Evaluation of the simple sequence repeats to work as a phase variation with the Neisseria meningitidis genome

Abstract

Background: The most crucial mechanism of genetic variation in *N. meningitidis* is the slipped strand mispairing, this mechanism generates Phase variation using simple sequence repeat (SSR) and is commonly used by the *N. meningitidis* to escape the immune system despite its function in eradicating the pathogenic and commensal bacteria. Some of simple sequence repeats (SSRs) that located within the genome works as phase variation while other SSRs have no role in generating phase variation mechanisms. Therefore,

Aim: the main goal of the current in silico study was to detect the probability of SSR to enroll with phase variation for the entire *N. meningitidis* genome.

Methods: Different criteria were used to judge SSR as it works in phase variation and these criteria were taken from the current literature. These criteria involve the Z score value of the synonymous shuffling model and Markov model of SSR, the position of SSR in the gene or the promoter, instability and polymorphism of SSR in genomes of different strains, and the length of SSR.

Results: The positive Z score value of SSR, SSR being variable in length among genomes of different strains and SSR location in 3 prime end of a gene or in the promoter indicates that the SSR generates phase variation in a particular gene.

Conclusion: 67 out of 327 putative phase variable genes located on *the N. meningitidis* genome were determined to fit these criteria. We assume, therefore, that SSR in these genes may be connected with phase variation mechanism. We recommend that experimental evidence should be generated to confirm these findings.

KeyWords: ***N. meningitidis*** ; **phase variation**; **simple sequence repeats**

Introduction

Neisseria meningitidis colonies the upper respiratory tract in form of carriage isolates while disease isolates enter the bloodstream and cause meningitis or septicemia with high mortality and morbidity [1- 2]. Neisseria meningitidis spreads quickly among South America, Europe, and the United Kingdom especially the serogroup type MenW, and this has increased the mortality due to *Neisseria meningitidis* in these countries [3]. Different strategies arise in the commensal and pathogenic bacteria to escape the immune attack of the host, one of them is phase variation [4]. Phase variation may refer to change in methylation status or hypermutation in SSR reversibly [5]. Hypermutation in SSR occurs in DNA replication which leads to gain or loss in repeat patterns [6]. Neisseria meningitidis possess phase variation through the presence of SSR in the genic and intergenic regions in a process called slipped strand mispairing [4]. Changing the SSR in the genic regions leads to the frameshift mutation or truncated the product of the gene in case of this changing codes for stop codon. The SSR change in the promoter region leads to an increase or decrease in the level of expression [4]. In the phase variation process, the gene can switch to an ON or OFF state and this helps Neisseria meningitidis to adapt to the severe stress conditions [1]. In Neisseria meningitidis, the phase variation mainly occurs in the Opa, NadA, PorA, pilli, and capsule and enhances the changing of the proteins located on the outer membrane to resist the immune system [7]. However, it has been

observed that phase variation occurs in NadA higher than in other genes. NadA gene was found in 22.3% of the MenB isolates and found in a high percentage in MenW isolates [8]. The MenW isolates cause invasive meningococcal disease in most countries [3]. In phase variation of Neisseria meningitidis, changing lipooligosaccharides leads to resisting antibodies and neutrophil cells, while the phase variation in polysaccharide capsule mainly leads to resisting antibodies [9]. Genome sequencing of MC58 and Z2491 helps in the detection of the presence of SSR in whole genic and intergenic regions of Neisseria meningitidis [10]. Our study intended to identify putative phase variable genes from 12 strains, relying on criteria provided from the current literature. This in silico approach is going to predict if each SSR, carried on genic or intergenic regions, may enroll in the process of On/OFF state, increase or decrease the gene expression.

Material and methods

The criteria were used for predicting phase variation genes relying on two different methods which are comparative and probabilistic analysis approaches [11- 3]. The study was approved by the Baghdad University Ethics Committee, and written informed consent was obtained from all volunteers.

These criteria are listed below:

Detection SSR stability and polymorphism [11-12-13- 4].

MICAS program was used to detect SSR in the whole genome of an invasive strain with the ID: 20026. This strain was collected from BIGSdp [14].

