

*Review*

## Genomic and Global Approaches to Unravelling How Hypermutable Sequences Influence Bacterial Pathogenesis

Fadil A. Bidmos and Christopher D. Bayliss \*

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK;  
E-Mail: fab11@le.ac.uk

\* Author to whom correspondence should be addressed; E-Mail: cdb12@le.ac.uk;  
Tel.: +44-116-2523465; Fax: +44-116-2523465.

*Received: 4 December 2013; in revised form: 6 January 2014 / Accepted: 13 February 2014 /  
Published: 25 February 2014*

---

**Abstract:** Rapid adaptation to fluctuations in the host milieu contributes to the host persistence and virulence of bacterial pathogens. Adaptation is frequently mediated by hypermutable sequences in bacterial pathogens. Early bacterial genomic studies identified the multiplicity and virulence-associated functions of these hypermutable sequences. Thus, simple sequence repeat tracts (SSRs) and site-specific recombination were found to control capsular type, lipopolysaccharide structure, pilin diversity and the expression of outer membrane proteins. We review how the population diversity inherent in the SSR-mediated mechanism of localised hypermutation is being unlocked by the investigation of whole genome sequences of disease isolates, analysis of clinical samples and use of model systems. A contrast is presented between the problematical nature of analysing simple sequence repeats in next generation sequencing data and in simpler, pragmatic PCR-based approaches. Specific examples are presented of the potential relevance of this localized hypermutation to meningococcal pathogenesis. This leads us to speculate on the future prospects for unravelling how hypermutable mechanisms may contribute to the transmission, spread and persistence of bacterial pathogens.

**Keywords:** phase variation; simple sequence repeat; phasevariome; phasotype; localised hypermutation; bacterial genome

---

## 1. Introduction

Adaptability is a common feature of bacterial pathogens. The rapidly changing milieu of an inflammatory host environment and traversing between host tissues with their differing nutrient contents and innate/adaptive immune effectors requires the induction of a diverse set of defence/attack molecules and metabolic pathways by the bacterial cells. Additionally, strong selective pressures are also exerted on these pathogens by transmission within an immunologically and genetically variable host population. Rapid, stochastic generation of phenotypic variation by “hypermutable” mechanisms is a widespread adaptive phenomenon that has evolved within the genomes of several bacterial pathogens as a response to these forces [1–4]. The prevalence of these mechanisms in disease isolates and the amounts of variation manifest during disease is amenable to high throughput global analysis. This review will consider the nature of one of these hypermutable mechanisms—simple sequence repeat tracts (SSRs)—examining the experimental techniques utilised for analysis, particularly the utility of next generation sequencing (NGS), and examples of the potential contributions of this mechanism to disease.

## 2. Localised Hypermutation, Phase Variation and the Mechanisms of Hypermutation

The mutability of genomes varies as a function of the DNA sequence and the complement of DNA replication and repair enzymes. The average mutation rate of a bacterial gene is  $1 \times 10^{-9}$  mutations/division, but there are certain sequences, such as microsatellites, whose mutation rates are significantly higher than this average [5,6]. Evolution has acted on these sequences, sometimes in combination with trans-acting factors, to evolve loci with mutation rates up to  $1 \times 10^{-3}$  per division [7]. This heightened mutability is referred to as localised hypermutation and is contained in ‘contingency’ loci [8]. These loci are thought to have evolved for adaptation to the fluctuating pressures acting on bacterial pathogens.

The major mechanisms of localised hypermutation are site-specific recombination, homologous recombination of tandem duplications of DNA sequences, simple sequence repeats and G-quartet-mediated gene conversion [3,4,8–10]. Only one example of this last mechanism has been described, and it mediates antigenic variation in the major pilin sub-unit of *Neisseria* [11,12]. The variation is produced by the insertion of sequences into an expression locus (*i.e.*, the *pilE* gene) from silent partial copies of the gene, producing a vast array of allelic variants. The other mechanisms generate reversible mutations, meaning that there is no loss of genetic information. These hypermutable mechanisms often result in reversible switches in the expression of the associated genes and phenotypes. These stochastic, high frequency, reversible switches in gene expression are responsible for the phenomenon of phase variation (PV), which is widely observed in bacterial pathogens [8,13].

The majority of contingency loci identified to date encode outer membrane proteins (OMPs), enzymes involved in the modification of surface molecules or restriction-modification systems. The known and potential functions of these loci have been extensively reviewed elsewhere [1–4,10,14]. The main functions of these loci are, however, in mediating interactions between the pathogen and their hosts or altering resistance to phage infection. The contributions of contingency loci to the pathology, course of disease, host persistence, transmission and spread of bacterial pathogens is an

expanding area of research relevant to a diverse range of species from *Mycoplasmas*, *Campylobacters*, *Helicobacters*, *Neisseria*, *Bacteroides* and coliforms. This review will consider only SSR-mediated hypermutation and highlight some examples of how research into these types of contingency loci is improving our understanding of aspects of bacterial pathogenesis.

### 3. Detection of Hypermutable Sequences in Genomes

#### 3.1. A Definition of Functional SSR Classes

One of the most widespread mechanisms of localised hypermutation involves hypermutable SSR sequences. These sequences are easily detected in genomes; however, there is no consensus on the definition of a ‘functional SSR’ or for classification into different functional classes. We set out below a scheme for separating SSR into different classes based on their potential to frequently and reproducibly mediate switching (*i.e.*, PV). Mutation rate is an important determinant of the ability of an SSR to mediate PV and to facilitate the adaptation of a bacteria clone. Mutation rate can be determined experimentally and hence provides a non-arbitrary approach to separating SSRs into different classes. The determination of mutation rates is, however, difficult and can be confounded by under- or over-estimating the number of divisions (*i.e.*, bacterial generations) relative to mutational events [5]. Nevertheless, the mutation rate of an SSR is proportional to the repeat number and influenced by the repeat unit length/sequence [15]. Hence, the repeat number can be used to separate SSRs into different ‘functional’ classes, which we have delineated as hypermutable, mutable and proto-mutable. Although, the boundaries for these classes are not easily defined and are partially arbitrary, the classifications provide a framework for distinguishing the evolved loci subject to localised hypermutation from those SSRs whose occurrence is influenced by genome composition. These classes are described below, with Table 1 providing predictions of the mutation rates and SSR tract lengths for each class.

Hypermutable SSRs consist of multiple repeat units that reproducibly generate insertions and deletions in the SSR at high rates. These hypermutable SSRs are responsible for PV. Phenotypic and molecular characterisation correlating switches in gene expression with alterations in the SSR are required to definitively ascribe a locus or specific tract length to this class of SSRs. These studies should be performed in the absence of selection, as a strong selection can amplify low frequency events, implicating short SSRs as being hypermutable and involved in PV, whereas in natural infections, such mutations would rarely mediate adaptation. Hypermutable SSRs are thought to have evolved due to frequent selection for alterations in the phenotypes controlled by mutations in the SSR, resulting in secondary selection for higher rates of mutation in the SSR (*i.e.*, longer tract lengths). These loci are likely to have evolved from other hypermutable SSRs or to be due to an increase in the tract length of a mutable SSR.

Mutable SSRs have intermediate length SSR tracts and, hence, lower mutation rates. Reversible switches in phenotypes due to these tracts may be rare, require strong selection for detection and may not always involve mutations in the SSR [16,17]. Mutable SSRs do not arise due to selection for mutability, but as a result of genome composition and the molecular forces acting on proto-mutable SSRs [18].

Thus, a higher propensity for insertions in short SSRs may result in an increase in tract length and convert a proto-mutable SSR into a mutable SSR that occasionally produces adaptive mutations.

