Driving forward genotyping of foodborne protozoan parasite species

COST Short Term Scientific Mission Report

13th- 17th March 2017

Dr Guy Robinson, Cryptosporidium Reference Unit, Public Health Wales

Host Investigator and Institution: Dr Karin Troell, National Veterinary Institute, Uppsala
Dr Jessica Beser, Public Health Agency for Sweden, Stockholm

Introduction

Subtyping foodborne protozoan parasite species for epidemiological purposes is generally not sufficiently discriminatory and has not been standardised. For example, only the single locus gp60 sequencing tool is comparable across laboratories for differentiating within *Cryptosporidium* spp., but a single locus method is problematic when sexual recombination is an integral part of the life-cycle. Additionally, although the gp60 genotyping method is used worldwide, there have been variations that have crept into the nomenclature fuelling confusion and name duplication. During the COST funded EURO-FBP workshop on "working towards a consensus on genotyping schemes for surveillance and outbreak investigations of *Cryptosporidium*" in Berlin in June 2016 (Chalmers and Caccio, 2016), the outcomes included the need to identify and validate suitable markers for a multi-locus scheme, and compare computational analytical approaches. Several researchers at the meeting had been working towards developing markers, some based on variable numbers of tandem repeats (VNTR) with a view to fragment sizing or sequencing, and others characterising single nucleotide polymorphisms (SNPs) using sequencing. The aim of a robust, harmonised subtyping scheme would be for the same outcome by either method, allowing participants to use their most amenable platform.

The Cryptosporidium Reference Unit (CRU) has recently undertaken the systematic discovery and evaluation of new VNTR markers for the discrimination of *Cryptosporidium parvum* by fragment sizing (Perez et al., 2016). The validation of seven VNTR markers across five chromosomes is currently underway. Simultaneously, the National Veterinary Institute (SVA) and Public Health Agency of Sweden (FoHM) have, through a joint project, evaluated seven SNP markers across four chromosomes, which has been shown to have high discriminatory power and provided evidence of transmission between calves and humans in Sweden (unpublished data). Each group used their own recently generated whole genome sequences for marker discovery and are further developing technologies for whole genome sequencing of *Cryptosporidium* and other foodborne parasites such as *Giardia* and *Cyclospora*.

Aims and Objectives

To strengthen the network initiated at the Berlin workshop and to foster collaborations between the groups with mutual interests but differing approaches, this short term scientific mission (STSM) was undertaken from the CRU to Dr Troell's lab at SVA, Uppsala and FoHM, Stockholm. The contemporary importance of this STSM is to explore the most appropriate way forward for

genotyping of *Cryptosporidium* and other foodborne protozoan parasites. The visit took place across two sites and involved expert researchers from all three organisations.

Specific objectives included:

1. allow a multi-attribute assessment and comparison of VNTR fragment sizing and SNP-based sequence analysis
2. assess the broader utility of SNP analysis originally designed for Swedish *Cryptosporidium parvum* with a range of gp60 subtypes from human and animal samples from the UK and Ireland; can they be differentiated by either method (MLFT and SNP analysis)
3. facilitate exchange of laboratory and analytical methods for the purposes of technology transfer including the bioinformatic analysis of multilocus data
4. automate analysis of the gp60 sequences using the Swedish bespoke computer program and investigate the expanding nomenclature of gp60 subtyping including a web-based platform to enable a standard approach and avoid naming confusion and duplications
5. explore the use of specific bioinformatic pipelines in the analysis of NGS data from protozoan foodborne parasites
6. use *in silico* analysis of whole genome sequences to compare variation by different markers

In addition to these objectives, the STSM enabled several meetings between staff at the SVA, FoHM and CRU to discuss wider aspects of research into foodborne parasites, particularly *Cryptosporidium*, including clarification of methods and techniques, sharing of experiences and brainstorming sessions.