The id-20026 was selected out of 500 invasive isolates depending on two criteria. It contained the highest number of phase variable genes with G or C repeat tract (168) and the lowest number of this tract located at the end of contig (10). The length of simple sequence repeats was selected with specific cut off (homopolymeric with 7 bp or more for G or C, homopolymeric 8 bp or more for A or T, dinucleotide and trinucleotide with four or more copies, tetranucleotide and pentanucleotide with three or more copies and the motif with (5-9pb) with three copies or more). The Artemis program was utilized to detect SSR (have been identified using MICAS program) from seven carriage isolates belonging to different clonal complex (CC1157, CC167, CC174, CC23-ST1655, CC23, CC60, and Serob-N119) which was collected from MRF Meningococcus Genome Library. Then BLAST search was used to find the orthologous genes in different invasive strains (N417, Nng63, E934, MC58, and FAM18) that were collected from MRF Meningococcus Genome Library. The quality of data was checked thereby the putative phase variable genes with poor alignment were excluded, Moreover, The SSR located at the end of the contigs was also excluded. Finally, polymorphism and stability were detected for each particular SSR within the 12 compared strains.

Z score detection by Markov model [11-13-15].

Markov model was used to estimate the Z score for SSR in phase variation genes through calculating expected value. For example, the expected value for ABCDE is calculated using the following formula:

Markov chain is used to calculate the Z score for each SSR. The method counted the expected value as a null model. The following example is using the word “ABCDE”, in order to provide an explanation about how this formula works:

$$E\big(C(ABCDE)\big) = \frac{C(ABCD)C(BCDE)}{C(BCD)}$$

Independent observations on a model that is a sum of C(ABC), can be used with the binomial distribution to calculate variance from the following equation;

$$\sigma = \frac{C(ABCD)C(BCDE)}{C(ABC)C(BCD)} \ldots\ldots\ldots\ldots\ldots 2$$

The difference between expected and observed was quantified by calculating the Z score by using the formula;

$$Z = \frac{expected - observed}{\sigma^2} \ldots\ldots\ldots\ldots 3$$

If the value of the Z score was positive they are over-represented or have high density. The Z score for the entire repeat tract in id-20026 invasive isolate was identified.

Z score detection by Synonymous shuffling model

A synonymous shuffling model was used to calculate the z score [16]. SSR was counted in original, global shuffling thereby a previous formula of Z score was applied to identify the SSR for expected and observed values. The previous formula of the Z score calculated the probability of repeat tracts being phase variable. If the value of the Z score was positive they are over-represented or have high density. The Z score for the entire repeat tract in id-20026 invasive isolate was also identified.

Determination of frameshift of target sequence holding SSR [17]

The Mega program was utilized to identify the frameshift of each target sequence holding SSR in each strain. The sequences of each putative gene for all 12 genomes were aligned, then the start and stop codon for all strains were highlighted. The position, number of repeats tract, and stability within the 12 strains for each SSR were identified. Then, the DNA sequences were translated into amino acids. After that, we looked for stop codons in the whole amino acids. If a stop codon was found at the end of the protein and there were no other stop codons in the rest of the protein, then the gene frame is (ON). On another hand, if there were stop codons in other locations in protein and the protein is truncated then we looked for the case that truncation of protein was in due to the repeat tract or indels. We repeatedly changed the tract's frame and translated it. If stop codons were still present, the process was repeated thrice. Finally, DNA was aligned again and indels were searched for. If the stop codon disappeared from the middle and shifted to the 3' end, we can say the frame was (OFF) due to the repeat tract.

Detection SSR location within the gene in *N. meningitidis* genome [18- 4- 17]

The location of the repeat tract was checked within the entire putative phase variable gene. If the repeat tracts are positioned towards the 3' end of the gene, they are less likely to alter gene expression.

Determination of Location SSR between -10 and -35 patterns of a promoter of an intergenic region of different isolates [19-20]

BPROM program - Prediction of bacterial promoters used to predict the position of -10 and -35. Then we looked to identify the location of repeat patterns within -10 and -35 patterns of promoters.

KEGG refers to Kyoto Encyclopedia of Genes and Genomes was used to detect function schemes for each putative phase variable gene.

