# Screening DNA repeat tracts of phase variable genes by fragment analysis

**Freda E-C Jen, Kate L Seib*, Aimee Tan**

Institute for Glycomics, Griffith University, Gold Coast, Queensland, Australia

* Address correspondence to k.seib@griffith.edu.au

**Key Words** Fragment analysis, polymeric DNA tract, phase variable gene, DNA repeat sequence

**Running Title** Fragment analysis of DNA repeats

**Abstract**

Fragment analysis (or fragment length analysis) is a PCR based method which allows quantification of the size and proportion of a DNA repeat tract length of a phase variable region. Primers are labelled with a fluorescent dye, the resulting amplicons are processed by capillary electrophoresis and results are analysed for amplicon size and proportion by associated software (such as Peakscanner). Here we describe the process of designing primers and controls to screen for the number of repeats in a polymeric tract of a phase variable gene in *Neisseria meningitidis* (the DNA methyltransferase ModA is used as an example, but this method can be applied to other phase variable genes).

## 1. Introduction

*Neisseria meningitidis* is a host-adapted pathogen and many of its virulent genes are phase-variable ***(1-6).*** Phase variation is the high frequency, reversible switching of gene expression. In many cases, phase-variation is caused by the high frequency mutation of DNA due to DNA slippage in tandem repeat tracts, causing gene and protein expression to vary. Both homopolymeric and hetropolymeric DNA repeat tracts have been seen in the pathogenic *Neisseria* ***(7)***. DNA tracts may be located within the gene, which can result in switching between phase ON (expressed) and phase OFF (not expressed) due to the introduction of frameshift mutations (8), or upstream of the gene at promoter regions, which can lead to altered levels of expression ***(9)***.

Phase variation is a random process, and random switching can be observed in independent single colonies isolated during cloning or from different growth conditions. This can have significant consequences for experimental analysis, as without knowledge of whether phase-variation has taken place, phenotypes may be misattributed to the wrong determinants ***(10)***. Various methods can be used to determine the ON/OFF status of particular phase variable genes: (i) PCR based fragment analysis, that will be discussed further in this chapter ***(8)***, (ii) sequencing of the phase-variable region ***(7)***, or (iii) Western blot analysis if antibodies are available ***(11)***. Of these methods, fragment analysis is the only method that allows simultaneous quantification of both the number of repeats in the phase variable DNA region and the relative proportion of each repeat length within the bacterial population being studied ***(11)***. This is achieved by using a standard PCR with a primer labelled with fluorescent dye at the 5-prime end to produce labelled amplicons. The PCR should amplify all repeat sizes in a non-biased manner, and so amplicons reflect the proportion of repeat tract lengths in the template population. The amplified PCR products are electrophoresed through a polymer-filled capillary along with appropriate size standards to determine the exact size of the labelled amplicons, and the fluorescence intensity of each amplicon can be used to determine the proportion of each tract length. The output can be viewed using software such as Peak Scanner.

Correlation of peak size, repeat number and phase ON/OFF may require further determination (see section 3.6). However, once this calibration has been established, fragment analysis may be used to reliably determine the expression of the gene based on the fragment size determined.

## 2. Materials

### 2.1. Design of oligonucleotide primers

1. Programs for genetic analysis/sequence viewing: e.g. MacVector, Geneious, Bioedit.
2. Programs for primer design: primer3 (often incorporated into commercial software); online design programs are also available (e.g. NCBI Primer-BLAST). Many of these programs will also analyse secondary structure and self-interactions.

### 2.2. Oligonucleotide Primers

1. Commercially available from vendor of choice. Primers can be ordered with 5' fluorescent dye labels, but some dyes may be dependent on supplier (e.g. VIC, NED and PET are propriety to ABI, ThermoFisher Scientific). Your vendor may be able to recommend appropriate replacements.