**Table 1.** Classification of bacterial sequence repeat tracts (SSRs).

Type of SSR	Mutation Rate (Mutations/Division)	Predicted Range of Repeat Numbers (References <sup>1</sup> )
Proto-mutable	$10^{-9}$ to $10^{-7}$	<6G [19]
		2-4CA [20]
		2-4TAAA
		2-4AGTC
Mutable	$10^{-7}$ to $10^{-5}$	GACGAGAAGA
		6G-8G [21]
		4-10CA [20]
		3-9TAAA
		3-12AGTC [15]
Hypermutable	$10^{-5}$ to $10^{-2}$	2GACGA
		>7G [21–23]
		>10CA[20]
		>9 TAAA[24]
		>10 AGTC [15]
		3+GACGA [25]

<sup>1</sup> The indicated references contain experimental data on mutation rates for tracts of a particular type.

Proto-mutable SSRs consist of small numbers of repeat units and have mutation rates marginally higher than the genome average. These tracts are numerous in bacterial genomes and can arise by chance as a result of the nucleotide composition and codon usage of a species. These short SSRs are prone to generating mutations capable of inactivating genes, and hence, the distribution is uneven with higher numbers in the ends of genes [18]. These proto-mutable SSRs rarely produce adaptive mutations, but are the precursors for the formation of the longer tracts.

### 3.2. Detection of Multiple SSR in the Bacterial Genomes

SSR are easily detected in bacterial genomes by simple word searches and through widely-available scripts. Numerous websites offer online searches for SSRs (e.g., [26]). One of the first whole genome sequences, *Haemophilus influenzae* strain Rd, was analysed for the presence of SSRs, resulting in the identification of 12 loci subject to SSR-mediated PV; this finding represented a significant advance in our knowledge of the repertoire of phase variable genes in this organism [27]. These loci contained tetranucleotide repeat tracts consisting of five or more repeats, with the majority containing 15+ repeats. These tracts were longer than predicted to occur based solely on genome composition, suggesting their presence was due to selection for mutability.

Whole genome sequences of other important pathogens similarly identified the presence of multiple loci whose expression may be controlled by SSRs [28–30]. The types of SSRs vary between species, suggesting different evolutionary pathways resulting in convergence towards localised hypermutation as a mechanism of adaptation.

A major area of research is to catalogue and compare the number and distribution of SSRs and functions of SSR loci within and between species [31–33]. SSRs within specific loci of a species exhibit a variable presence. Thus, 55 SSR-associated loci were found in four complete *C. jejuni* genomes, but only 17 were conserved in all the genomes; and only five of these loci contained an SSR in all strains [22]. The dispersed nature of SSR-containing hypermutable genes within a species is indicative of a variable selection for these loci, due to the redundancy of their functions or a requirement for these loci for infrequently encountered selective pressures (e.g., an unusual niche or resistance to a subset of bacteriophages). Alternatively, differences in the presence or length of the SSR in a particular locus may suggest weak selection for hypermutation. More detailed analyses of genome sequences will begin to unravel the complex link between the repertoire of phase variable loci and the selective pressures driving their evolution and distribution.

#### 4. Expression States, Phasotypes and the Phasevariome

##### 4.1. Analysis of Expression States of Phase Variable Genes

SSRs mediate PV by altering either the transcription or the translation of a gene [4]. Alterations in repeat number of SSRs located in coding sequences result in a gain/loss of expression due to frameshifts (referred to herein as ‘translational PV’). These SSRs have to contain non-triplet repeat units, otherwise the changes have no effect on expression (although they can alter antigenicity and function, due to the insertion of extra amino acids). The majority of translational switching is between the ON and OFF, with one ON state for every two OFF states. Exceptions have been described wherein the SSR is located within the 5' end of the gene, and there are multiple initiation codons upstream of the SSR in different reading frames with differential rates of translation from these codons, producing variations in gene expression [34]. There are also cases where the SSR is at the end of the open reading frame, such that changes in the SSR alter protein length with unpredictable effects on protein function [35].

For those SSRs mediating PV through alterations in transcription, there is a different association between tract length and expression status that is not directly predictable from genomic data. The simplest type of mechanism occurs when the repeat is located between the –10 and –35 binding sites for the RNA polymerase holoenzyme. Expression levels are dependent on an optimal distance (17/18 nucleotides) between these elements, and SSRs modulate expression by altering this distance [36,37]. The result is that optimal expression is usually centred around one tract length. However, uncertainty in the actual start and end sites of the –10 and –35 motifs necessitates experimental evaluation of the expression levels of phase variants with differing tract lengths for the accurate correlation of SSR number and expression state [36]. This complicates ascertaining predictions of expression status directly from genomic sequences.

In some cases, SSRs are located upstream or downstream of the core promoter. The *nadA* gene is an example of an upstream SSR, where variation modulates interactions between repressor proteins/activator sequences and the RNA polymerase holoenzyme, thereby modulating expression [38]. A recent dissection of a downstream SSR in *Helicobacter pylori* has demonstrated a role for this SSR in modulating interactions with a small untranslated RNA (sRNA) that regulates transcript stability

(S. Pernitzsch, D. Beier and Cynthia M. Sharma; Targeting of a homopolymeric G-repeat by a small RNA mediates the repression of a chemotaxis receptor in *Helicobacter pylori*. Abstract from the Campylobacter, Helicobacter and Related Organisms meeting, CHRO2013, in 2013 at Aberdeen, Scotland).

#### 4.2. Phasotypes

For both translational and transcriptional PV, rules can be established linking repeat number with a particular expression status. Thus, the genotypes (repeat number) can be turned into an expression status (e.g., ON or OFF), and these states can be coded (Table 2). These expression states contain phenotypic information, but are not directly correlated with phenotype, as the expression of some of these loci can be regulated at other levels. Thus, some of the phase variable meningococcal genes are only expressed under low iron conditions [39,40]. Other phase variable genes may act to sequentially add sugar moieties, such that the phenotypic expression of the late-acting genes is dependent on the expression status of an early-acting gene. An example of this type is with the neisserial *pglA* and *pglE* genes, which act sequentially to add galactose monomers during protein glycosylation [41,42]. These coded expression states are therefore referred to as phasotypes to indicate that they are predictions of the phenotypic state of a bacterial isolate, not the actual state.

**Table 2.** Generation of a meningococcal phasotype from SSR repeat numbers.

Gene	<i>fetA</i> <sup>1</sup>	<i>porA</i> <sup>1</sup>	<i>opc</i> <sup>1</sup>	<i>nadA</i> <sup>1</sup>	<i>hpuA</i> <sup>2</sup>	<i>nalP</i> <sup>2</sup>
Tract Length	9C	12G	11C	9TAAA	10G	11C
Expression State <sup>1</sup>	2	2	1	0	2	2
Phasotype	2-2-1-0-2-2					

<sup>1</sup> Genes subject to transcriptional switching such that expression states are coded into three states: high (2), intermediate (1) and low (0); <sup>2</sup> Genes subject to translational switching leading to coding into two expression states, ON (2) and OFF (0).

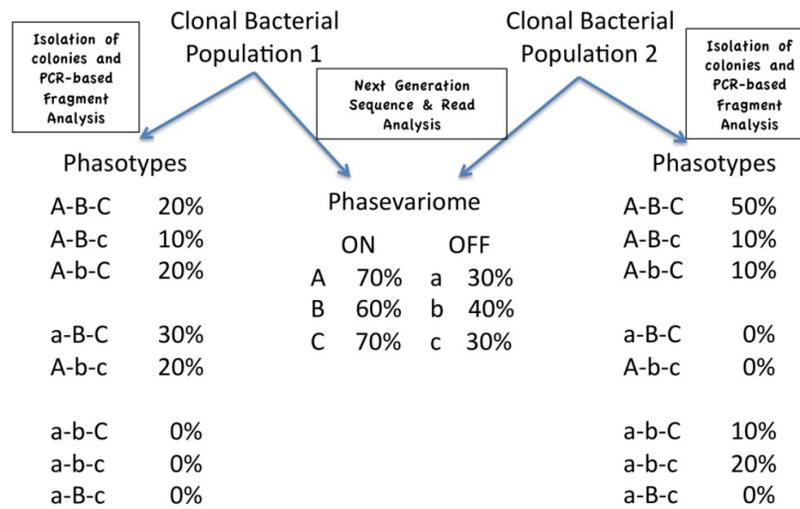
#### 4.3. Phasotypes and the Phasevariome

Analysis of the amounts of each state of a phase variable gene in a particular bacterial population is a key technique for understanding how this phenomenon contributes to adaptation and pathogenesis. In organisms with multiple phase variable genes, a population will have a particular combination of expression states for each phase variable gene—the phasevariome—whilst each cell will have a particular combination of expression states—the phasotype. As Figure 1 illustrates, the phasotypes cannot necessarily be directly predicted from the phasevariome. The analysis of the phasevariome can proceed by extracting DNA from a sample of the population and typing of the SSRs for each phase variable gene. The phasotypes, however, can only be examined by the typing of SSRs in individual cells or a surrogate—colonies grown from individual cells with the assumption that the SSRs do not change in the majority of the colony during growth. This latter approach requires plating dilutions of the population and then picking and analysing individual colonies.