Results

All the types of SSR were extracted from whole genome using MICAS program with 45 types. Relying on the length of repeat tracts (See cut off for the length of repeat in material and methods), we detected 200, 216, 213, 208, 200, 2014, and 217 putative phase variable genes in genomes of Serob-N119, CC60, CC23 CC23-ST1655, CC23, CC174, CC167, and CC1157 isolates respectively. In addition, we detected 57 new putative phase variable genes in invasive isolates (N417, Nng63, E934, MC58, and FAM18) (Table 1).

Table 1: The overall putative phase variable genes in invasive isolates. Repeat tract: is the type of repeat tract. Gene: is the name of gene that contained the repeat tracts

repeat tract	Gene	repeat tract	Gene	repeat tract	Gene
CAAT3	No mutch	GC6	NEIS1634-NMB1716	GGC(4-5)	NMB0870
AGTC3	No mutch		NEIS1634-NMB1716		NEIS1176-NMB1036
CAAC3	No mutch		NMB0800NIE0752		NMB1270
AGCC3	N73-00567		NEIS1133		NMB1461
AAGC3	DOWNSTREAM NMB0311		NMB2061-NEIS2042		NMB1363-NEIS1298
TAAA3	DOWNSTREAM NIES0182		NIES2000		NMB1590-NEIS1512 GGC5
GAAA3	NMB1077		NES1742		NMB1614-NEIS1535
TTCC3	DOWNSTREAM LOIP		NEIS1831-NMB0339		NMB111
GGCA3	No mutch		NEIS1903		NMB0950
TGCG3	No mutch		NMB0195-NEIS0186 6CG		NMB0460
CTTCT3	DOWNSTREAM NEIS0612-NMB0663		NEIS0001-NMB0017 6GC		NMB2005
CCCAA3	No mutch				NMB0385
CAAAT3	No mutch		NEIS0315-NMB1908		NMB1947
GCCAA3	DOWNSTREAM N73-00200		NEIS0343-NMB1876		NEIS0185-NMB0194 GGC5
ATAACAA A3	No mutch		NEIS0835-NMB0895		NEIS0185-NMB0194
CAAACAA 3	No mutch	GGAC3	No mutch		NMB1818
TAGGCT3	NEIS1297 NMB1362	CG6	NEIS2000CG6		NMB1511
GGCAG3	No mutch				NMB2064 GGC5
GGCGC 3	NEIS1525-NMB1605		NMB1188		NMB0576 GGC5
	NEIS0103 NMB0110		NMB1348-NEIS1284		NEIS0311
	NEIS2009-NMB2030		NEIS0671		NEIS0357
AC5	NMB1693		NIES1742-NMB0422	ACGGC3	No mutch
	DWONSTREAM NEIS0395 NMB1823		NEIS0204-NMB0212	AAACAAACAA AC	N73-01522
	INTG NEIS1786 NMB0379		NEIS0191-NMB0199	CGCGC3	NEIS2009
GGATT3	No mutch		NEIS0186-NMB0195	CCAG(4-28)	N73-00567
TCAAA3	No mutch		NEIS0342-NMB1877	GGCGC 3	NMB1605-NEIS1525
AGAA3	No mutch		NESI0343		NEIS0103-NMB110
AAAT(5-14)	DOWNSTREAM NMB1994		NMB0045		NIES2009

			NMB0088	TGTTGA 2	NMB1379-NEIS1315
		GGAAGG 2	No mutch		NEIS0560NMB0617
				AGTTG 3	No mutch

The polymorphism mediating different lengths of the SSR was detected in all 12 strains (Table 2).

Table 2: Number of polymorphism for each SSR in all putative phase variable genes that predicted from carrier's isolates. CC1157, CC167, CC174, CC23-CC23-ST1655, CC23, CC60 and Serob are the carrier's isolates. Example, G7: means the number of G in SSR is seven (GGGGGGG).