### 2.3. PCR

1. Thermal cycler (i.e. PCR machine).
2. PCR reagents: DNA polymerase (Polymerases will be sold as kits or master mixes, with buffer and $MgCl_2$ components), dNTPs (may be included with kit for specific concentrations, but may also be purchased from independent vendors), oligonucleotide primers (commercially available, as described above), and $dH_2O$.
3. Template: Template preparation reagents will depend on the requirements of the experiment. Some examples include:
    a. Genomic DNA (gDNA): gDNA extraction kit.
    b. Colony PCR: $dH_2O$, heat block.

c. Quality control template (positive control/calibration): synthetic DNA with exact numbers of repeats may be ordered from commercial suppliers (e.g. gBlocks (Integrated DNA Technologies, IDT)).

### 2.4. Agarose gel electrophoresis

1. Agarose.
2. 0.5M EDTA (ethylenediamine tetra-acetic acid): adjust to pH 8 with NaOH (*see* **Note 1**).
3. Tris-acetate EDTA (TAE) buffer (X50): 242 g Tris-base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA per L $dH_2O$. Dilute to 1x with $dH_2O$ for use. Final pH should be around 8.5 (*see* **Note 2**).
4. Tris-borate EDTA (TBE) buffer (X5): 54 g Tris-base, 27.5 g boric acid, 20 mL 0.5 M EDTA per L $dH_2O$. Dilute to 1x with $dH_2O$ for use. Final pH should be around 8.0 (*see* **Note 2**).
5. Ethidium bromide or SYBR SAFE (*see* **Note 3**).
6. Horizontal electrophoresis gel tank.
7. Gel trays and well combs.
8. Gel caster (optional).
9. Power pack.
10. DNA ladders/markers.
11. DNA loading dye (if not supplied as part of PCR master mix): 30% (v/v) glycerol; 0.25% (w/v) bromophenol blue (or other dye, e.g. orange G) in $dH_2O$. Filter before use. May be purchased commercially or made in the lab.
12. Visualisation: UV (or blue light) transilluminator. This may be embedded in a unit with mounted camera for documentation.

### 2.5. Capillary electrophoresis

1. Commercial services for fragment analysis are routinely used. If not, minimal requirements are:
   a. Capillary electrophoresis system (e.g. Genetic Analyzer, ABI).
   b. Size standards (e.g. GeneScan 500 LIZ).

### 2.6. Data analysis program

E.g. Peakscanner, Genemapper (Thermo Fisher Scientific).

## 3. Methods

### 3.1. Design oligonucleotide primers (*see* Note 4)

1. Identify conserved regions flanking the DNA repeat tract by comparison of sequence data from different genomes (e.g. Fig. 1) (*see* **Note 5**).
2. Design primers in conserved regions flanking the DNA repeat tract, manually or using programs such as primer3 (Fig. 1) (*see* **Note 6**).

   2.1. One primer of each pair should be labelled with a fluorescent dye (e.g. 6FAM) (*see* **Note 7**).

### 3.2. Fragment length analysis PCR

1. Prepare template DNA: Template DNA may consist of purified DNA, crude cell extract or whole bacterial cells (colony PCR) (*see* **Note 8**). Template preparation for colony PCR is as follows: a colony (or several colonies) of bacteria to be analysed is/are suspended in 100 µl sterile water and vortexed briefly to mix. The sample is boiled for 10 minutes or incubated at 60 ˚C for 1 hour to kill the bacteria. 1 µl of this is used as template DNA.
2. Set up fragment length PCR: A master mix is made up for PCR for n+1 samples, as per manufacturer's instructions (*see* **Note 9**). Dispense master mix into 0.2 mL PCR tubes or plate, then add 1 µl template per well and mix. Run PCR as per manufacturer's instructions.

### 3.3. Visualise amplicons using agarose gel electrophoresis

1. Make agarose gels with 1.5% agarose (w/v) in 1x TAE or TBE buffer (*see* **Note 10**): e.g. add 1.5 g to 100 mL TAE buffer in an erlynmeyer flask, microwave to dissolved agarose (60-90 seconds on high, mix periodically) and add 0.2-0.5 µg/mL ethidium bromide (or 1x SYBR SAFE) to the dissolved agarose (*see* **Note 3**).