Comparison of VNTR and SNP-based genotyping methods

Methods and Materials

Samples for initial SNP and VNTR method comparison

As the SNP method was developed specifically to investigate the common genotypes of *Cryptosporidium* found in Sweden, 20 *C. parvum*-positive isolates were selected from the national archive at the CRU to compare with the results generated by the VNTR method. The isolates were selected in contrast to the Swedish situation and to challenge their method. The panel was made up of anonymised isolates from sporadic and outbreak cases, some of which were epidemiologically linked and others completely geographically and temporally separate (Table 1).

Oocysts were semi-purified by saturated salt flotation and DNA extracted using QIAamp DNA mini kit (Qiagen) as described previously (Method 1 in Elwin et al., 2012). All samples were previously characterised by sequencing the gp60 gene following amplification using a primer cocktail described in Chalmers *et al.* (2016a). DNA samples were sent in advance of the STSM to SVA for processing to enable data analysis to be undertaken during the visit.

CRU VNTR analysis

DNA from each of the 20 *C. parvum* samples was amplified at seven loci spread across five chromosomes (1_470_1429, 4_2350_796, 5_4490_2941, 6_4290_9811, 8_4440_NC_506, cgd5_10_310 [also known as MSF] and MM19), that were previously identified as showing potential for use in a multi-locus VNTR typing scheme (Perez et al. 2016, Chalmers et al., 2016a). The PCRs for these markers have been developed and evaluated at the CRU, but the details of which have yet to be published and so are not shown in this report. To obtain the amplicon sizes, the PCR products were run through a high resolution cartridge on a QIAxcel capillary gel electrophoresis platform (Qiagen). Fragment sizes were converted to the number of repeat units based on *a priori* knowledge of the DNA sequences of previous samples. Multi-locus genotypes (MLGs) were determined for each sample by comparison of the number of repeat units present at each locus.

Table 1. Details of the 20 *C. parvum* samples tested by the SNP and VNTR methods.

Reference	gp60 genotype	Description
UKP90	IIaA18G3R1	Human sample from part of Outbreak A - a common Irish subtype
UKP91	IIaA20G5R1	Human sporadic case identified from Galway
UKP92	IIaA19G4R1	Human sample from part of Outbreak B
UKP93	IIaA18G2R1	Lamb sample from outbreak C
UKP94	IIaA18G2R1	Human sample from outbreak C
UKP95	IIaA19G1R1	Human sample from outbreak D
UKP96	IIdA15G1	Human sporadic case identified from Barnsley
UKP97	IIdA19G1	Human sporadic identified from Gateshead , but this gp60 subtype has previously shown different MLST from different geographical locations (Wang et al., 2014)
UKP98	IIdA21G1	Human sporadic case identified from Leicester
UKP99	IIcA5G3a	Human sporadic case identified from HIV +ve patient from Nottingham
UKP100	IIcA5G3j	Human sporadic case identified from Manchester
UKP101	IIaA15G2R1	Lamb sample from part of outbreak E – a common gp60 genotype
UKP102	IIaA15G2R1	Human sample from part of outbreak E (onset date 07/03/16) – a common gp60 genotype
UKP103	IIaA15G2R1	Human sample from part of outbreak E (onset date 24/03/16) – a common gp60 genotype
UKP104	IIaA15G2R1	Human sample from part of outbreak F (Sample date 12/04/16) – a common gp60 genotype
UKP105	IIaA15G2R1	Human sporadic case initially thought to be part of outbreak F, but subsequently removed from case list as epidemiology showed that was not part of outbreak (Sample date 18/05/16) – a common gp60 genotype
UKP106	IIaA15G1R2	Human sample from part of mixed outbreak G – each gp60 subtype (IIaA15G1R2 & IIaA17G1R1) has the same length microsatellite and only differs by a single SNP in the microsatellite to create the two different gp60 subtypes and an additional SNP downstream. The similar gp60 genotype IIaA15G2R1, from outbreaks F & G, has a longer microsatellite and additional SNPs compared to these.
UKP107	IIaA17G1R1	Human sample from part of mixed outbreak G – each gp60 subtype (IIaA15G1R2 & IIaA17G1R1) has the same length microsatellite and only differs by a single SNP in the microsatellite to create the two different gp60 subtypes and an additional SNP downstream. The similar gp60 genotype IIaA15G2R1, from outbreaks F & G, has a longer microsatellite and additional SNPs compared to these.
UKP108	IIdA24G1	Human sample from a national outbreak (H) from Tameside (NW England), a different region to UKP109. This subtype also caused an outbreak in Sweden.
UKP109	IIdA24G1	Human sample from a national outbreak (H) from Canterbury (SE England), a different region to UKP108. This subtype also caused an outbreak in Sweden.

