Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

This document describes two assays, the first for MLVA profiling and the second for instrument calibration, peak binning setup and confirmation of new alleles (Robinson *et al.*, 2022; Risby *et al.*, 2023).

Assay name: CRU *C. parvum* Multiple locus variable-number tandem repeat analysis (Cp MLVA)
Test target: *Cryptosporidium parvum*
Assay description: Subtyping of *C. parvum* by MLVA profiling, using repeat numbers inferred from Applied Biosystems (ABI) fragment sizing targeting a panel of seven variable-number tandem repeat (VNTR) loci in two multiplex PCR reactions.

This assay has been validated as described; any modification of these conditions or reagents may alter its performance and require optimisation.

Assay design - labelled MLVA PCR primers (as combined in the four- and three-plex PCRs):

PCR Element (locus)	Primer Name	Fluorophore and Sequence 5'-3'
Four-plex		
Forward primer (cgd1)	cgd1_470_1429F	**PET**-CTCAGGAAGAGGAAGATACGG
Reverse primer (cgd1)	cgd1_470_1429R	GGAAGGTATGGCAGCAAAAG
Forward primer (cgd4)	cgd4_2350_796F	**VIC**-GGGTCAATCAGGCATGAGC
Reverse primer (cgd4)	cgd4_2350_796R	TTGCATGTTTATCATATTATTTCCCAT
Forward primer (cgd8)	cgd8_4440_NC_506F	**NED**-CTCAATATTTTTTCCACACCTGAAC
Reverse primer (cgd8)	cgd8_4440_NC_506R	ACTGCCTGAGAAAGGAACCA
Forward primer (MM19)	cgd8_4840_6355F	**FAM**-GTTCCAGGAATATTTGATTCTGC
Reverse primer (MM19)	cgd8_4840_6355R	CTCCTACGCCAACTCCTA
Three-plex		
Forward primer (MSF)	cgd5_10_310F	**VIC**-AAGGTGAAGGAATCAAAGGC
Reverse primer (MSF)	cgd5_10_310R	TTTGTCCTTCTTGCCCTCGG
Forward primer (cgd5)	cgd5_4490_2935F*	**FAM**-CAGTGAATAACTCTGAACGGAAC
Reverse primer (cgd5)	cgd5_4490_2935R*	TTGATTTTGGGTTCGGTATTG
Forward primer (cgd6)	cgd6_4290_9811F	**NED**-CATTGGAACGTAAACAAAACCA
Reverse primer (cgd6)	cgd6_4290_9811R	CTAGCCGAATCTGGCGGTAT

NOTE: This is not the order of loci for expression of results as a MLVA profile (see Interpretation on page 2)
* The locus cgd5_4490_2935 is the same as that originally named cgd5_4490_2941. The name has been updated to reflect the true start position of the counted repeat units, which is located 2935 bp downstream of the start of gene cgd5_4490.

Assay design – PCR reaction mix:

Component	Specific reagent	Volume (µl)
PCR master mix	2x Type-it Microsatellite PCR master mix (Qiagen)	12.5
10x primer mix (2µM of each primer)	Stock concentration of each primer provided at 100 µM (ABI). To prepare 500 µL of working concentration (2 µM): • Four-plex: 10 µL of each primer, and 420 µL of water. • Three-plex: 10 µL of each primer, and 440 µL of water.	2.5
nfH_2O	Any nuclease free / molecular grade	8
Template DNA		2
Total volume:		25

Reaction conditions:

Stage	Temperature	Duration (sec)
Hold (Hot start)	95 °C	300
Cycling (30 cycles): Denaturation Annealing Extension	 95 °C 63 °C 72 °C	 30 90 30
Final extension	60 °C	1800
Hold	4 °C	∞

Assay format: 8-well strips or 96-well plate

Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

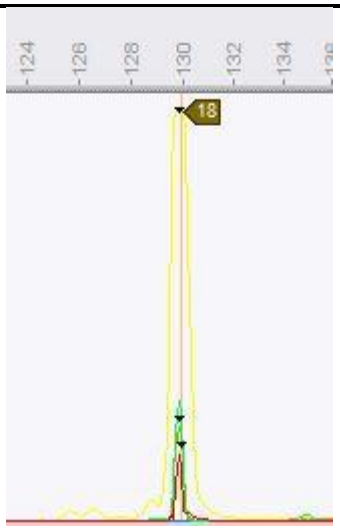
Interpretation:

Fragment sizing on ABI platform:

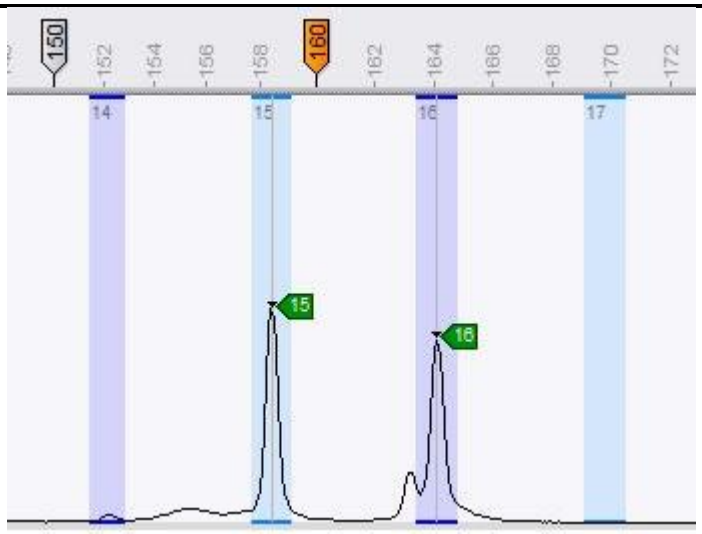
- Interpret allele sizes according to the software available, for example BioNumerics (bioMérieux) or Geneious (Biomatters)
- The allele is identified based on the size of the fragment, indicating the number of repeat units, calibrated against sequenced reference standards.
- For initial instrument calibration and setting up peak bins, as well as confirmation of new alleles, a suitable set of unlabelled primers has been designed to verify the number of repeats by sequencing (see page 3 for the confirmatory sequencing assay description).
- All new alleles must be sequenced and aligned to the reference library (see pages 5 to 7) to confirm the true number of repeats and accurately assign allele bins in the analysis software.
- The full MLVA profile is compiled by listing the calculated number of repeats (not sizes) at each allele (e.g. 4-14-5-7-27-28-16) in chromosomal order (i.e. cgd1-cgd4-MSF-cgd5-cgd6-cgd8-MM19).

Caution should be taken regarding:

- **Differentiation of a true low-level peak from baseline signal noise:** Minimum threshold of values (peak height) may be established, although there are no standardised rules. ABI suggests for most of their instruments an ideal peak height of ≥150 RFU, but that peak heights can be ≥50 RFU and those <50 RFU considered to be noise (Applied Biosystems, 2014). DNA samples containing peaks below this threshold can be concentrated (e.g. by desiccation) and repeated.
- **Stutter peaks**: Can be ignored – they are easily identified as they typically appear 1-2 repeats shorter than the corresponding allele peak. They are usually seen in loci with repeats that are ≤6 bp.
- **Wider and less intense peaks than true alleles**: Wide and low artefactual peaks that appear at an apparent size unique for each dye and are evident in all samples on a run, including the NTC, can be ignored.
- **Intense peaks that bleed through to the fluorescence channel of interest**: The bleed through peaks should be ignored and can be checked by turning on the other dyes from that multiplex PCR (Figure A below).
- **Double peaks**: Apply a Peak Height Ratio (PHR) (height of lower peak : height of higher peak) of >0.25 to consider the lower peak as an additional true peak (Figure B below). NOTE the occasional presence of a second reverse primer site in MM19 may generate double peaks at this locus that need confirmation by sequencing.



A: Example of intense yellow peak being detected in the green and red channels



B: Example of a mixed sample showing true double peaks

Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

Assay name: CRU VNTR Seq
Test target: *Cryptosporidium parvum*
Assay description: Simplex PCRs for amplification and Sanger sequencing of each locus, used for initial instrument calibration and peak bin setup, and to confirm new alleles identified by ABI fragment sizing.

This assay has been validated as described; any modification of these conditions or reagents may alter its performance and require optimisation.

Assay design - unlabelled sequencing PCR primers:

Element	Primer Name	Sequence 5'-3'
Forward primer	cgd1SeqF	AGGAGGAAGTGGCGAATCAGG
Reverse primer	cgd1SeqR	GATAGGGTAGTTTTACCGGGGTTG
Forward primer	cgd4SeqF	CAATGGATGCCAGACAAGCT
Reverse primer	cgd4SeqR	AAGGCTACGAGCAGATTAACG
Forward primer	MSFSeqF	TGACCCTTCTATTGAGCCAC
Reverse primer	MSFSeqR	ACTTCTTCCTCATCAGTAGC
Forward primer	cgd5SeqF	ATCCAGTAATTTCTGACATTTCTGAG
Reverse primer	cgd5SeqR	TAATTAACATTTTCAGGGTCTCTGG
Forward primer	cgd6SeqF	GACTTGGATTTGGACTTACACC
Reverse primer	cgd6SeqR	TCCAAATGATGTAAATACTCCAGC
Forward primer	cgd8SeqF	CCATATTGAATGCAATGCCAAATC
Reverse primer	cgd8SeqR	GCATGGATTAACGACCAGTTG
Forward primer	MM19SeqF	GCGGAGAAGGAGGATTTAATTC
Reverse primer	MM19SeqR	TTACAACTCCAACTCCACCAC