2. Cool mix until it Pour into gel tray and allow to set. Gel will become more opaque when set. When set, transfer gel to tank filled with corresponding buffer and flush out wells (*see* **Note 11**).
3. If no loading dye is included in the PCR master mix buffer, mix 1 µl 6x loading dye with 5 µl PCR product in a fresh tube.
4. Load 5 µl PCR product into wells on agarose gel, with DNA ladder of appropriate size or standardised concentration (e.g. 100 bp ladder) (*see* **Note 12**).
5. Run 30-60 minutes at 90-110V (TBE vs TAE) (maximum recommended voltage is 10V/cm), or until the dye front has migrated at least halfway down the gel (*see* **Note 13**).
6. Visualise gel on gel imaging system – an example of a visualised gel in provided in Fig. 2.

### 3.4. Analysis of size by capillary electrophoresis

1. Capillary electrophoresis is carried out by most DNA sequencing services, using Sanger sequencing machines such as the Genetic Analyzer (ABI), as per manufacturer instructions.
2. PCR reactions do not need to be cleaned prior to electrophoresis, but concentration should be adjusted to 5-10 ng/µl with distilled water (*see* **Note 14**), and 1 µl run.
3. Size standard: reactions should be run simultaneously with an internal size standard such as Genescan LIZ500 (*see* **Note 15**).

### 3.5. Results analysis

1. Import .fsa files into the analysis program (*see* **Note 16**), and set size standard and analysis method (PP – with primers, vs NPP – without primers), then select analyse (Fig. 3).
2. Under review data, view peaks in region of expected amplicon (Fig. 3).
3. Export peak size, height and area values from relevant peaks: edit table to display file name, size, height and area in point only, highlight peaks in region of interest and select 'show selected peaks', then export data to excel.
4. Calculate the relative ratios of peaks as the height of the peak divided by total height (see Fig. 3 for example calculation). Area under curve can be used instead of height, and should give similar proportions (*see* **Note 17**).

### 3.6. Further work

Qualification of whether peak sizes correlated with phase ON or OFF requires further experimental work that is beyond the scope of this method. However, potential assays include quantitative real time PCR (variation in transcription) or Western blot (variation in protein production). For these assays, isolate strains for which fragment length analysis produces different peak sizes (corresponding to different numbers in the phase variable repeat tract), and conduct comparative assays for mRNA or protein production.

## 4. Notes

1. EDTA will not go into solution unless the concentration is correct. Use NaOH to adjust – solid pellets can be used at first, then switch to concentrated solution as pH approaches 8.
2. The pH of the TAE and TBE solutions should be correct without further adjustment. However, it may be prudent to check pH if issues are seen with gel electrophoresis – incorrect pH is indicative of insufficient buffering and causes abnormal migration of DNA in gels. Correctly pHed buffers can be reused for several gels, but will need to be replaced periodically.
3. Ethidium bromide is mutagenic and nitrile gloves (not latex) should be used when handling. SYBR safe is considered to be a less hazardous replacement.
4. Identification of repeat tracts. This method starts from primer design, assuming the phase variable gene and repeat tract have already been identified. If this is not the case, bioinformatic identification of DNA repeat tracts that may lead to phase variable expression can be carried out ***(4, 12-16)***. Minimal repeat numbers for phase variation to potentially occur are 7 for single polynucleotide repeats ***(17)*** , or 3 for tri-nucleotide repeats and above ***(7)*** . Potential for phase variation is assessed by comparing sequences for the region e.g. from different genomes, sample times or from comparing reads from a genome sequence (Fig. 1).