Swedish SNP analysis

DNA from all 20 samples was amplified at seven loci (E1, E2, E3, E4, E5, E6, E7), across four chromosomes. This method was developed and evaluated at the Swedish agencies, and is yet to be published so the details are not shown in this report. Amplicons were sequenced using NGS technology on an Ion Torrent platform (Thermo Fisher Scientific) and SNPs identified using CLC Genomic Workbench (Qiagen).

Comparative analysis between SNP and VNTR results

Initially, the comparative performance of the genotyping methods was described in terms of the number of differences identified and how the differentiation aligned with the epidemiological data. The sequence-based SNP analysis also allowed for construction of a neighbor-joining tree with Jukes-Cantor correction and 100 bootstrap replicates using CLC Genomic Workbench following concatenation of all seven loci. To compare the number of differences between isolates typed by each method and how they separated the samples, minimum spanning trees with crosslinks were produced in Bionumerics (Applied Maths). Finally, the two methods were compared in relation to factors that the scientific community value from a genotyping scheme specifically for the purpose of *Cryptosporidium* surveillance and outbreak investigation. To achieve this, a multi-attribute analysis was started during the STSM using nine specific factors (hands on time, specialist equipment, cost of consumables, typability, discriminatory power, biological robustness, expertise, turn-around time, portability) that the expert participants of the COST workshop in Berlin (June 2016) were asked to rank in order of importance. By using their answers it is possible to put a numeric weighting on each attribute to allow an objective comparison to be undertaken. Some of the information to complete this is still being generated, but an initial table was completed.

Results and Discussion

Comparison and multi-attribute assessment of VNTR and SNP methods

Both methods have shown promise in the settings that they were designed for, but have not really been tested outside of those parameters. This STSM allowed us to perform both methods on an initial sample set containing a variety of samples found in the UK. This challenged the SNP method with samples from outside of Sweden and allowed the VNTR method, which is still in the process of being evaluated, to be tested with a selection of related and unrelated samples. The results of the two methods are shown in Table 2, but caution must be taken during the interpretation due to the limited number of samples tested. The positivity was superior in the SNP method with all samples giving a full set of results at each locus compared to the VNTR method which failed to amplify a product in 5 of the PCRs. However, the discrimination was sufficiently good in the other amplified loci for these samples that it was still possible to assign separate MLGs.

The Swedish SNP method produced 11 unique MLGs from this sample set compared to 17 in the VNTR assay. The VNTR method separated all of the different gp60 genotypes that were included. The SNP method separated most of the different gp60 genotypes, with the exceptions of UKP91

Table 2. Results of the 20 *C. parvum* samples by each method showing the variation at each locus and the multi-locus genotype (prefixed SNP_ or VNTR_). Each unique multi-locus genotype is numbered from 1 onwards based on this study alone as the final nomenclature for genotypes produced by these methods is still to be decided upon.