Figure 1 shows the extra information contained in an analysis of phasotypes as compared to the phasevariome. The phasevariome can be obtained by next generation sequencing and describes the proportions of variants in an ON and OFF state for each gene in a population of bacterial cells.

Phasotypes are obtained by the analysis of single-cell derivatives (e.g., a number of colonies grown from the population) and a PCR-based fragment analysis of each phase variable gene. Each single-cell derivative will have a particular combination of ON and OFF states for a set of phase variable genes (the phasotype). This figure depicts the analysis of a bacterial population with three phase variable genes and, hence, eight possible phasotypes. Critically, two populations with the same phasevariome can have different proportions of the eight possible phasotypes.

**Figure 1.** Phasevariome and phasotypes.



## 5. Analysis of SSR Diversity by Next Generation Sequencing

### 5.1. Utility of NGS for Comparison of SSR Prevalence in Bacterial Genomes

Comparisons of the prevalence of SSRs in bacterial genomes, providing an improved understanding of the evolutionary pressures acting on these sequences, has been facilitated by the NGS-driven increase in genome sequences from a wide range of species [22,33]. Similarly, strain-to-strain comparisons have also been transformed with an explosion in the numbers of genome sequences, providing the raw materials for the examination of the intra-species variability in SSR prevalence and conservation. Some studies utilising these large databases are outlined below. Nevertheless, there are problems with older and current NGS approaches, limiting the utility of some of the data for SSR tracts.

### 5.2. Methodological Impact of NGS on the Determination of Tract Length

A critical feature of an SSR tract is the repeat number, as this is a determinant of mutability [15,23]. NGS methodologies have an impact on the determination of repeat number when tracts contain large numbers of repeats or are composed of mononucleotide repeat tracts.

Long tracts of SSRs with large repeat units cannot be determined by short-read technologies, as a single read cannot span the entire repeat tract. This was a particular problem for early Illumina data, when reads were only 30 nucleotides, and so, some of the older NGS genome sequences have contig breaks in these sequences. Improvements in this technology, such as Illumina HiSeq, and the use of

Roche 454 in combination with novel bioinformatics search programs have overcome this limitation, and most SSR tracts are now encompassed in NGS datasets [43].

Mononucleotide repeat tracts are, however, still a particular problem. A study of the Illumina read data for a *Drosophila* genome demonstrated that accuracy was similar to non-repetitive regions for pentanucleotide repeats, but decreased with shorter repeat units [44]. Mononucleotide tracts had the highest error rates, with internal concordance falling below the accuracy cut-off for tracts of 13 or more nucleotides. One outcome of these limitations of NGS data is that contig breaks occur frequently within phase variable genes. The analysis of meningococcal genomic data generated by Illumina indicates high error rates occurring in the polyC or polyG tracts of eight or more nucleotides, as 22 of 31 tracts were found to occur at the ends of contigs [45]. Comparisons of other recently developed platforms have also found variability in error rates and suggest that Illumina's MiSeq may provide the greatest accuracy on repetitive tracts as compared to PacBio and Ion Torrent [46]. Thus, the determination of the actual tract length for these loci by NGS is impacted by the technology. These loci are, however, also difficult to sequence by standard Sanger-based technologies, which fail, due to a combination of the stalling of the DNA polymerase and the slippage of the DNA strands during re-initiation of DNA polymerisation, producing over-lapping sequence reads. As some NGS technologies are not reliant on a polymerase, advances in the methodology will eventually increase the accuracy across these tracts.

### 5.3. Methodological Impact of Assembly on Analysis of Multiple-Copy Loci

The potential for antigenic variation generated by phase-variable loci is enhanced by redundancies in function. The presence of two or more copies of different alleles of a phase-variable gene within a genome is common in bacterial genomes. The most extreme example is the presence of 12 alleles of the *opa* gene in *N. gonorrhoeae* genomes, which all have identical N- and C-termini, but differ in the antigenic hyper-variable regions [47]. These repetitive regions can be compromised by the NGS and/or assembly method. Distributed loci, such as the *opa* genes, can be assembled from paired-end reads, as the unique read can anchor gene reads to the correct part of a genome. *De novo* assembly may also be better than mapping to a reference genome, as mapping forces reads from different genes together. Duplicated genes in close proximity are more difficult to assemble, as the distances between paired end reads are not accurate enough to localise the exact position of a gene *versus* a unique read. These types of phase variable genes are likely to be missing from or poorly assembled in genome sequences derived using short-read technologies. The degree to which this occurs will depend on the length of the duplicated region and may need long reads of 300+ for proper assembly to occur.

## 6. Analysis of SSRs by NGS *versus* PCR-Based Approaches

### 6.1. NGS Approach

Analysis of an SSR by NGS will provide data on the median repeat number for a population and the relative amounts of other repeat numbers when coverage is performed to a high read depth (Figure 2). Only reads containing a unique sequence from both sides of the repeat tract can be utilised for this type of analysis. The quality of the data will be influenced by the type of repeat unit, the length of the tract and the NGS methodology, as discussed above. There has, however, been no validation of such an

NGS approach to compare NGS distributions of tract length to either a PCR-based approach or phenotypic characterisation (e.g., with a reporter of expression). Thus, it is unclear if systematic errors are produced by the NGS methodology. High read depths and direct analyses of samples without further growth may, however, overcome some of the caveats associated with a PCR-based approach.

**Figure 2.** The use of next generation sequencing for the analysis of SSRs in isolates and bacterial populations. NGS, next generation sequencing. PV, phase variation.

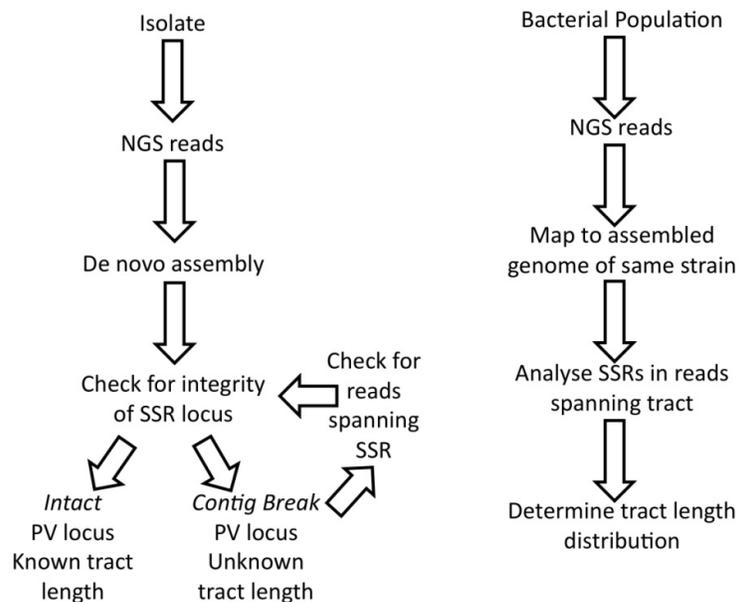


Figure 2 depicts a flow chart for the analysis of SSRs in bacteria. The analysis of one isolate aims to determine the number of SSRs present in the genome and the tract lengths of these SSRs. Variation in these tracts may occur depending on how the isolate is grown prior to sequencing. Analysis of a bacterial population may be utilized following the exposure of a bacterial population to a selective pressure. The goal of this analysis is to determine the numbers of cells with particular tract lengths for a given SSR.