Assay design - PCR reaction mix: Each locus is run as a simplex PCR reaction

Component	Specific reagent (Supplier)	Volume (µl)
10x Reaction Buffer	HotStar Taq Buffer (Qiagen)	2
2 mM dNTP	(Biogene)	2
5 mg/mL BSA	Non-acetylated (Sigma-Aldrich)	2
10 µM Forward Primer	(IDT)	0.4
10 µM Reverse Primer	(IDT)	0.4
5 units/µL Taq	HotStar Taq (Qiagen)	0.5
Template DNA		2
nfH_2O	Any nuclease free / molecular grade	10.7
Total Volume:		20

Reaction conditions:

Stage	Temperature	Duration (sec)	Comments
Hold (Hot start)	95 °C	900	
Cycling (34 cycles): Denaturation Annealing Extension	 94 °C 58 °C 72 °C	 30 90 60	No final extension used.
Hold	4 °C	∞	

Sequencing:

- Bi-directional
- Align to reference library for each specific locus (see pages 5 to 7)
- Count repeats as described below

Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

Interpretation:

The allele is identified by the number of repeat units counted in the consensus sequence.

Alignment of sequences (see table and examples below):
Example sequences from all loci have been submitted to GenBank (OM832408-OM832443) to aid in the production of a reference library alignment.

It is easier to see the repeat units if the sequences are translated into amino acids as the sequence length is reduced and the motifs are easier to visualise. This includes the cgd8 non-coding locus. The reading frame may need adjusting to see the listed amino acids.

Flanking sequences tend to be conserved (also more apparent when the alignment is translated to amino acids) helping to define the repeat region.

Identification of rare sequence anomalies:
cgd1_470_1429 – Upstream sequence variation has been detected in some samples that may result in non-amplification of the VNTR region with the MLVA primers. These samples tend to also have additional mini- or microsatellite repeats upstream of the VNTR region.
MSF – gp60 subtype IIcA5G3a isolates may contain a TQEG and a PQEG amino acid repeat (see 4 rpt in MSF figure below)
cgd6_4290_9811 – Rare absence of the "SL" motif in the "LSL" start flanking sequence results in a sizing of 20 repeats, instead of the true 22.
MM19 – Some isolates seem to have an unusual "GV" motif within the repeat region that sometimes generates dual peaks. Dual peaks at 12 + 28 repeats (see IIcA5G3a in MM19 figure below) and 39 + 46 repeats have been detected so far. Where dual peaks are detected in MM19 sequence confirmation may be warranted.

Description of the repeat features at each locus for when confirming the number of repeat units:

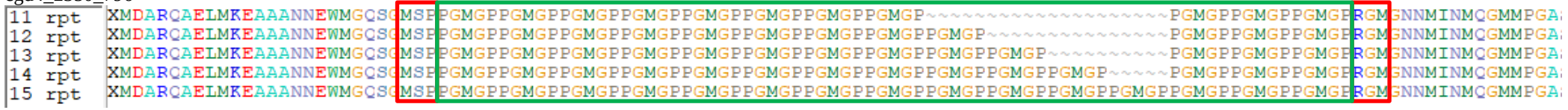
Locus	Start Flanking Seq [Amino Acids]	Repeat Unit (size in bp)	Repeat Unit Amino Acids	Stop flanking Seq [Amino Acids]
cgd1_470_1429	GGTTCTAGT [GSS]	TC(T/G)GAT (6 bp)	SD	TCTAATTCT [SNS]
cgd4_2350_796	ATGAGCCCG [MSP]	CC(T/C)GGTATGGG(T/C)CC(A/G) (15 bp)	PGMGP	CGCGGAATG [RGM]
cgd5_10_310 (MSF)	GTTAAGGAAGGA [VKEG]	GCTCAGGAAGGA (12 bp)	AQEG	CCTGTAGAG [PVE]
cgd5_4490_2935*	GAACGGAAC [ERN]	(C/A)A(A/G)AGC (6 bp)	(Q/K)S	AAAAACCAG[KNQ]
cgd6_4290_9811	TTATCATTA [LSL]	TC(T/C) (3 bp)	S	ACTTCTTCT [TSS]
cgd8_NC_ 4440_505	CTTTGGCTT [LWL]	GAGC(T/C)T (6 bp)	E(L/P)	AAGCTTAAA [KLK]
cgd8_4840_6355 (MM19)	AATTCAGTA [NSV]	GGA(G/T)C(A/T/C) (6 bp)	G(A/S)	GGAGTTGGT [GVG]

Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

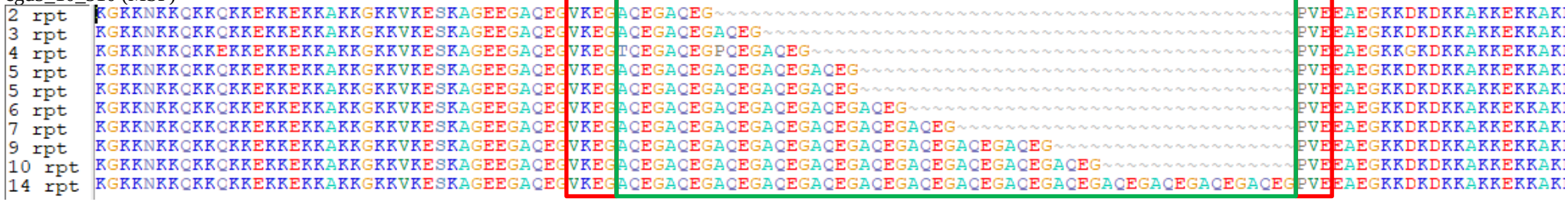
Examples of how the amino acid sequences align in the reference libraries used to confirm the number of repeat units:

cgd1_470_1429

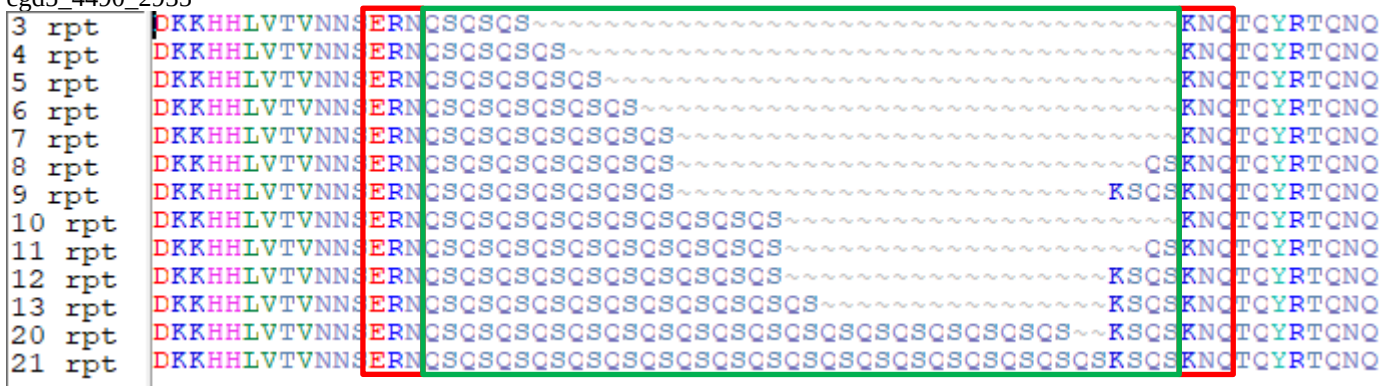
cgd4_2350_796



cgd5_10_310 (MSF)