Sample	gp60	Swedish SNP method								UK VNTR method							
		MLG	E1	E2	E3	E4	E5	E6	E7	MLG	1_470_1429	4_2350_796	cgd5_10_310	5_4490_2941	6_4290_9811	8_4440_NC_506	MM19
UKP90[A]	**IIaA18G3R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_1**	4	16	5	7	3	35	15
UKP91	**IIaA20G5R1**	**SNP_4**	1	2	3	5	3	7	8	**VNTR_2**	4	15	5	6	2	31	16
UKP92[B]	**IIaA19G4R1**	**SNP_4**	1	2	3	5	3	7	8	**VNTR_3**	4	15	5	7	3	33	18
UKP93[C]	**IIaA18G2R1**	**SNP_11**	2	4	3	5	3	7	8	**VNTR_4**	4	14	5	7	3	23	16
UKP94[C]	**IIaA18G2R1**	**SNP_11**	2	4	3	5	3	7	8	**VNTR_5**	4	14	5	7	3	32	16
UKP95[D]	**IIaA19G1R1**	**SNP_10**	7	3	1	1	3	3	1	**VNTR_6**	5	14	3	20	2	10	20
UKP96	**IIdA15G1**	**SNP_5**	11	2	5	1	3	2	3	**VNTR_7**	5	15	6	Neg	2	11	21
UKP97	**IIdA19G1**	**SNP_2**	11	4	5	1	3	2	3	**VNTR_8**	Neg	16	14	6	Neg	11	28
UKP98	**IIdA21G1**	**SNP_1**	1	3	5	1	3	2	3	**VNTR_9**	Neg	15	3	4	3	10	23
UKP99	**IIcA5G3a**	**SNP_8**	6	2	1	1	1	6	2	**VNTR_10**	6	16	4	3	2	6	28
UKP100	**IIcA5G3j**	**SNP_2**	11	4	5	1	3	2	3	**VNTR_11**	5	14	5	Neg	2	10	21
UKP101[E]	**IIaA15G2R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_12**	4	14	5	7	3	33	15
UKP102[E]	**IIaA15G2R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_12**	4	14	5	7	3	33	15
UKP103[E]	**IIaA15G2R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_12**	4	14	5	7	3	33	15
UKP104[F]	**IIaA15G2R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_13**	4	15	5	8	3	32	15
UKP105[(F)]	**IIaA15G2R1**	**SNP_7**	2	2	3	5	3	7	8	**VNTR_14**	4	15	5	7	3	25	13
UKP106[G]	**IIaA15G1R2**	**SNP_6**	1	3	5	1	4	2	3	**VNTR_15**	5	14	3	13	2	10	27
UKP107[G]	**IIaA17G1R1**	**SNP_9**	7	3	5	2	4	2	3	**VNTR_16**	6	15	3	9	2	10	23
UKP108[H]	**IIdA24G1**	**SNP_3**	7	4	4	2	4	2	7	**VNTR_17**	6	14	3	12	2	16	4
UKP109[H]	**IIdA24G1**	**SNP_3**	7	4	4	2	4	2	7	**VNTR_17**	6	14	3	12	2	16	4

[A-H] Superscript letters indicate samples related to various outbreaks (see Table 1 for further description).

[(F)] Sample initially thought to be part of Outbreak F, but subsequent epidemiological information ruled it out of the investigation.

(IIaA20G5R1) from UKP92 (IIaA19G4R1), UKP97 (IIdA19G1) from UKP100 (IIcA5G3j), and UKP90 (IIaA18G3R1) from UKP101-105 (IIaA15G2R1).

As the SNP method is sequenced based it was possible to concatenate all of the loci sequences and produce a phlyogenetic tree (generated by Dr Olov Svartström) to compare the sequence relationship between all of the 20 samples (Figure 1). This tree clearly shows which of the samples cluster together and which were separated. It is interesting to see how most of the gp60 genotypes belonging to family IIa all cluster together on the top branches with the exception of IIa samples UKP106, UKP107 and UKP95 which all cluster with the IId samples. The IIcA5G3j sample (UKP100) also clusters with the IId samples, but the similar gp60 genotype IIcA5G3a sample sits separately from all the other samples on its own branch.

In order to compare the two methods it was also possible to produce minimum spanning trees by using the results at each locus as character data (Figures 2a and 2b). The minimum spanning trees display the number of variations between each multilocus genotype and sample. The colours in the trees represent the samples that are epidemiologically linked so it is easier to see which samples have been properly separated. The gp60 genotypes are also displayed to show the separation between seemingly closely related genotypes.

The separation of outbreak samples was better with the VNTR method with linked isolates from each outbreak grouping together (UKP102-105, UKP108-109) and separate from other isolates, even with the same gp60 genotype (UKP102-105 versus UKP106), whereas the SNP assay grouped all of the outbreak samples with the same gp60 genotype (IIaA15G2R1) together regardless of whether they were from separate outbreaks. Interestingly, the VNTR method separated UKP104 and UKP105, which initially thought to be part of the same outbreak and had the same gp60 genotype, but when the epidemiological questionnaire was returned for UKP105 they had not visited the open farm in question and were removed from the case list. As expected, both methods separated two samples from a mixed outbreak (UKP106 and UKP107) where two different gp60 genotypes were circulating.