5. Primers should be designed in conserved regions flanking the repeat tract. Design of primers least 30 bp away from the repeat tract being analysed may be beneficial in case sequencing is desired. Regions should also be checked for indel variations as this will alter the size of the amplicon to be analysed, and these should be avoided if possible. Some issues that complicate fragment analysis are highlighted in Fig. 5, using the *opa* genes as an example.
6. Oligonucleotide primers for fragment length analysis should conform to standard primer requirements for the DNA polymerase to be used. Typically, primers should be between 18-25 base pairs in length and contain 50% GC content, spread evenly throughout the primer (i.e. no polynucleotide stretches). Checks for secondary structure formation (e.g. hairpin loops), self-annealing, or primer dimer formation will be carried out in online primer calculators and when ordering from most manufacturers. Amplicons of up to 1 kB can be analysed by fragment analysis, but generally the amplified region is between 100 – 200 base pairs in length.
7. Labels: one primer from each oligonucleotide primer pair is labelled at the 5' end with a fluorescent dye. 6FAM is the most commonly used for fragment length analysis, but alternatives may be used including VIC, NED, PET, HEX. Choice of dye is contingent on the machine to be used for analysis; and relative intensity of dyes. NOTE: In general, 6FAM is the more frequently used due for this work, and oligos with this label can be ordered from major suppliers. VIC, NED and PET are propriety dyes from Thermo Fisher, but substitutes (such as the ATTO dyes) are available.
8. Quantity and quality of DNA needed for fragment length analysis is contingent on the sample being examined, however a wide range of templates have been used successfully, including purified gDNA (by commercial kit or crude extract), whole cell samples (colony PCR) and mixed samples (e.g. tissue culture cells lines infected with bacteria). However, the quantity of DNA required will vary with quality and the general requirements for the polymerase being used. Purified gDNA can be used at concentrations as low as 1 ng/µl, but mixed or unpurified samples, may require more template than usual – for example, if using template from an infected sample (e.g. homogenised infected tissue), there will be a significant amount of eukaryotic DNA present that can interfere with the reaction. Calibrate the amount of template used by serially diluting the template DNA to determine amount

needed for PCRs. Note that amplification efficacy may be lowered by the fluorescent label relative to unlabelled DNA oligos.

9. Both proof-reading and non proof-reading polymerases can be used for fragment length analysis. Selection of the polymerase is up to the researcher's needs and budget. There are additional considerations for Taq selection. For example, samples may contain components that inhibit PCR reactions – e.g. blood. In these cases, it may be necessary to use specialised polymerases and kits to optimise reactions (e.g. Phusion blood direct PCR kit). Also, non-proof reading enzymes may not be ideal when analysing single nucleotide repeats, as A-tailing (addition of adenine on the 3' end of DNA) may give erroneous readings on repeat tract size. To overcome possible issues, extend the final extension time on fragment length reactions, and process amplicons through an additional A-tailing reaction to ensure all amplicons are completely A-tailed; alternately a proof-reading enzyme may be used.
10. Either TAE or TBE buffer can be used for agarose gel electrophoresis of amplicons for visualisation. Mobility and resolution of small fragments may be better in TBE, however use of TAE is sufficient for this work as size and amounts are quantified by capillary electrophoresis. As fragments are small, an agarose content of between 1.5-2% may be desirable (Fig. 2). Gels can alternatively be post-stained with ethidium bromide/SYBR safe by adding the stain to the buffer used (in concentrations as described in methods), and gently shaking the gel in this solution until sufficient staining is achieved. Note that these dyes are light sensitive and incubation should be carried out in the dark – e.g. by wrapping the container in foil. Due to the mutagenic nature of ethidium bromide, it is also desirable to carry this out in a waterproof container.
11. Gels can be set without a gel caster by taping ends of trays with masking tape. If this is being done, take care that the agarose gel has cooled sufficiently before pouring to avoid leakage.
12. Use a DNA ladder/marker with known loading concentrations, as specified by manufacturer's instructions. This will allow the amplicon intensity to be used to estimate DNA concentration in the sample.