Gene name	CC1157	CC167	CC174	CC23-CC23-ST1655	CC23	CC60	Serob	Number of polymorphism
N114-00492	G7							Poly=1
N64-01702	G7	G8	G9	G10	G11			Poly=5
NMB1969-2	C7	C8	C9	C10	C11	C12	C13	Poly=7
N188-00894		C8	C9	C10	C11	C12	C13	Poly=8
N199-00635	C7	C8	C9	C10	C11	C12	C13	Poly=7
N258-00214	G7	G8	G9	G10				Poly=4
N258-01303	G7	G8	G9	G10	G11	G12	G13	Poly=7
N59-01791	G7	G8	G9	G10	G11			Poly=5
N59-01936		2C8	18C9	C10	C11	C12	C13	Poly=7
N64-00871	C7	C8	C9	C10	C11	C12	C14	Poly=8
N64-01769		C8	C9	C10	C11	C12	C13	Poly=7
N73-00241	C7	C8						Poly=2
N73-01693	no match							
N64-00342	G7	G8	G9	G10				Poly=4
NMB-0218	C8							Poly=1
NMB0841	C7	C8	C9	C10	C11			Poly=5
NMB1541-1		C8						Poly=1
NMB1543	C7							Poly=1
NMB1668 -1	C7	C8	C9	C10	C11	C12	C15	Poly=7
NMB1797	C7	C8	C9					Poly=3
NMB1836	C7	C8	C9	C10	C11	C12		Poly=6
NMB1882-1		C8	C8					Poly=2
NMB1969	C7	C8	C9	C10	C11	C12	C13	Poly=7
NMB1931	G7	G8						Poly=2
NMB2132-1	no match							
ybiP	G7	G8	G9	G10	G11			Poly=5
NMB1443	no match							

Further work has been carried out by detecting the frameshift for all the putative phase variable genes (Table 3) and the Z score for each repeat tract was estimated by Markov model and synonymous codon shuffling model as illustrated in (Table 4).

Table 3: The frame shift for the genes in carrier's isolates. Indel: means frame shift found in the gene due to indel (insertion and deletion), off: means frame shift found in the gene due to SSR, On: means there is no any frame shift in the gene. CC1157, CC167, CC174, CC23-CC23-ST1655, CC23, CC60 and Serob are the carrier's isolates.

gene	CC1157	CC167	CC174	CC23-ST1655	CC23	CC60	Serob-N119
N73-00567	On	indel	indel	indel	indel	indel	indel
mboIIM	indel	indel	indel			Off	
N59 00037	indel	indel	indel	indel	indel	Off	indel
N188-01821							
N114 01371		indel	indel	indel		On	indel
N199-01562	indel	Off	Off	indel	Off	indel	Off
NMB1077	On					3/on	
NMB1913		On	On	On	On		On
NMB0663	indel	Off	indel	indel	indel	indel	Off
NMB0961						On	On
N199-01208	indel		indel			indel	On
epsH	On	indel	indel	indel	indel	indel	indel
NEIS1297 NMB1362	On	On	On	On	On	On	On
NEIS0103-NMB111	On				On	On	On
NMB 2030	On		On			On	On
NMB1693	On		On		On	On	On
NMB0961						On	On
NMB1895		On	On	On	On		
NEIS2135-NMB2157		On		On	On	On	On
NEIS0001-NMB0017 6GC	On	On	On	On	On	On	On
NMB0195	Off	On	Off	On	On	Off	Off
NMB1716	On	On	On	On	On	On	On
NMB0208		On	On				
NMB0352						On	
NEIS2000CG6	On	On	On	On	On	On	On
NMB0878			On				
NMB1590-NEIS1512 GGC5	On			On	On		

Table 4: Z score calculated by Markov model for all SSR

Repeat tract	Z score	Repeat tract	Z score	Repeat tract	Z score	Repeat tract	Z score
CCTG3/cagg3	-1.4	ACG4	0.7	A9	-1.73	G14	illegal
AAAT3	-0.9	GAAC3/GTTC3	0.7	A8	-3.21	CAAACAA3	-0.1