### 6.2. PCR-Based Approach

A PCR-based approach can be used for direct analysis of a population without growth in order to determine the commonest tract length in a bacterial population. However, slippage in the SSR during PCR amplification means that this method is inaccurate for the analysis of the relative proportions of minor tract lengths. The alternative is to analyse a subset of isolates from the population following growth in non-selective conditions. This sampling approach has obvious limitations, as a sampling error is introduced, with the magnitude of this error influenced by the number of analysed isolates. The potential for alterations in SSRs during growth is another limitation of this type of data, and the degree of selection occurring in this stage should be considered and, where possible, controlled for using comparable samples [48].

SSR analysis is performed by PCR amplification of the tract using flanking primers and then either dideoxy-sequencing of the product or sizing by a GeneScan approach. The latter method utilises one

fluorescent primer and detection/sizing of fluorescent PCR products on an automated sequencing machine relative to a fluorescently-labelled DNA size standard. Repeat numbers are determined by comparison to a control sample of known repeat number (determined by sequencing) and with the same flanking sequence. This technique can detect differences of a single nucleotide, such that even mononucleotide repeat-mediated variation can be accurately quantified [22,48].

The minor advantages of this PCR-based approach are the high throughput and low cost. The major advantage is, as outlined above, the generation of information on the combined genotypes and expression states of phase variable genes, leading to the generation of phasotype data [22].

### 6.3. Use of NGS versus PCR to Analyse *C. jejuni* Phase Variable Genes

The genome of *C. jejuni* is AT-rich, but possesses a significant number of polyG/polyC SSR-associated genes with 12–27 in the genomes of *C. jejuni* strains, NCTC11168, RM1221, 81–176 and 81116 [22,49]. Only a subset of these genes, mainly those involved in capsule and lipooligosaccharide (LOS) biosynthesis, have been shown to be phase variable [50–57]. The lack of a mismatch repair system in *C. jejuni* translates into very high mutation rates at the polyG/C loci, ranging from  $4.2 \times 10^{-4}$  to  $4.1 \times 10^{-3}$ , up to 100 times the rates of equivalent tract lengths in *N. meningitidis* [22]. The presence of multiple SSR loci and the associated high rates of switching in *C. jejuni* could potentially produce thousands of phasotypes (e.g.,  $2^{27}$  for strain NCTC11168), a huge amount of phenotypic diversity in one population. The use of both NGS and PCR-based approaches is starting to reveal the impact of this variability on the commensal and disease behaviour of this major cause of gastroenteritis.

A study by Jerome *et al.* [49] investigated alterations in the phase variable genes of *C. jejuni* strain NCTC11168 following adaptation to the infection of mice, as a model of disease. Input and output populations were subject to NGS. The prevalence of different tract lengths for each of the 28 SSRs was determined by analysis of the read data, and relative expression states were determined. An average read depth of 107 was achieved for 22 of the tracts. A subset of six loci could not be analysed by NGS, due to the high similarity between the sequences of some of these loci, preventing mapping of reads to a specific locus. These loci were analysed by PCR amplification from colonies and Sanger sequencing. Twelve genes exhibited significant differences in the expression state during adaptation to the mouse intestinal tract. Overall, this data provided the first comprehensive analysis of the changes in the *C. jejuni* phasevariome during adaptation to an animal host.

A comparable study was performed by Bayliss *et al.* [22]. This study investigated the changes in the phase variable genes of *C. jejuni* during the colonization of four-week old chickens. Six genes were analysed by a PCR-based analysis of colonies sampled from input ( $n = 30$ ) and output ( $n = 150$  or  $30/\text{bird}$ ) populations. Three were found to exhibit alterations in the expression state. Due to the colony-based approach, this study also examined the phasotypes. A total of nine and sixteen phasotypes (out of a possible 64), with non-overlapping distributions, were detected in the input and output populations, respectively. A modelling approach indicated that changes in these distributions were not due to mutational drift alone, but that selection and/or population bottlenecks must have caused the changes in the population structure. This approach has now been extended to incorporate all

27 phase variable genes of *C. jejuni* strain NCTC11168, greatly extending the potential utility of this technique [58].

The complementary nature of the NGS and PCR-based approaches suggest that both will have utility in dissecting the contributions of the phasevariome to adaptation by bacterial species with multiple genes subject to SSR-mediated PV. The additional benefit of the PCR-based derivation of phasotypes has yet to be fully explored, but has the potential to detect interacting networks of genes and provides a critical input for the modelling of the behaviour of phase variable genes.

## 7. Case Studies Illustrating Global Approaches to SSR Analysis in Clinical Isolates

A combination of genomic and experimental studies can be utilized to study the contributions of phase variable genes to bacterial virulence. The four examples described below illustrate a variety of approaches to studying how SSRs influence the asymptomatic carriage and disease attributes of bacterial pathogens.

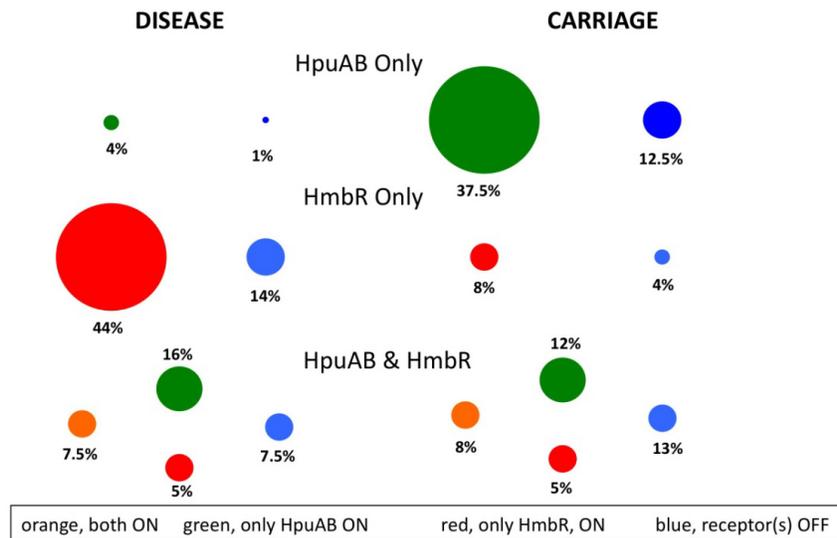
### 7.1. Comparison of Hb-Receptor Prevalence and Expression in Meningococcal Isolate Collections

In the pathogenic *Neisseria* species, *N. gonorrhoeae* and *N. meningitidis*, scavenging of iron, an essential co-factor of many biochemical processes in living cells, is achieved by iron-regulated, surface-expressed receptors that sequester iron molecules from host glycoproteins, such as transferrin (Tf) [59,60] and haemoglobin (Hb) [61,62]. Unlike the sequences of the well-studied Tf and lactoferrin (Lf) receptors, *tbpBA* and *lbpBA*, which lack hypermutable SSRs characteristic of phase variable genes, the haemoglobin receptors, *hpuAB* and *hmbR*, contain polyG tracts that control gene expression by ON-OFF switching [63]. Genetic epidemiology studies on separate meningococcal carriage and disease isolate collections supported a direct link between the presence of the Hb receptors and the virulence of *N. meningitidis*. These studies were performed by a combination of whole genome sequencing by Illumina sequencing and PCR-based amplification/Sanger dideoxy sequencing of products. In one study by Harrison *et al.* [64], genotypic analysis of 761 meningococcal isolates (314 disease, 447 carriage) from three separate collections detected an over-representation of *hmbR* in disease isolates, especially the hyper-invasive ST-4, ST-5, ST-8, ST-11, ST-18 and ST-32 lineages, *versus* a significantly higher proportion of *hmbR*-negative isolates in the carriage group. These results implied an important role for *hmbR* and, by extension, Hb-utilization during systemic infection. Tauseef *et al.* [40] extended this understanding of the association of Hb-utilization with meningococcal disease by using a PCR-based approach to detect the presence of either only HmbR or both Hb receptors in 99% of 214 disease isolates, but an under-representation of an HpuAB only phenotype in disease as compared to carriage isolates. Finally, Harrison *et al.* [65] analysed 218 genome sequences generated by Illumina sequencing to demonstrate that HmbR is mainly restricted to pathogenic *Neisseria*, whilst HpuAB is widely distributed in both commensal and pathogenic *Neisseria* species.