13. Run conditions: applied voltage of between 80-110V can be used (generally, 10V/cm between electrodes is recommended), with run times inversely proportional to the voltage (i.e. the higher the voltage, the shorter the running time). Note that excessive voltages may significantly heat gels and cause them to melt – this is of more concern with gels with lower agarose content. Run time can be determined by observing the position of the dye front; migration between halfway and two-thirds of the way down the gel should give good resolution.
14. Before sending for analysis adjust concentration to approximately 5-10 ng/µl, based on amplicon intensity on the agarose gel relative to the DNA ladder used. Note that attempts to estimate concentration by spectrophotometer (such as Nanodrop) will not work due to the presence of primers and dNTPs. It is not necessary to clean reactions, as the assay will only detect fluorescently labelled molecules, and labelled primers should be cleanly separated from amplicon peaks by size (Fig. 3).
15. Commonly used size standards are LIZ500 and LIZ600, however fragment length service providers will be able to advise more specifically on the dyes and standards they use. Note that the standard used may affect the choice of fluorescent label for primers.
16. Analysis programs: various programs are available to analyse .fsa files generated by capillary electrophoresis, such as Genemapper (ABI) and peakscanner (ThermoFisher). The analysis presented in methods has been carried out using Peakscanner 2. For .fsa files, it may be necessary to set size standard (i.e. LIZ500), and analysis method (PP – with primers, vs NPP – without primers). If capillary electrophoresis was carried out by a DNA sequencing facility, this analysis may be carried out as part of the service.
17. Note that the size of the fragment may differ from the expected size of the amplicon, due to the fluorescent dye attached altering migration. Use of a synthetic version of the region with a known number of repeats (e.g. a gBlock, IDT) can allow precise calibration of peak size to repeat number. Depending on the polymerase used, A-tailing may cause lesser peaks of ± 1 bp either side of an expected repeat number (Fig. 4). If the repeat tract is of sufficient size (tetrameric repeat or higher), these peaks may be grouped with the major peak when calculations are carried out.

**References**


1. Livorsi DJ, Stenehjem E, Stephens DS (2011) Virulence factors of gram-negative bacteria in sepsis with a focus on *Neisseria meningitidis*. Contrib Microbiol 17:31-47.

2. Rotman E, Seifert HS (2014) The genetics of *Neisseria* species. Annu Rev Genet 48:405-31.

3. Srikhanta YN, Fox KL, Jennings MP (2010) The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol 8(3):196-206.

4. Saunders NJ, Jeffries AC, Peden JF et al (2000) Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. Mol Microbiol 37(1):207-15.

5. Peng J, Yang L, Yang F et al (2008) Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone. Genomics 91(1):78-87.

6. Parkhill J, Achtman M, James KD et al (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature 404(6777):502-6.

7. Power PM, Roddam LF, Rutter K et al (2003) Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. Mol Microbiol 49(3):833-47.

8. Srikhanta YN, Dowideit SJ, Edwards JL et al (2009) Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. PLoS Pathog 5(4):e1000400.

9. Tauseef I, Ali YM, Bayliss CD (2013) Phase variation of PorA, a major outer membrane protein, mediates escape of bactericidal antibodies by *Neisseria meningitidis*. Infect Immun 81(4):1374-80.

10. Anderson MT, Seifert HS (2013) Phase variation leads to the misidentification of a *Neisseria gonorrhoeae* virulence gene. PLoS One 8(8):e72183.

11. Seib KL, Jen FE, Tan A et al (2015) Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N(6)-adenine DNA methyltransferases of *Neisseria meningitidis*. Nucleic Acids Res 43(8):4150-62.

12. Hood DW, Deadman ME, Jennings MP et al (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. Proc Natl Acad Sci U S A 93(20):11121-5.

13. Saunders NJ, Peden JF, Hood DW et al (1998) Simple sequence repeats in the *Helicobacter pylori* genome. Mol Microbiol 27(6):1091-8.

14. Jordan PW, Snyder LA, Saunders NJ (2005) Strain-specific differences in *Neisseria gonorrhoeae* associated with the phase variable gene repertoire. BMC Microbiol 5:21.

15. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573-80.

16. Martin P, van de Ven T, Mouchel N et al (2003) Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: evidence for a novel mechanism of phase variation. Mol Microbiol 50(1):245-57.

17. Jen FE, Warren MJ, Schulz BL et al (2013) Dual pili post-translational modifications synergize to mediate meningococcal adherence to platelet activating factor receptor on human airway cells. PLoS Pathog 9(5):e1003377.