Two samples from the same outbreak UKP93 and UKP94, one from a lamb and the other from a human case, which were grouped together using the SNP method, were separated by the VNTR method at a single locus. The significance of this is still unclear, whether this locus could be too discriminatory in linked samples, could the variation be driven by the host or how much variation should be allowed when looking for epidemiologically linked cases all still needs to be investigated.

The nine attributes for consideration in a genotyping scheme for the purpose of surveillance and outbreak investigation were tabulated and scored (Table 3). Based on the current knowledge about the two methods, the VNTR approach looks more suited for this epidemiological application, but it is too early to make a final judgement on this as there are still some factors to be fully assessed on a much large set of isolates. The 20 isolates that were included in this STSM originated from the UK collection and may introduce bias into the results as the VNTR loci were selected from genomes also from the UK collection, whereas the Swedish SNP method was specifically designed to separate common genotypes within the Swedish setting. It will be interesting to see the VNTR approach when challenged with samples previously characterised by the SNP method including those from the same incident but from different hosts, and also from unrelated incidents but seen to be conserved by the current gp60 genotyping and separated by SNPs.

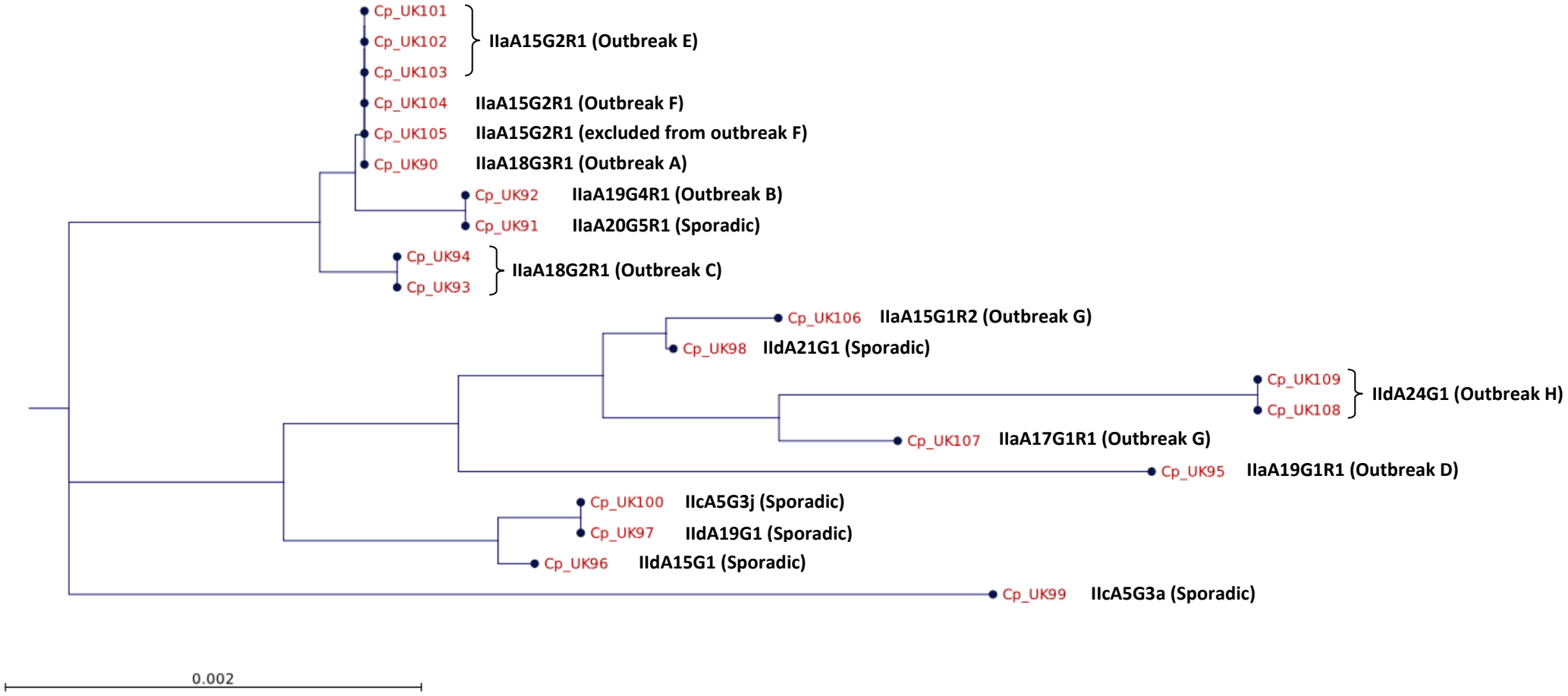
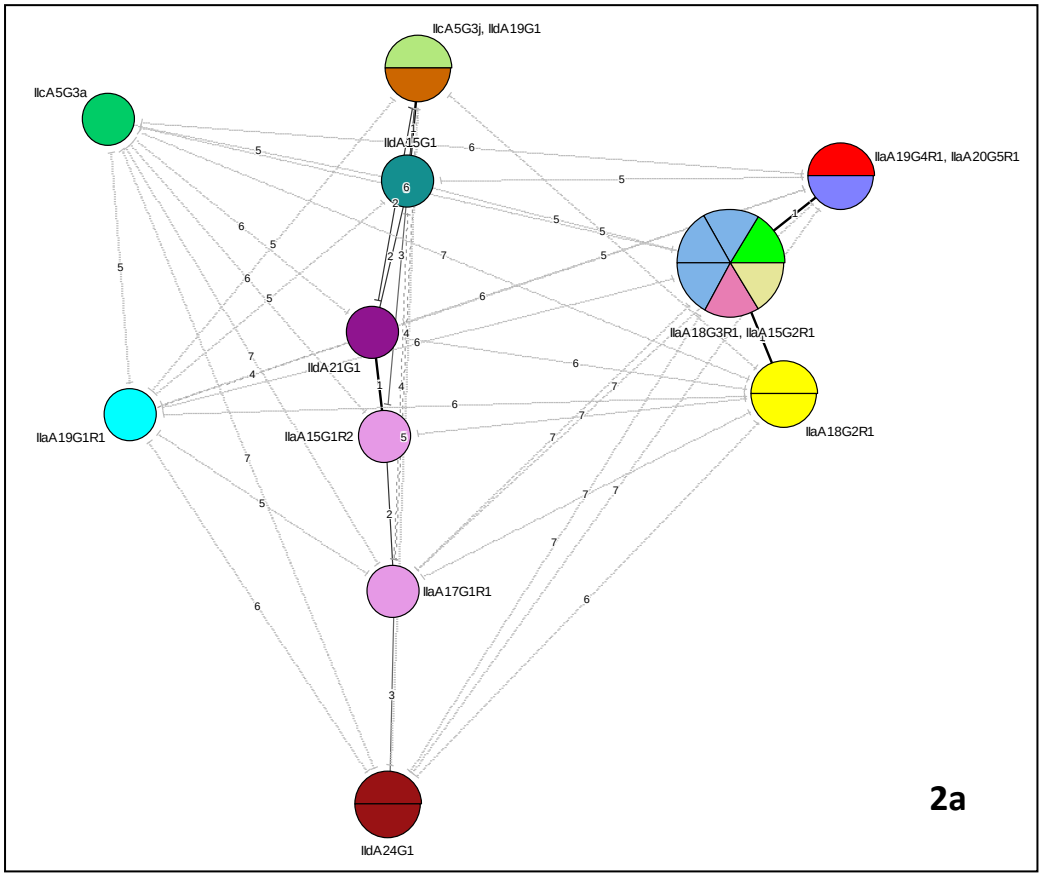


Figure 1. A neighbour-joining tree showing the clustering between the 20 samples based on the concatenated sequence of all 7 SNP loci.