These studies revealed the prevalence of the Hb receptors (Figure 3). However, the SSRs could not be analysed, due to frequent contig breaks in the polyG tracts. Tauseef *et al.* [40] examined these tracts by a combination of dideoxy sequencing and GeneScan analysis of PCR products. These studies provided evidence of the importance of Hb receptor PV, showing that one or both receptors were in a PV-ON state (genotypic prediction of gene expression) in 91% of 90 disease isolates as compared to

71% of 103 carriage isolates. A subsequent analysis (see below) weakened this finding, as only 76% of disease isolates were found to have an Hb receptor in the ON state [48]. Differences in the strain composition of these surveys may explain these differing results.

**Figure 3.** Analysis of the distributions and expression states of phase variable haemoglobin receptors in disease and carriage isolates of meningococci.



Finally, whole genome sequencing, by a combination of whole genome shotgun sequencing and Illumina, was utilized to examine isolates from an accidental infection of a laboratory worker with an *hpuAB* and *hmbR*-positive mutator strain (*hpuAB*-OFF, *hmbR*-ON,  $\Delta mutS$ ) of *N. meningitidis* [66]. This infection led to a systemic infection of the bloodstream. Variants from this infection exhibited a loss of expression of *hmbR*, but a gain of expression of *hpuAB* via changes in the repeat tracts, thus demonstrating that SSR-associated genotypic switching of the Hb receptors occurs during infection. This switch in the homopolymeric SSRs of both genes could have been a consequence of the hypermutator phenotype, characteristic of mismatch repair-deficient ( $\Delta mutS$ ) strains [21,67], working in conjunction with the selective pressures encountered by the infecting strain during intravascular colonization of the host. The HpuAB receptor binds both Hb and Hb-haptoglobin complexes, whereas HmbR binds only Hb [68]. Perhaps the lack of free Hb molecules, due to sequestration by haptoglobin and other host Hb-binding glycoproteins, favoured the expression of *hpuAB* as disease progressed in this patient. This case provides an example of how hypermutable SSRs influence gene expression and, potentially, the virulence of *N. meningitidis*.

Two phase variable Hb-receptors, HpuAB and HmbR, are present in *Neisseria meningitidis* and are subject to translational PV, due to hypermutable polyG tracts. The majority of meningococcal strains can be separated into strains carrying both receptors or only one (*i.e.*, HpuAB only or HmbR only). This figure depicts the experimental data on the distribution and expression status of these receptors in disease ( $n = 80$ ) and carriage ( $n = 103$ ) isolates derived from Lucidarme *et al.* and Tauseef *et al.* [40,48], respectively. Note that the expression status was indeterminate in 1% of the disease isolates. The strain collections are not directly comparable, as serogroup B (MenB) strains predominate in the disease

isolates and serogroup Y (MenY) strains in the carriage isolates, due to a paucity of analyses of MenB carriage isolates and infrequent association of MenY strains with disease.

### 7.2. Analysis of Clinical Samples from Meningococcal Patients for Hb-Receptor Expression Status

A fundamental issue in the aforementioned genetic studies on SSR-mediated PV of meningococcal OMPs is the use of isolates that have been passaged in laboratory media. This transition from an *in vivo* to an *in vitro* environment could force adaptive changes, some of which may involve PV genes. While the studies described above utilized isolates that had been minimally passaged, the likelihood of obtaining “false” *in vivo* PV states was one issue needing clarification for proper interpretation of data generated from genetic analyses. The use of clinical specimens from meningitis or septicaemia sufferers, *i.e.*, cerebrospinal fluid (CSF) or blood samples, as a substitute for laboratory-cultured isolates was pursued by Lucidarme *et al.* [48]. A nested PCR protocol was devised to amplify target PV genes from blood and CSF samples. Using this method, it was shown that minimally passaged isolates are good predictors of the *in vivo* PV status of a gene, as the data from isolates matched results from cognate clinical specimens. The genotypic predictions of *in vivo* gene expression were also largely consistent with data from phenotypic analyses. However, genetic studies often analyse a single isolate from a clinical sample and doubts have arisen as to whether data from a single isolate is representative of the entire population of meningococcal cells within a patient. In the absence of clinical specimens to confirm data generated from the analysis of a single isolate, the analysis of multiple isolates will be required in order to determine the *in vivo* ratios of PV states (e.g., relative amounts of ON and OFF states for a phase variable gene). Employment of advanced NGS methods to directly analyse clinical samples may also facilitate the generation of this type of data.

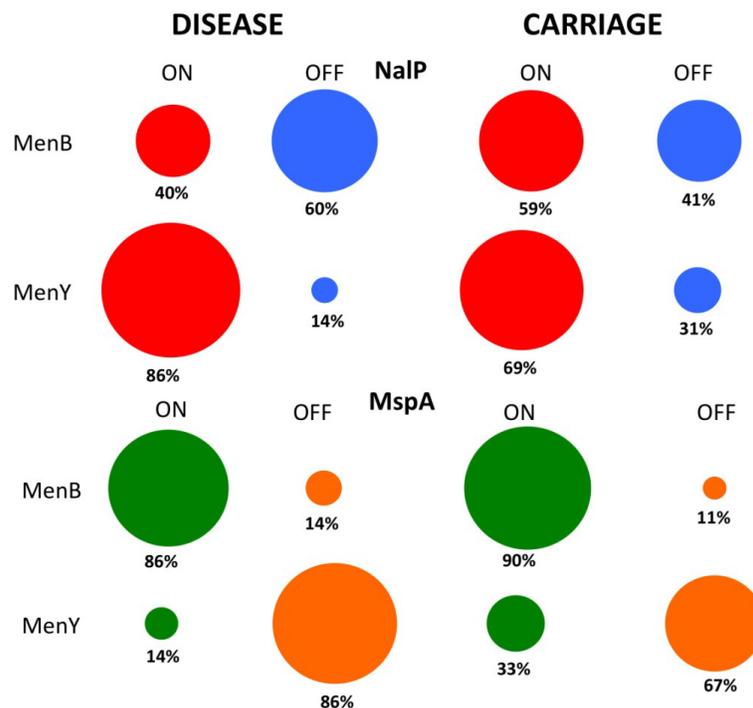
### 7.3. Phase Variable Autotransporters in Clinical Meningococcal Isolates

Further genetic epidemiology studies on other PV genes of *N. meningitidis* have utilized genome libraries, in addition to conventional PCR and dideoxy sequencing methods, to produce interesting data that may impact on vaccine design and development. For example, a recent study by Oldfield *et al.* [69] on the autotransporters, *mshA* and *nalP*, showed high prevalences and differing expression states of either or both genes in carriage ( $n = 127$ ) as compared to disease ( $n = 514$ ) isolates (Figure 4). The epidemiological data on these autotransporters in disease isolates was derived by bioinformatic studies of the 500+ meningococcal genomes present in the Meningitis Research Foundation Meningococcus Genome Library. These sequences were generated using an Illumina system, and each genome is fragmented into 150–250 contigs. As expected, use of these incompletely assembled genomes translated into an inability to obtain all of the required sequences, such as for assigning all *nalP*-negative isolates to one of four *nalP* deletion classes. Crucially, SSR tract lengths were obtained from 500 *mshA* and 430 *nalP* genes from these genome sequences of invasive isolates. This high level of coverage of these phase variable genes and their associated SSRs demonstrates the potential for using these genome libraries to search for SSRs and to derive putative expression states of phase variable virulence factors.