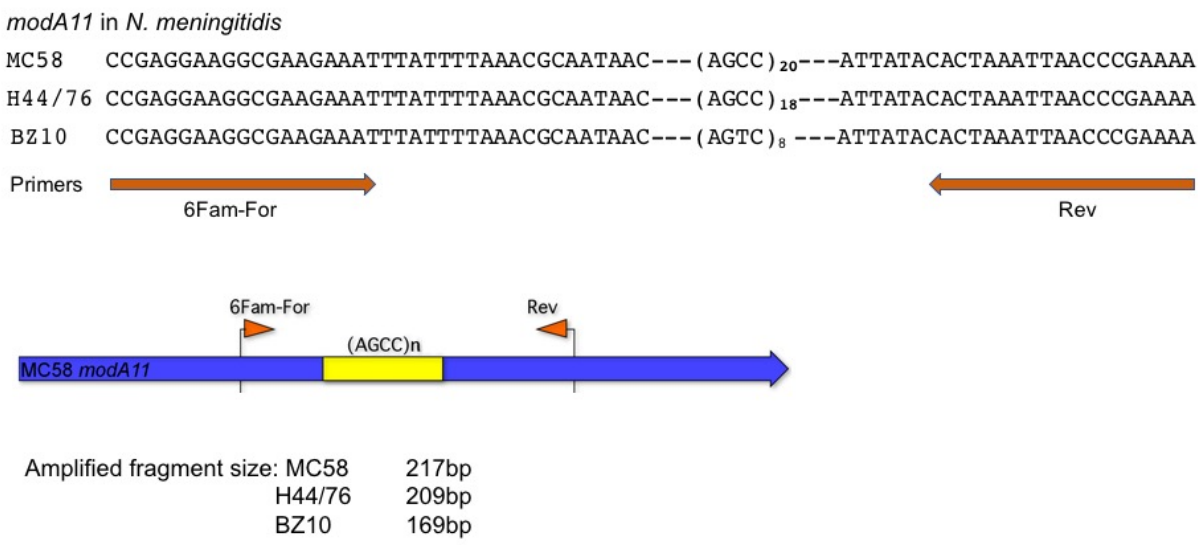



**Figure 1. Primers designed for fragment analysis.**

MC58 *modA11* as an example. Primers bind to the conserved region flanking the repeat region. Due to the difference in numbers of repeat in different *N. meningitidis* strains, the PCR amplified fragment size is different.

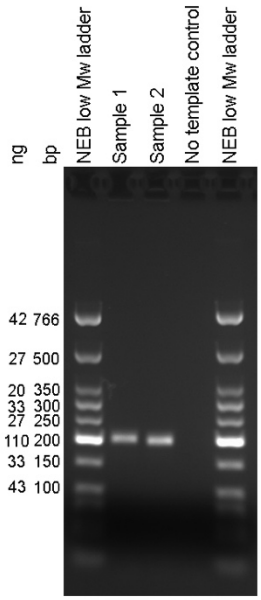



**Figure 2. Fragment analysis sample in DNA agarose gel.**

5ul of 30ul PCR reaction were run in the gel. This gel was run for 50 min 110V in 1x TBE.