GAAA3/CTTT3	Illegal/-0.7	GCAG3	0.7	A11	1.79	CCCAA3	0
CAG4/ctg4	-0.7	CAAAT3	0.71	T8	-1.51	TGCG3	-0.5
TCCG3	-0.7	GCAG3	0.73	C4	-1.1	TTCC4	0
GGAC3	-0.6	AC5	1.3	C5	4.49	GGCA3	0
AAAC3/GTTT3	-0.5	GATG3/catc3	1.7	C7	1.59		
CGGCG3	-0.35	TGT4	1.78	C8	1.6		
GGC(5)	-0.28	GAA4 /ttc4	0.1/0.7	C9	-1.21		
CAAACAA3	-0.1	CGGG3/cccg3	0.6/-0.12	C10	illegal		
CAA4	-0.1	AAGC3/GCTT	0/0	C11	illegal		
GGCGC 3	0	GCCAA3	illegal	C12	illegal		
CGCGC3	0	TAGGCT3	illegal	C13	illegal		
TGTTT3	0	AT5	0.03	C14	illegal		
TGTTT3	0	CCTCCC3	illegal	G4	1.3		
ACGCGC3	0	CGGTGG3	illegal	G5	1.7		
GC6	0.23	TATT3/AATA3	-0.1	G7	0.82		
CCG5	0.4	AGCC3	0.5	G8	1.5		
CGGT3	0.4	AAGC3/GCTT3	0/0	G9	0.58		
AGCC3	0.5	CTTCT3	0	G10	5		
TTCC3	0.5	GGCGC3	0.28	G11	-4		
GCC5	0.5	AC5	1.3	G12	-0.41		
CG6	0.54	AAAT3/ATTT3	(-0.9) -0.2	G13	illegal		

For the SSR located within the open frame, position of the SSR at 3 end or 5 end were identified for all the putative phase variable genes (Table 5). On the other hand, For the SSR located within the intergenic, the Location of SSR within the elements of the promoter was detected.

Table 5: Detection the position of SSR within the gene as it located at 3' end of gene or 5' end of gene.

repeat type	gene	position
AGCC3	N73-00567	3' end
0.5	mboIIM	5' end
	N59-01623	
	N114 01898	
AAGC3/GCTT3		
0/0	N73-01522 END CONTIG	
	N59 00037	5' end
	N188-01821	
	N114 01371	5' end
	N199-01562	5' end
GAAA3/CTTT3	NMB1077	3' end
illegal/-0.7		
TTCC3	NMB1913	5' end

0.5		
CTTCT3	NMB0663	5′ end
0		
CAAAT3	NMB0961	5′ end
0.71		
GCCAA3	N199-01208	3′ end
CAAACAA3		
-0.1	epsH	3′ end
TAGGCT3	NEIS1297 NMB1362	5′ end
GGCGC 3		
0.28	NEIS0103-NMB111	5′ end
	NMB 2030	5′ end
AC5	NMB1693	5′ end
1.3		
TCAAA3	NMB0961	5′ end

Relying on the length of the repeat tract that fits our cut off (as mentioned previously), the number of genes was 327 for all 12 strains. However, the number of genes that fits all the criteria which are considered as strong putative phase variable genes were 67 out of 327.

Finally, the function of the strong putative phase variable genes was determined from National Center for Biotechnology Information website for each gene. Then, each gene is assigned to functional groups by different schemes Kyoto Encyclopedia of Genes and Genomes (KEGG).

The number of strong putative phase variable genes for proteins with unknown function or hypothetical, metabolism processing, process of environmental information, and process of genetic information were 22, 20, 10, and 6 respectively while the rest was pseudogenes. However, the number of genes in the whole-genome for proteins with unknown function or hypothetical, metabolism processing, process of environmental information, and process of genetic information is different therefore we had to normalize the gene number in scheme by the gene number in the overall genome. Therefore, we calculated the proportional effect of function of each gene (The proportional effect of each scheme was calculated by dividing the number of gene for each scheme on number of gene in whole genome). The proportions for proteins with unknown function or hypothetical and process of environmental information putative phase-variable were very high with 10.5 and 4.8 respectively, while the proportions for metabolism processing and process of genetic information were 2.8 and 1.5 respectively (Figure 1).

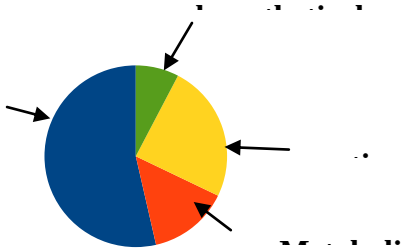

Figure 1: The proportional effect of a function of putative phase variable genes. Proteins with unknown function or hypothetical: blue colour, metabolism processing: yellow colour, process of environmental information: orange colour, process of genetic information: green colour.