The study by Oldfield and co-workers uncovered a potential bias for the PV-ON state of *mshA* in serogroup B strains (86%) and *nalP* in serogroup Y strains (86%), suggesting an important role for *mshA*

or *nalP* during meningococcal disease, due to serogroup B and Y strains, respectively (Figure 4). While the biological significance of *mspA* expression and activity is still unclear in the meningococcus [70,71], *nalP* PV was recently found to exert a significant impact on biofilm formation and complement resistance in the meningococcus. NalP proteolytically cleaves the eDNA-binding 4CMenB vaccine target *Neisseria* heparin binding antigen (NHBA) [72], thereby reducing the eDNA-binding capacity of NalP-expressing meningococcal cells [73]. Recently, NalP was also shown to cleave C3, thereby increasing the serum resistance of meningococci [74]. Having high amounts of NalP on the meningococcal surface, as a consequence of the PV-ON state of *nalP*, may therefore have a profound effect on biofilm formation and the survival of complement-mediated killing of these serogroup Y strains, resulting in changes in their ability to persist in hosts and to cause disease. Furthermore, the expression of NalP in these serogroup Y strains and 40% of invasive MenB strains could impact on vaccine efficacy by reducing the surface expression of NHBA. However, Oldfield *et al.* maintain that the apparent lack of *nalP* expression in the majority of serogroup B strains, either by complete deletion of the gene or by PV, should not affect the use of 4CMenB for the prevention of serogroup B disease, as there is evidence that NalP deletion mutants are still susceptible to bactericidal NHBA antibodies.

**Figure 4.** Contrasts in the expression states of two phase variable autotransporters in serogroup B and Y strains of *Neisseria meningitidis*.



The data for Figure 4 was derived from an analysis by Oldfield *et al.* [69] using a database of genomic sequences to determine expression states in disease strains (312 MenB and 73 MenY isolates for NalP; 381 MenB and 73 MenY isolates for MspA) and direct experimental analysis for carriage isolates (17 MenB and 52 MenY isolates for NalP; 19 MenB and 52 MenY isolates for MspA). Both of these autotransporters are subject to translational PV, due to a polyC tract.

#### 7.4. Alterations in Outer Membrane Protein of *Neisseria meningitidis* and *Haemophilus influenzae*

Key to increased intra-host persistence of both commensal and pathogenic bacterial strains in specific niches of their hosts is the generation of a phenotypically-diverse population of cells via PV. Evidence for this assertion has come from studies of phase variable surface proteins in *N. meningitidis* and *H. influenzae* that have utilized PCR-based approaches to analyse multiple isolates from healthy or diseased individuals. For example, the expression of the immunogenic high-molecular-weight adhesins of *H. influenzae*, *hmw1A* and *hmw2A* is transcriptionally controlled by a heptanucleotide tract [75]. During chronic infections of the respiratory tract by *H. influenzae*, the expression levels of HMW1 and HMW2 were found to decrease as the bacteria persisted within a host. These decreases in protein expression correlated with increases in the lengths of the repeat tracts within the promoters of these genes. This decreased expression was correlated with elevated levels of anti-HMW1 and anti-HMW2 antibody titres [76]. Data from this study supports the theory that a PV-mediated reduction in gene expression allows for immune avoidance, *in vivo*. Similarly, longitudinal isolates obtained from healthy volunteers in a six-month meningococcal carriage study [77] were examined for differences in the expression of phase variable OMPs. These isolates differed in the SSR tract lengths and expression status of these OMPs with a trend of switching towards phenotypes with lower expression levels for combinations of eight phase variable surface proteins during persistent carriage [78]. The study by Alamro *et al.* analysed up to 20 isolates per time point, thereby, providing a thorough assessment of the changes in the meningococcal phenotypes and phasevariomes during carriage. The association between immune avoidance and phase variable expression state during meningococcal carriage, and indeed disease, is yet to be expressly confirmed, but the selection for lower expression states of some meningococcal OMPs is most likely due to the activity of antibodies targeting these OMPs, either individually or synergistically.

## 8. Summary

The first stage of pathogenomics has placed the emphasis on comparisons between bacterial isolates from outbreaks and epidemics, disease *versus* carriage, different isolation sites within a host and alternate types of disease presentation. These samples have begun to be exploited to determine the influence of phase variable genes on the pathogenesis and commensal behaviour of a range of bacterial species. However, the population diversity inherent in mechanisms of localized hypermutation requires a population-based approach. Future studies will place a greater emphasis on isolation and analysis of total DNA from a particular sampling point or genomic studies of multiple isolates from these samples. These studies are required to dissect the contributions of the phasevariome and phenotypes to bacterial pathogenesis. The utility of NGS approaches will assist in studies of this type of population diversity, leading to rapid and detailed analyses of the dynamic changes occurring in bacterial genomes during infections.

## Author Contributions

Both authors contributed to the writing of this manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References and Notes

1. Deitsch, K.W.; Lukehart, S.A.; Stringer, J.R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat. Rev. Microbiol.* **2009**, *7*, 493–503.
2. Moxon, E.R.; Bayliss, C.D.; Hood, D.W. Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Ann. Rev. Genet.* **2007**, *40*, 307–333.
3. Van der Woude, M.W.; Baumler, A.J. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **2004**, *17*, 581–611.
4. Bayliss, C.D. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. *FEMS Microbiol. Rev.* **2009**, *33*, 504–520.
5. Hall, L.M.; Henderson-Begg, S.K. Hypermutable bacteria isolated from humans—a critical analysis. *Microbiology* **2006**, *152*, 2505–2514.
6. Drake, J.W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7160–7164.
7. Bayliss, C.D.; Moxon, E.R. Repeats and variation in pathogen selection. In *The Implicit Genome*, Caporale, L.H., Ed.; Oxford University Press: Oxford, UK, 2005; pp. 54–76.
8. Moxon, E.R.; Rainey, P.B.; Nowak, M.A.; Lenski, R. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **1994**, *4*, 24–33.
9. Cahoon, L.A.; Seifert, H.S. Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in *Neisseria gonorrhoeae*. *PLoS Pathog.* **2013**, *9*, e1003074.
10. Yogev, D.; Browning, G.; Wise, K.S. Mechanisms of surface variation. In *Molecular biology and pathogenicity of mycoplasma*, Razin, S., Herrmann, R., Eds.; Kluwer Academic/Plenum Publishers: New York, NY, USA, 2002; pp. 417–443.
11. Cahoon, L.A.; Manthei, K.A.; Rotman, E.; Keck, J.L.; Seifert, H.S. *Neisseria gonorrhoeae* RecQ helicase HRDC domains are essential for efficient binding and unwinding of the *pilE* guanine quartet structure required for pilin antigenic variation. *J. Bacteriol.* **2013**, *195*, 2255–2261.
12. Criss, A.K.; Kline, K.A.; Seifert, H.S. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. *Mol. Microbiol.* **2005**, *58*, 510–519.
13. Andrewes, F.W. Studies in group-agglutination I. The salmonella group and its antigenic structure. *J. Path. Bacteriol.* **1922**, *25*, 505–521.
14. Henderson, I.R.; Owen, P.; Nataro, J.P. Molecular switches—the ON and OFF of bacterial phase variation. *Mol. Micro.* **1999**, *33*, 919–932.
15. De Bolle, X.; Bayliss, C.D.; Field, D.; van de Ven, T.; Saunders, N.J.; Hood, D.W.; Moxon, E.R. The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol. Microbiol.* **2000**, *35*, 211–222.
16. Hendrixson, D.R. Restoration of flagellar biosynthesis by varied mutational events in *Campylobacter jejuni*. *Mol. Microbiol.* **2008**, *70*, 519–536.