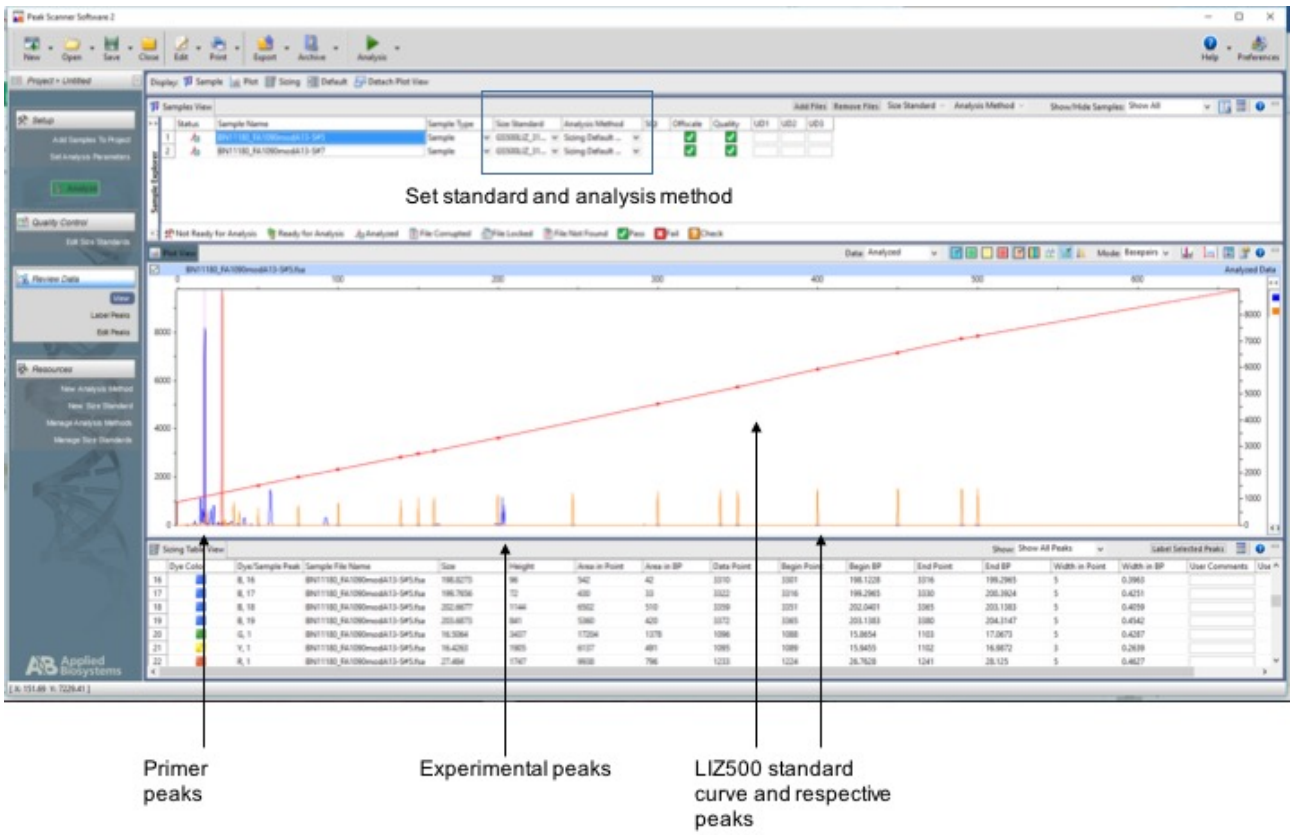
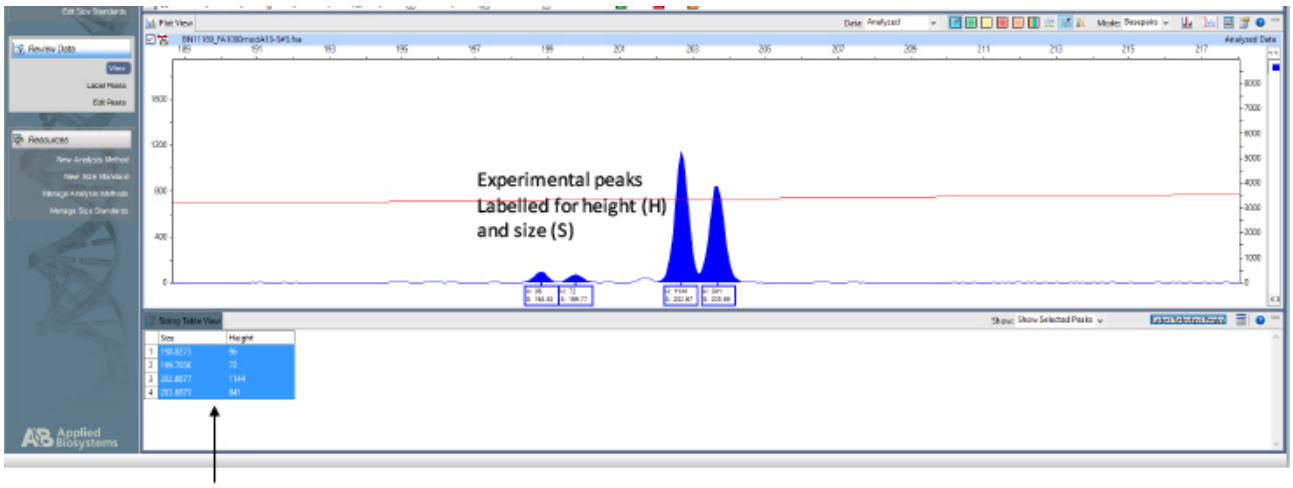



**Figure 3. Peakscanner overview for *modA* example.**

Size standard and Analysis methods are set in the setup window, before proceeding to analysis, which will generate the plot (middle window). Selection of appropriate filters shows size standard peaks in orange and the curve in red. Experimental peaks for this experiment are shown in blue (6FAM label).




