941–950.
36. **Ward, H., C. Ison, S. Day, I. Martin, A. Ghani, G. Garnett, G. Bell, G. Kinghorn, and J. Weber.** 2000. A prospective social and molecular investigation of gonococcal transmission. Lancet **356:**1812–1817.
37. **Wiertz, E., A. Delvig, E. Donders, H. Brugghe, L. van Unen, H. Timmermans, M. Achtman, P. Hoogerhout, and J. Poolman.** 1996. T-cell responses to outer membrane proteins of *Neisseria meningitidis*: comparative study of the Opa, Opc, and PorA proteins. Infect. Immun. **64:**298–304.