17. Tran, H.T.; Keen, J.D.; Krickler, M.; Resnick, M.A.; Gordenin, D.A. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell. Biol.* **1997**, *17*, 2859–2865.
18. Lin, W.H.; Kussell E. Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Res.* **2012**, *40*, 2399–2413.
19. Gawel, D.; Jonczyk, P.; Bialoskorska, M.; Schaaper, R.M.; Fijalkowska, I.J. Asymmetry of frameshift mutagenesis during leading and lagging-strand replication in *Escherichia coli*. *Mutat. Res.* **2002**, *501*, 129–136.
20. Strauss, B.S.; Sagher, D.; Acharya S. Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. *Nucleic Acids Res.* **1997**, *25*, 806–813.
21. Martin, P.; Sun, L.; Hood, D.W.; Moxon, E.R. Involvement of genes of genome maintenance in the regulation of phase variation frequencies in *Neisseria meningitidis*. *Microbiology* **2004**, *150*, 3001–3012.
22. Bayliss, C.D.; Bidmos, F.A.; Anjum, A.; Manchev, V.T.; Richards, R.L.; Grossier, J.P.; Wooldridge, K.G.; Ketley, J.M.; Barrow, P.A.; Jones, M.A.; *et al.* Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. *Nuc. Acids Res.* **2012**, 5876–5889.
23. Richardson, A.R.; Yu, Z.; Popovic, T.; Stojiljkovic, I. Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6103–6107.
24. Martin, P.; Makepeace, K.; Hill, S.A.; Hood, D.W.; Moxon, E.R. Microsatellite instability regulates transcription factor binding and gene expression. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3800–3804.
25. Zaleski, P.; Wojciechowski, M.; Piekarowicz, A. The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiology* **2005**, *151*, 3361–3369.
26. Sreenu, V.B.; Ranjitkumar, G.; Swaminathan, S.; Priya, S.; Bose, B.; Pavan, M.N.; Thanu, G.; Nagaraju, J.; Nagarajaram, H.A. MICAS: A fully automated web server for microsatellite extraction and analysis from prokaryote and viral genomic sequences. *Appl. Bioinformatics* **2003**, *2*, 165–168.
27. Hood, D.W.; Deadman, M.E.; Allen, T.; Masoud, H.; Martin, A.; Brisson, J.R.; Fleischmann, R.; Venter, J.C.; Richards, J.C.; Moxon, E.R. Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis. *Mol. Microbiol.* **1996**, *22*, 951–965.
28. Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Davies, R.M.; Feltwell, T.; Holroyd, S.; *et al.* The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **2000**, *403*, 665–668.
29. Saunders, N.J.; Jeffries, A.C.; Peden, J.F.; Hood, D.W.; Tettelin, H.; Rappouli, R.; Moxon, E.R. Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol. Micro.* **2000**, *37*, 207–215.
30. Saunders, N.J.; Peden, J.; Hood, D.; Moxon, E.R. Simple sequence repeats in the *Helicobacter pylori* genome. *Mol. Microbiol.* **1998**, *27*, 1091–1098.

31. Martin, P.; van de Ven, T.; Mouchel, N.; Jeffries, A.C.; Hood, D.W.; Moxon, E.R. Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: Evidence for a novel mechanism of phase variation. *Mol. Microbiol.* **2003**, *50*, 245–257.
32. Power, P.M.; Sweetman, W.A.; Gallacher, N.J.; Woodhall, M.R.; Kumar, G.A.; Moxon, E.R.; Hood, D.W. Simple sequence repeats in *Haemophilus influenzae*. *Infect. Genet. Evol.* **2009**, *9*, 216–228.
33. Snyder, L.A.S.; Butcher, S.A.; Saunders, N.J. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp. *Microbiology* **2001**, *147*, 2321–2332.
34. Dixon, K.; Bayliss, C.D.; Makepeace, K.; Moxon, E.R.; Hood, D.W. Identification of the functional initiation codons of a phase-variable gene of *Haemophilus influenzae*, *lic2A*, with the potential for differential expression. *J. Bacteriol.* **2007**, *189*, 511–521.
35. Snyder, L.A.; Loman, N.J.; Linton, J.D.; Langdon, R.R.; Weinstock, G.M.; Wren, B.W.; Pallen, M.J. Simple sequence repeats in *Helicobacter canadensis* and their role in phase variable expression and C-terminal sequence switching. *BMC Genomics* **2010**, *11*, doi: 10.1186/1471-2164-11-67.
36. Van der Ende, A.; Hopman, C.T.; Dankert J. Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. *Infect. Immun.* **2000**, *68*, 6685–6690.
37. Van Ham, S.M.; van Alphen, L.; Mooi, F.R.; van Putten, J.P. Phase variation of *H. influenzae* fimbriae: Transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **1993**, *73*, 1187–1196.
38. Metruccio, M.M.; Pigozzi, E.; Roncarati, D.; Berlanda Scorza, F.; Norais, N.; Hill, S.A.; Scarlato, V.; Delany, I. A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements. *PLoS Pathog.* **2009**, *5*, e1000710.
39. Carson, S.D.; Stone B.; Beucher, M.; Fu, J.; Sparling, P.F. Phase variation of the gonococcal siderophore receptor FetA. *Mol Microbiol* **2000**, *36*, 585–593.
40. Tauseef, I.; Harrison, O.B.; Wooldridge, K.G.; Feavers, I.M.; Neal, K.R.; Gray, S.J.; Kriz, P.; Turner, D.P.; Ala'Aldeen, D.A.; Maiden, M.C.; *et al.* Influence of the combination and phase variation status of the haemoglobin receptors HmbR and HpuAB on meningococcal virulence. *Microbiology* **2011**, *157*, 1446–1456.
41. Power, P.M.; Roddam, L.F.; Rutter, K.; Fitzpatrick, S.Z.; Srikhanta, Y.N.; Jennings, M.P. Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol. Microbiol.* **2003**, *49*, 833–847.
42. Aas, F.E.; Vik, A.; Vedde, J.; Koomey, M.; Egge-Jacobsen, W. *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol. Microbiol.* **2007**, *65*, 607–624.
43. Churbanov, A.; Ryan, R.; Hasan, N.; Bailey, D.; Chen H.; Milligan, B.; Houde, P. HighSSR: high-throughput SSR characterization and locus development from next-gen sequencing data. *Bioinformatics* **2012**, *28*, 2797–2803.
44. Fondon, J.W., III; Martin, A.; Richards, S.; Gibbs, R.A.; Mittelman, D. Analysis of microsatellite variation in *Drosophila melanogaster* with population-scale genome sequencing. *PLoS One* **2012**, *7*, e33036.

