JOURNAL OF BACTERIOLOGY, Mar. 2009, p. 1878–1890 Vol. 191, No. 6
0021-9193/09/$08.00+0 doi:10.1128/JB.01518-08

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-

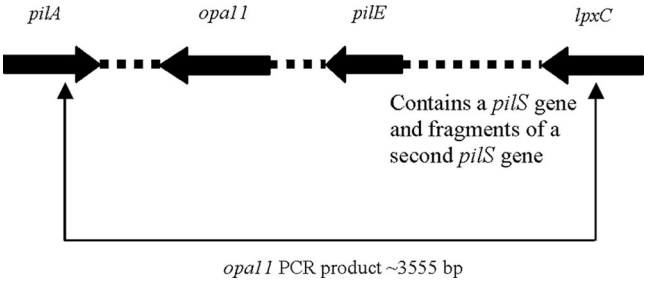


FIG. 1. Region of chromosome amplified for sequencing of the *opa11* gene. In *N. gonorrhoeae* strain FA1090, the 3.5-kb amplified fragment (vertical arrows) includes the *pilE* expression locus and a silent pilin gene, *pilS*. The *pilA* and *lpxC* genes encode a FtsY homolog and a deacetylase involved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

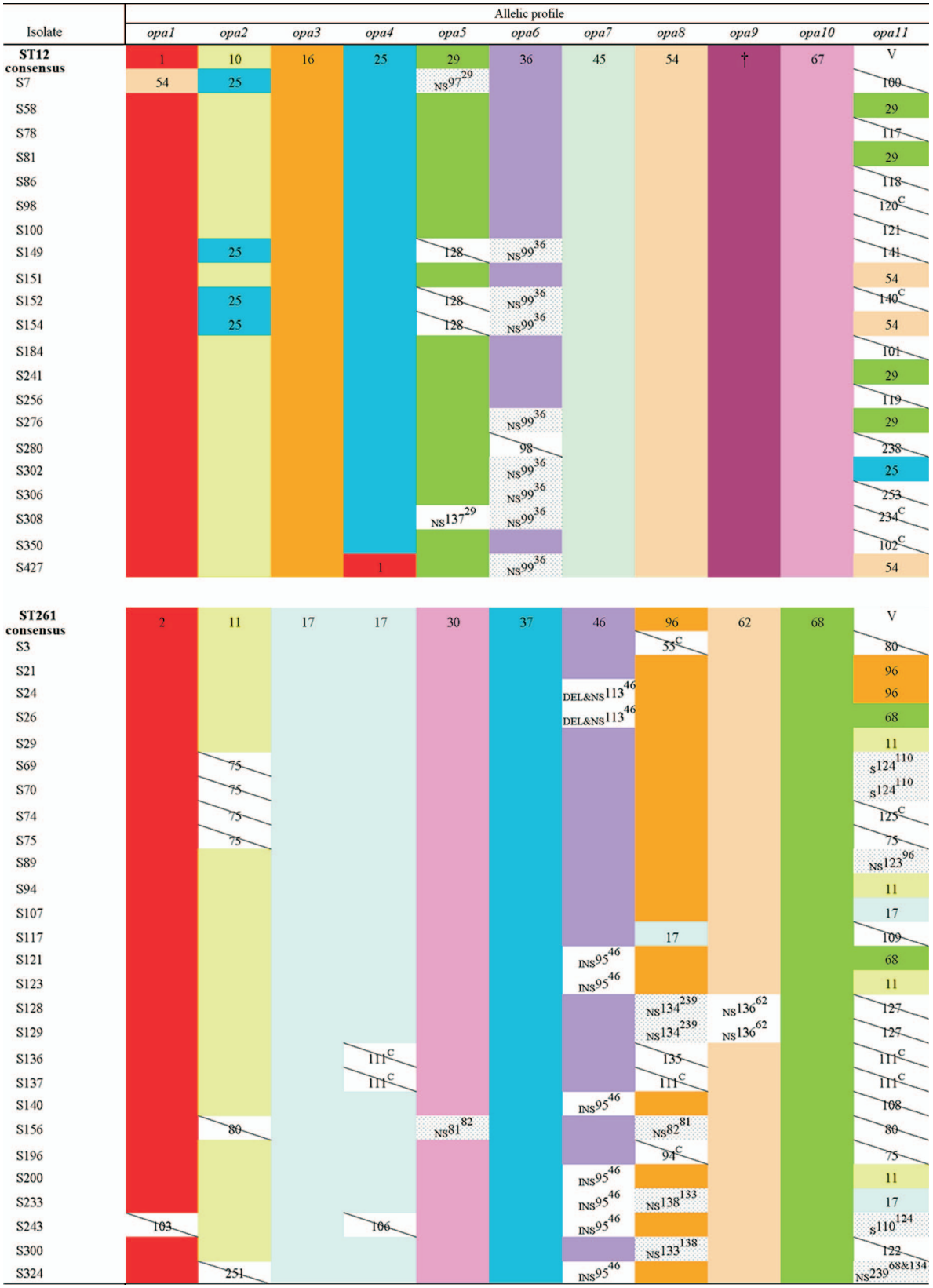
The differences in *opa* allelic profiles of isolates within the ST12/*opa* type 1 and ST261/*opa* type 4 clusters were unexpected, as the fragment patterns obtained by *opa* typing had previously been shown to be indistinguishable. In silico *opa* typing was carried out (data not shown) by using the 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	s211[223]			159		183	187			s223[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.