Figure 2. Minimum spanning trees of the SNP (2a) and VNTR (2b) methods showing the amount of variation between each sample. The sample colours represent epidemiologically linked samples from the same outbreak investigations. Branch numbers reflect the number of variant loci.



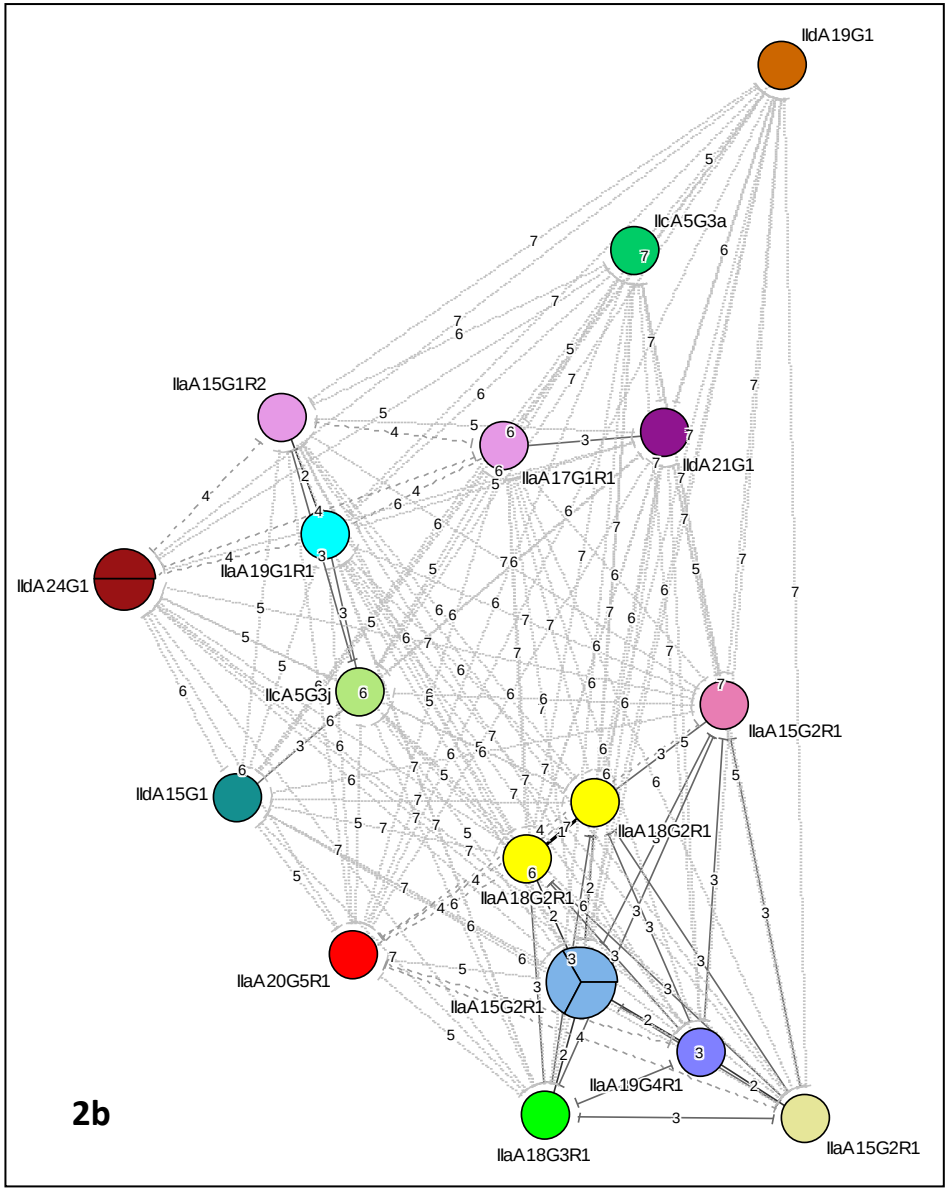


Table 3. Multi-attribute analysis based on the desirability weightings of nine factors of a genotyping scheme specifically for the purpose of surveillance and investigating outbreaks as determined from the opinions of 22 experts gathered at the June 2016 COST workshop in Berlin. Some information is still to be determined so the overall scores will change, but this current version is produced from the information gathered during the STSM.