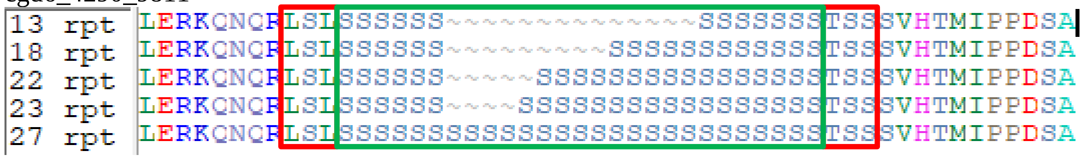


cgd5_4490_2935*

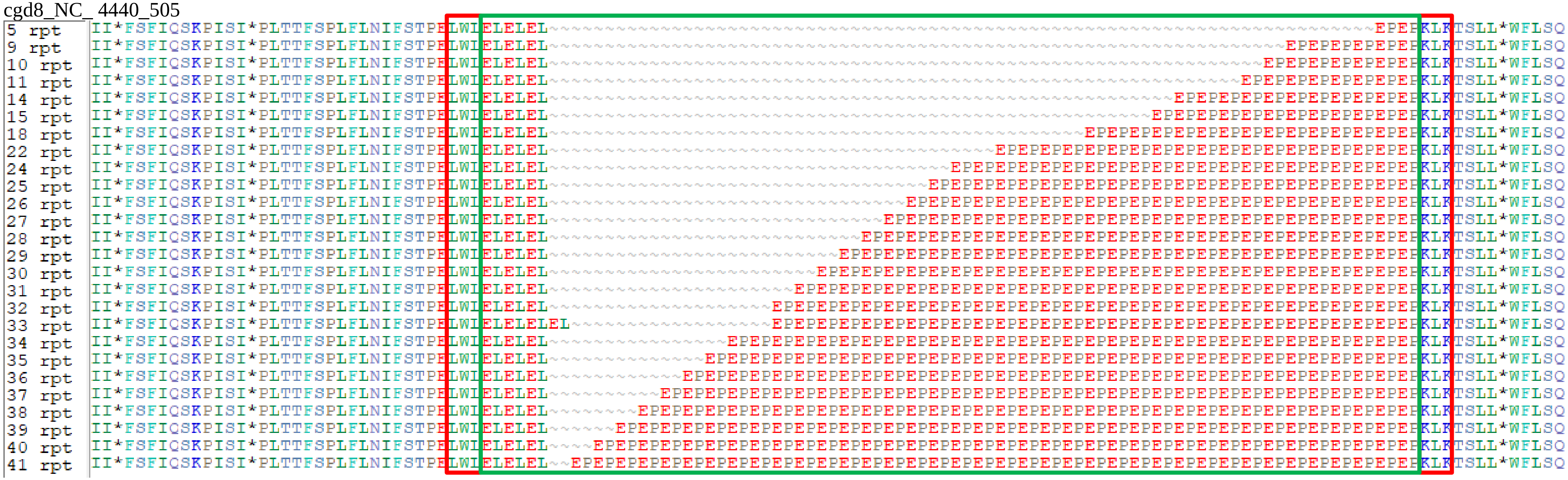


Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

cgd6_4290_9811



cgd8_NC_ 4440_505



Tel: +44 (0)1792 285341
Email: Swansea.Crypto@wales.nhs.uk

cgd8_4840_6355 (MM19)

```
4 rpt     IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGSGVGVGVGVGAGVGSGV
8 rpt     IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
13 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
14 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
15 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
16 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
17 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
18 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
19 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGSGVGVGVGVGAGVGSGV
19 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
20 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
21 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
23 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
24 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
25 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
26 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
27 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
28 rpt    IVPGIFDSATINSV~~~~~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
30 rpt    IVPGIFDSATINSV~~~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
31 rpt    IVPGIFDSATINSV~~~~~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
33 rpt    IVPGIFDSATINSV~~~~~~GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
36 rpt    IVPGIFDSATINSVGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGSGAGAGAGSGAGAGSGAGAGAGSGVGVGVGVGAGVGSGV
IIcA5G3a  IVPGIFDSATLNSV~~~~~~~~~~~~~~~~GVGAGSGSGAGSGSGAGAGAGVGSGAGVGVGVGAGVGAGVGVGAGVGAGAGAGVGSGAGVGVGVGAGVGAGV
```

References:

Applied Biosystems. User Guide: DNA Fragment Analysis by Capillary Electrophoresis, 2014; Publication number 4474504, Revision B, page 104.

Robinson G, Pérez-Cordón G, Hamilton C, Katzer F, Connelly L, Alexander CL, Chalmers RM. Validation of multilocus variable number tandem repeat analysis (MLVA) as a subtyping scheme for *Cryptosporidium parvum*. Food and Waterborne Parasitology 2022; 27, e00151, https://doi.org/10.1016/j.fawpar.2022.e00151

Risby H, Robinson G, Chandra N, King G, Vivancos R, Smith R, Thomas D, Fox A, McCarthy N, Chalmers RM. Application of a new multi-locus variable number tandem repeat analysis (MLVA) scheme for the seasonal investigation of *Cryptosporidium parvum* cases in Wales and the northwest of England, spring 2022. Current Research in Parasitology & Vector Borne Diseases 2023; 4, 100151. https://doi.org/10.1016/j.crpvbd.2023.100151