Discussion

Initially, we designed our analysis to look for a simple repeat tract in 12 strains depending on the length of the repeat tract with specific cut-off (cut-off mentioned previously). We detected 327 putative phase variable genes. However, 119 putative phase variable genes were detected relying on the length of the repeat tract with the same cut-off only comparing two different strains [13]. Then, we applied our criteria to detect putative phase variable genes. We started to look for polymorphism in the SSR located in a particular gene among different strains and we searched if the SSR was stable or variable with a particular number among different isolates. The more polymorphism of SSR the more acts as putative phase variable. We detected 67 genes out of 327 showed polymorphism in the SSR. However, another study confirmed only 36 genes showing polymorphism in the SSR within different isolates [11]. It became necessary to consider the frameshift and the instability as important factors for the possibility of repeat to work as phase variables. Therefore, we looked for the frameshift in the gene that was caused due to the presence of a change in the SSR located within the phase variable genes. 67 out of 327 putative phase variable genes showed frameshift in the gene due to SSR however, [11] showed around 45 genes associated with frameshift while [13] revealed around 68 genes associated with frameshift.

The location of the repeat tract is central in classifying repeat tracts as phase variables or not. If the repeat tracts are positioned towards the 3' end of the gene, they are less likely to alter gene expression [18-4]. We found 67 out of 327 genes had their SSR positioned towards the 5' end of the gene and this result was compatible with a study carried out by [17]. One of the most important mechanisms of variation in phase variable genes is related to the variation seen in the distance between a component of promoters which are -10 and -35 from the translation initiation site. Depending on the

fact that the repeat tract which is located between -10 and -35 has a responsibility to change the distance between the -10 and -35, and this, in turn, leads to a change in the expression of a gene product. We found a high number of variable SSR located between components of promoter and this was compatible with the study achieved by [19]. In order to complete all our criteria to predict the possibility of SSR to generate phase variation we conducted further analysis to calculate the Z score using a Markov chain and shuffling models. Markov chain analysis is considered as another test for phase variation through estimation of the number of observed values of repeat tracts and their expected values within the genome sequence.

We extended our filtration for the putative phase variable genes through a search for the alignment of each gene. The alignment of loci was used as an indicator to characterize the repeat tract. Perfect alignments indicate the presence of a corresponding locus and an identical SSR. Imperfect alignments indicate high variability or repeat at ambiguous bases or the end of a contig.

From all the criteria above that have been used to filter the less likely of repeat tract to work as phase variable, we detected that there were 67 out of 327 putative phase variable genes fit our criteria, finding which may need further confirmation through experimental work.

The proportion effect of functional of putative phase variable genes showed that the environmental information processing was higher than metabolism and genetic information processing schemes and this is because the genes enrolled with environmental information processing coded for outer membrane proteins. It is, therefore, easy for the immune system to generate antibodies against them that is why they undergo phase variation mechanism to switch OFF the gene or level of transcription is low.

Conclusion

This study intended to predict the phase variable genes that carried SSR on their genic or intergenic regions using different criteria that have been taken from previously published works. We predicted 67 out of 327 putative phase variable genes that fit our criteria. These genes mainly coded outer membrane proteins. This is the reason that the immune system can recognize them easily and produce antibodies against them, which is why these genes carried SSR to enroll with phase variation and switch gene OFF or produce less amount of protein resulting in less likely adherence by antibodies and therefore resistance towards the attacking immune system. Further experimental work has to be conducted in order to provide strong evidence about these 67 genes and the phase variation mechanism.

References

1 - Oldfield NJ, Matar S, Bidmos FA, Alamro M, Neal KR, Turner DP, Bayliss CD, Ala'Aldeen DA. Prevalence and phase variable expression status of two autotransporters, NalP and MspA, in carriage and disease isolates of Neisseria meningitidis. PloS one. 2013 Jul 25;8(7):e69746.

2 - Arenas J, Paganelli FL, Rodríguez-Castaño P, Cano-Crespo S, van der Ende A, van Putten JP, Tommassen J. Expression of the gene for autotransporter AutB of Neisseria meningitidis affects biofilm formation and epithelial transmigration. Frontiers in cellular and infection microbiology. 2016 Nov 22;6:162.