45. Bayliss, C.D. A preliminary analysis of genome assemblies of meningococcal carriage isolates. **2014**, in preperation.
46. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **2012**, *13*, doi: 10.1186/1471-2164-13-341.
47. Bilek, N.; Ison, C.A.; Spratt, B.G. Relative contributions of recombination and mutation to the diversification of the *opa* gene repertoire of *Neisseria gonorrhoeae*. *J. Bacteriol.* **2009**, *191*, 1878–1890.
48. Lucidarme, J.; Findlow, J.; Chan, H.; Feavers, I.M.; Gray, S.J.; Kaczmarek, E.B.; Parkhill, J.; Bai, X.; Borrow, R.; Bayliss, C.D. The distribution and “*in vivo*” phase variation status of haemoglobin receptors in invasive meningococcal serogroup B disease: Genotypic and phenotypic analysis. *PLoS One* **2013**, *8*, e76932.
49. Jerome, J.P.; Bell, J.A.; Plovianich-Jones, A.E.; Barrick, J.E.; Brown, C.T.; Mansfield, L.S. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLoS One* **2011**, *6*, e16399.
50. Caldwell, M.B.; Guerry, P.; Lee, E.C.; Burans, J.P.; Walker, R.I. Reversible expression of flagella in *Campylobacter jejuni*. *Infect. Immun.* **1985**, *50*, 941–943.
51. Linton, D.; Gilbert, M.; Hitchen, P.G.; Dell, A.; Morris, H.R.; Wakarchuk, W.W.; Gregson, N.A.; Wren, B.W. Phase variation of a beta-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. *Mol. Microbiol.* **2000**, *37*, 501–514.
52. Bacon, D.J.; Szymanski, C.M.; Burr, D.H.; Silver, R.P.; Alm, R.A.; Guerry, P. A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81–176. *Mol. Microbiol.* **2001**, *40*, 769–777.
53. Guerry, P.; Szymanski, C.M.; Prendergast, M.M.; Hickey, T.E.; Ewing, C.P.; Pattarini, D.L.; Moran, A.P. Phase variation of *Campylobacter jejuni* 81–176 lipooligosaccharide affects ganglioside mimicry and invasiveness *in vitro*. *Infect. Immun.* **2002**, *70*, 787–793.
54. Karlyshev, A.V.; Champion, O.L.; Churcher, C.; Brisson, J.R.; Jarrell, H.C.; Gilbert, M.; Brochu, D.; St. Michael, F.; Li, J.; Wakarchuk, W.W.; *et al.* Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. *Mol. Microbiol.* **2005**, *55*, 90–103.
55. Szymanski, C.M.; Michael, F.S.; Jarrell, H.C.; Li, J.; Gilbert, M.; Larocque, S.; Vinogradov, E.; Brisson, J.R.; Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.* **2003**, *278*, 24509–24520.
56. Prendergast, M.M.; Tribble, D.R.; Baqar, S.; Scott, D.A.; Ferris, J.A.; Walker, R.I.; Moran, A.P. *In vivo* phase variation and serologic response to lipooligosaccharide of *Campylobacter jejuni* in experimental human infection. *Infect. Immun.* **2004**, *72*, 916–922.
57. Ashgar, S.S.; Oldfield, N.J.; Wooldridge, K.G.; Jones, M.A.; Irving, G.J.; Turner, D.P.; Ala’Aldeen, D.A. CapA, an autotransporter protein of *Campylobacter jejuni*, mediates association with human epithelial cells and colonization of the chicken gut. *J. Bacteriol.* **2007**, *189*, 1856–1865.

58. Lango-Scholey, L.; Aidley, J.; Akinyemi, N.M.; Woodacre, A.; Jones, M.A.; Bayliss, C.D. Development of a multiplex GeneScan assay for analysis of the phase variable genes of *Campylobacter jejuni* strain NCTC11168. **2014**, In preparation.
59. Schryvers, A.B.; Gonzalez, G.C. Receptors for transferrin in pathogenic bacteria are specific for the host's protein. *Can. J. Microbiol.* **1990**, *36*, 145–147.
60. Ala'Aldeen, D.A.; Borriello, S.P. The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* **1996**, *14*, 49–53.
61. Lewis, L.A.; Gray, E.; Wang, Y.P.; Roe, B.A.; Dyer, D.W. Molecular characterization of hpuAB, the haemoglobin-haptoglobin-utilization operon of *Neisseria meningitidis*. *Mol. Microbiol.* **1997**, *23*, 737–749.
62. Stojiljkovic, I.; Larson, J.; Hwa, V.; Anic, S.; So, M. HmbR outer membrane receptors of pathogenic *Neisseria* spp.: Iron-regulated, hemoglobin-binding proteins with a high level of primary structure conservation. *J. Bacteriol.* **1996**, *178*, 4670–4678.
63. Lewis, L.A.; Gipson, M.; Hartman, K.; Ownbey, T.; Vaughn, J.; Dyer, D.W. Phase variation of HpuAB and HmbR, two distinct haemoglobin receptors of *Neisseria meningitidis* DN2. *Mol. Microbiol.* **1999**, *32*, 977–989.
64. Harrison, O.B.; Evans, N.J.; Blair, J.M.; Grimes, H.S.; Tinsley, C.R.; Nassif, X.; Kriz, P.; Ure, R.; Gray, S.J.; Derrick, J.P.; *et al.* Epidemiological evidence for the role of the hemoglobin receptor, hmbR, in meningococcal virulence. *J. Infect. Dis.* **2009**, *200*, 94–98.
65. Harrison, O.B.; Bennett, J.S.; Derrick, J.P.; Maiden, M.C.; Bayliss, C.D. Distribution and diversity of the haemoglobin-haptoglobin iron-acquisition systems in pathogenic and non-pathogenic *Neisseria*. *Microbiology* **2013**, *159*, 1920–1930.
66. Omer, H.; Rose, G.; Jolley, K.A.; Frapy, E.; Zahar, J.R.; Maiden, M.C.; Bentley, S.D.; Tinsley, C.R.; Nassif, X.; Bille, E. Genotypic and phenotypic modifications of *Neisseria meningitidis* after an accidental human passage. *PLoS One* **2011**, *6*, e17145.
67. Richardson, A.R.; Stojiljkovic, I. Mismatch repair and the regulation of phase variation in *Neisseria meningitidis*. *Mol. Micro.* **2001**, *40*, 645–655.
68. Perkins-Balding, D.; Ratliff-Griffin, M.; Stojiljkovic I. Iron transport systems in *Neisseria meningitidis*. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 154–171.
69. Oldfield, N.J.; Matar, S.; Bidmos, F.A.; Alamro, M.; Neal, K.R.; Turner, D.P.; Bayliss, C.D.; Ala'Aldeen, A.A. Prevalence and Phase Variable Expression Status of Two Autotransporters, NalP and MspA, in Carriage and Disease Isolates of *Neisseria meningitidis*. *PLoS One* **2013**, *8*, e69746.
70. Turner, D.P.; Marietou, A.G.; Johnston, L.; Ho, K.K.; Rogers, A.J.; Wooldridge, K.G.; Ala'Aldeen, D.A. Characterization of MspA, an immunogenic autotransporter protein that mediates adhesion to epithelial and endothelial cells in *Neisseria meningitidis*. *Infect. Immun.* **2006**, *74*, 2957–2964.
71. Van Ulsen, P.; Adler, B.; Fassler, P.; Gilbert, M.; van Schilfgaarde, M.; van der Ley, P.; van Alphen, L.; Tommassen, J. A novel phase-variable autotransporter serine protease, AusI, of *Neisseria meningitidis*. *Microbes Infect.* **2006**, *8*, 2088–2097.

72. Serruto, D.; Spadafina, T.; Ciucchi, L.; Lewis, L.A.; Ram, S.; Tontini, M.; Santini, L.; Biolchi, A.; Seib, K.L.; Giuliani, M.M.; *et al.* *Neisseria meningitidis* GNA2132, a heparin-binding protein that induces protective immunity in humans. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3770–3775.
73. Arenas, J.; Nijland, R.; Rodriguez, F.J.; Bosma, T.N.; Tommassen J. Involvement of three meningococcal surface-exposed proteins, the heparin-binding protein NhbA, the alpha-peptide of IgA protease and the autotransporter protease NalP, in initiation of biofilm formation. *Mol. Microbiol.* **2013**, *87*, 254–268.
74. Del Tordello, E.; Vacca, I.; Ram, S.; Rappuoli, R.; Serruto, D. *Neisseria meningitidis* NalP cleaves human complement C3, facilitating degradation of C3b and survival in human serum. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 427–432.
75. Dawid, S.; Barenkamp, S.J.; St. Geme, J.W., III. Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1077–1082.
76. Cholon, D.M.; Cutter, D.; Richardson, S.K.; Sethi, S.; Murphy, T.F.; Look, D.C.; St. Geme, J.W., III. Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. *Infect. Immun.* **2008**, *76*, 4463–4468.
77. Bidmos, F.A.; Neal, K.R.; Oldfield, N.J.; Turner, D.P.; Ala'Aldeen, D.A.; Bayliss, C.D. Persistence, replacement, and rapid clonal expansion of meningococcal carriage isolates in a 2008 university student cohort. *J. Clin. Microbiol.* **2011**, *49*, 506–512.
78. Alamro, M.; Bidmos, F.A.; Chan, H.; Oldfield, N.J.; Newton, E.; Bai, X.; Aidley, J.; Care, R.; Mattick, C.; Turner, D.P.J.; *et al.* Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. *Infect. Immun.* **2014**, Submitted.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).