Data table

Sample calculation:

| Peak size[a] | Peak height[a] | % peak proportion[b] | Repeat #[c] | Phase[d] | |
|---|---|---|---|---|---|
| 198.8273 | 96 | 4.45 | 16 | ON | 7.8 % phase on |
| 199.7656 | 72 | 3.34 | 16 | ON | |
| 202.6677 | 1144 | 53.14 | 17 | OFF | 92.2% phase off |
| 203.6873 | 841 | 39.06 | 17 | OFF | |

[a] Peak height and size exported from Peakscanner, correlating to amplicon size (bp) and proportion (height)
[b] Peak proportion (%) calculated as (peak height)/(total peak height)*100, for each individual peak
[c] Number of repeats correlating to peak/amplicon size. Note *modA* had a four nucleotide repeat unit, and two peaks share the same repeat number
[d] Phase of expression, as determined previously by Western blot analysis. This information allows overall phase of the population to be determined as 7.8 % in phase on, and 92.2 % in phase off.

**Figure 4. Peakscanner peak analysis overview and sample calculation for *modA* example.**

Zoom in on experimental peaks in expected amplicon size region. Peaks can be selected and labelled (shown here with peak size (S) and height (H)). The respective data table can be exported for further analysis. Relative proportions of peak abundance and corresponding repeat number and phase are shown in the table below.

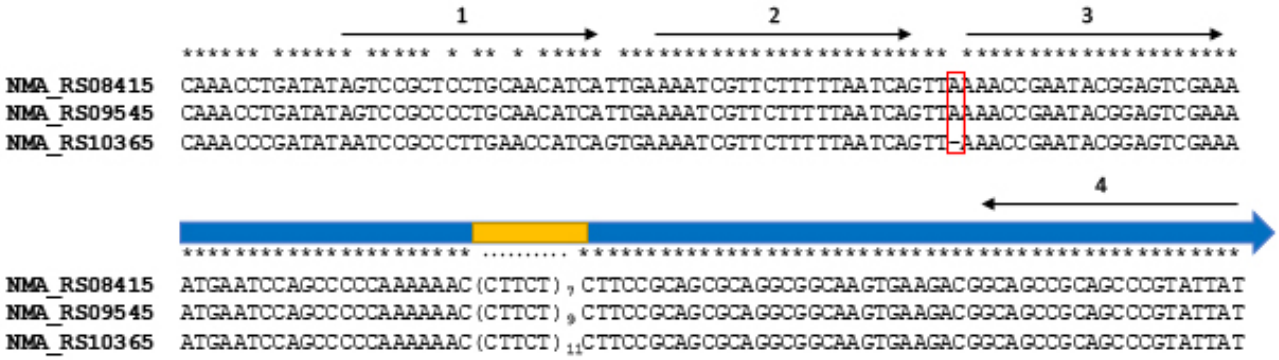



**Figure 5. Alignment of the upstream and N-terminal sequence of *opa* genes in *Neisseria meningitidis* Z2491.**

*N. meningitidis* genomes possess multiple copies and alleleic variants of the *opa* genes (start and N-terminal coding sequence shown as blue arrow). Whilst conserved regions (denoted by *) exist upstream and downstream of the repeat tract (yellow box), the presence of an indel variant (boxed in red) and SNP upstream mean that these genes are not easily analysed by fragment length analysis: placement of a primer (black arrow) in position 3 will allow consistent amplification of the repeat tract, however will not differentiate between allelic variants. Placement of primers upstream of the indel (positions 1 and 2) will potentially create issues with future calibration work to determine whether the gene is in phase on or off. A primer placed in the SNP-rich region of position 1 may amplify the genes with variable efficacy, causing inaccurate peak ratios.