3 - Green LR, Dave N, Adewoye AB, Lucidarme J, Clark SA, Oldfield NJ, Turner DP, Borrow R, Bayliss CD. Potentiation of phase variation in multiple outer-membrane proteins during spread of the hyperinvasive neisseria meningitidis serogroup W ST-11 lineage. The Journal of infectious diseases. 2019 Aug 30;220(7):1109-17.

4 - Janulczyk R, Masignani V, Maione D, Tettelin H, Grandi G, Telford JL. Simple sequence repeats and genome plasticity in Streptococcus agalactiae. Journal of bacteriology. 2010 Aug 1;192(15):3990-4000.

5 - Bayliss CD, Hoe JC, Makepeace K, Martin P, Hood DW, Moxon ER. Neisseria meningitidis escape from the bactericidal activity of a monoclonal antibody is mediated by phase variation of lgtG and enhanced by a mutator phenotype. Infection and immunity. 2008 Nov;76(11):5038-48.

6 - Metruccio MM, Pigozzi E, Roncarati D, Berlanda Scorza F, Norais N, Hill SA, 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 pathogens. 2009 Dec 24;5(12):e1000710.

7 - Sadarangani M, Hoe JC, Makepeace K, Van Der Ley P, Pollard AJ. Phase variation of Opa proteins of Neisseria meningitidis and the effects of bacterial transformation. Journal of biosciences. 2016 Mar 1;41(1):13-9.

8 - Green LR, Lucidarme J, Dave N, Chan H, Clark S, Borrow R, Bayliss CD. Phase variation of NadA in invasive Neisseria meningitidis isolates impacts on coverage estimates for 4C-MenB, a MenB vaccine. Journal of clinical microbiology. 2018 Sep 1;56(9):e00204-18.

9 - Phillips ZN, Tram G, Seib KL, Atack JM. Phase-variable bacterial loci: how bacteria gamble to maximise fitness in changing environments. Biochemical Society Transactions. 2019 Aug 30;47(4):1131-41.

10 - Siena E, D'Aurizio R, Riley D, Tettelin H, Guidotti S, Torricelli G, Moxon ER, Medini D. In-silico prediction and deep-DNA sequencing validation indicate phase variation in 115 Neisseria meningitidis genes. BMC genomics. 2016 Dec;17(1):1-3.

11 - Saunders NJ, Jeffries AC, Peden JF, Hood DW, Tettelin H, Rappuoli R, Moxon ER. Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58. Molecular microbiology. 2000 Jul;37(1):207-15.

12 - Martin P, Van De Ven T, Mouchel N, Jeffries AC, Hood DW, Moxon ER. Experimentally revised repertoire of putative contingency loci in Neisseria meningitidis strain MC58: evidence for a novel mechanism of phase variation. Molecular microbiology. 2003 Oct;50(1):245-57.

13 - Snyder LA, Butcher SA, Saunders NJ. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic Neisseria spp. Microbiology. 2001 Aug 1;147(8):2321-32.

14 - Sreenu VB, Ranjitkumar G, Swaminathan S, Priya S, Bose B, Pavan MN, Thanu G, Nagaraju J, Nagarajaram HA. MICAS: a fully automated web server for microsatellite extraction and analysis from prokaryote and viral genomic sequences. Applied bioinformatics. 2003 Jan 1;2(3):165-8.

15 - Klughammer J, Dittrich M, Blom J, Mitesser V, Vogel U, Frosch M, Goesmann A, Müller T, Schoen C. Comparative genome sequencing reveals within-host genetic changes in Neisseria meningitidis during invasive disease. PloS one. 2017 Jan 12;12(1):e0169892.

16 - Lin WH, Kussell E. Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. Nucleic acids research. 2012 Mar 1;40(6):2399-413.

17 - Orsi RH, Bowen BM, Wiedmann M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. BMC genomics. 2010 Dec;11(1):1-2.

18 - van Passel MW, Ochman H. Selection on the genic location of disruptive elements. Trends in Genetics. 2007 Dec 1;23(12):601-4.

19 - van der Ende A, Hopman CT, Dankert J. Multiple mechanisms of phase variation of PorA in Neisseria meningitidis. Infection and immunity. 2000 Dec 1;68(12):6685-90.

20 - Haseneyer G, Stracke S, Piepho HP, Sauer S, Geiger HH, Graner A. DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. BMC plant biology. 2010 Dec;10(1):1-1.