Attribute	Rank [2=Best]		Desirability Weighting	Adjusted score (Rank x Weighting)	
	SNP Rank (Detail)	VNTR Rank (Detail)		SNP	VNTR
Hands on time	? - (awaiting information from FoHM)	? - (2 hours for 10 samples, 8-10 hours for 96 samples)	0.0600	?	?
Specialist Equipment	2 – (NGS library prep and sequencing platform, but can be outsourced)	2 - (Capillary gel electrophoresis, but can be outsourced)	0.0690	0.1380	0.1380
Cost of consumables	1 – (awaiting costs from FoHM, but on fewer samples cost is higher than VNTR)	2 - (~£20 per sample)	0.1930	0.1930	0.3860
Typability (no. of samples typable)	2 - (To be determined on more samples, but based on 20 sample panel 100% were typable)	1 - (To be determined, but based on 20 sample panel 100% were typable, although a few loci did not amplify)	0.0580	0.1160	0.0580
Discriminatory power	1 – (To be determined on more samples, but based on 20 sample panel = 11)	2 - (To be determined, but based on 20 sample panel = 17 MLGs)	0.2255	0.2255	0.4510
Biological robustness (inc. reproducibility and repeatability)	? - (Still to be determined)	? - (Still to be determined)	0.0540	?	?
Expertise	1 – (High level, bioinformaticians required to analyse data)	2 – (Medium level, fairly automated, but size-calling can be subjective)	0.0515	0.0515	0.1030
Turn-around time	1 – (awaiting information from FoHM, but processing in NGS platform takes longer than the PCR/Capillary Electrophoresis)	2 – (1 day for 10 samples, 3-4 days for 96 samples)	0.1235	0.1235	0.2470
Portability (for standardisation and comparability)	2 - (High, awaiting information from FoHM, but as based on actual sequence data standardisation and comparability is better than fragment sizing)	1 - (Medium, may need to adjust for platform variations based on sequenced reference types)	0.1655	0.3310	0.1655
			Overall score	1.1785	1.5485

Bioinformatic analysis of *Cryptosporidium hominis* IbA10G2 genomes

Previously, the genomes of 14 *Cryptosporidium hominis* IbA10G2 (by far the most common genotype in western countries) that were part of national excessive summer increase in 2015 showed very little variation across the whole genome with a maximum of 50 SNPs between the most different isolates and 0-3 between the most similar (Chalmers et al., 2016b). One thing that was not determined at the time was where the limited differences are within the genomes were and whether there are any themes between them that may be useful in subtyping during epidemiological investigations.

Methods and Materials

During this STSM I was able to link up with the bioinformatician, Dr Bjorn Hallström, who originally did the genome assemblies using SPAdes, to further explore where these SNPs were. Genomes were compared by aligning to a reference genome using Burrows-Wheeler Aligner and individual base variants were called using GATK HaplotypeCaller. I was able to use these individual variants as character data to produce minimum spanning trees with crosslinks in Bionumerics (Applied Maths) in order to look for trends between the isolates.

Results and Discussion

The re-analysis of these samples allowed the individual variations to be identified at 68 different locations across the genome (sequence data not show as being prepared by CRU for publication). This allowed me to do a large multi-locus analysis between these isolates and produce a minimum spanning tree to show the clustering and extent of variation between the cases (Figure 3). While there is some apparent clustering between some samples, there does not appear to be any linking themes between them. However, this is a fairly small subset of the hundreds of excess cases that were seen during this unusual increase.

Despite the low number of SNPs across the genomes it may be possible to use this variation data to identify particular SNP loci that could help differentiate IbA10G2 isolates, which is something that we will examine further. During the STSM, my hosts at SVA and FoHM showed me a SNP method that was designed to discriminate between IbA10G2 isolates, using the same approach as the *C. parvum* method compared above but with different markers, which they are currently in the process of preparing for publication. It would be interesting to use our IbA10G2 genome data to perform their method *in silico* and see how it compares to the analysis that we did with the whole genome variants. This is something that we aim do in the future.

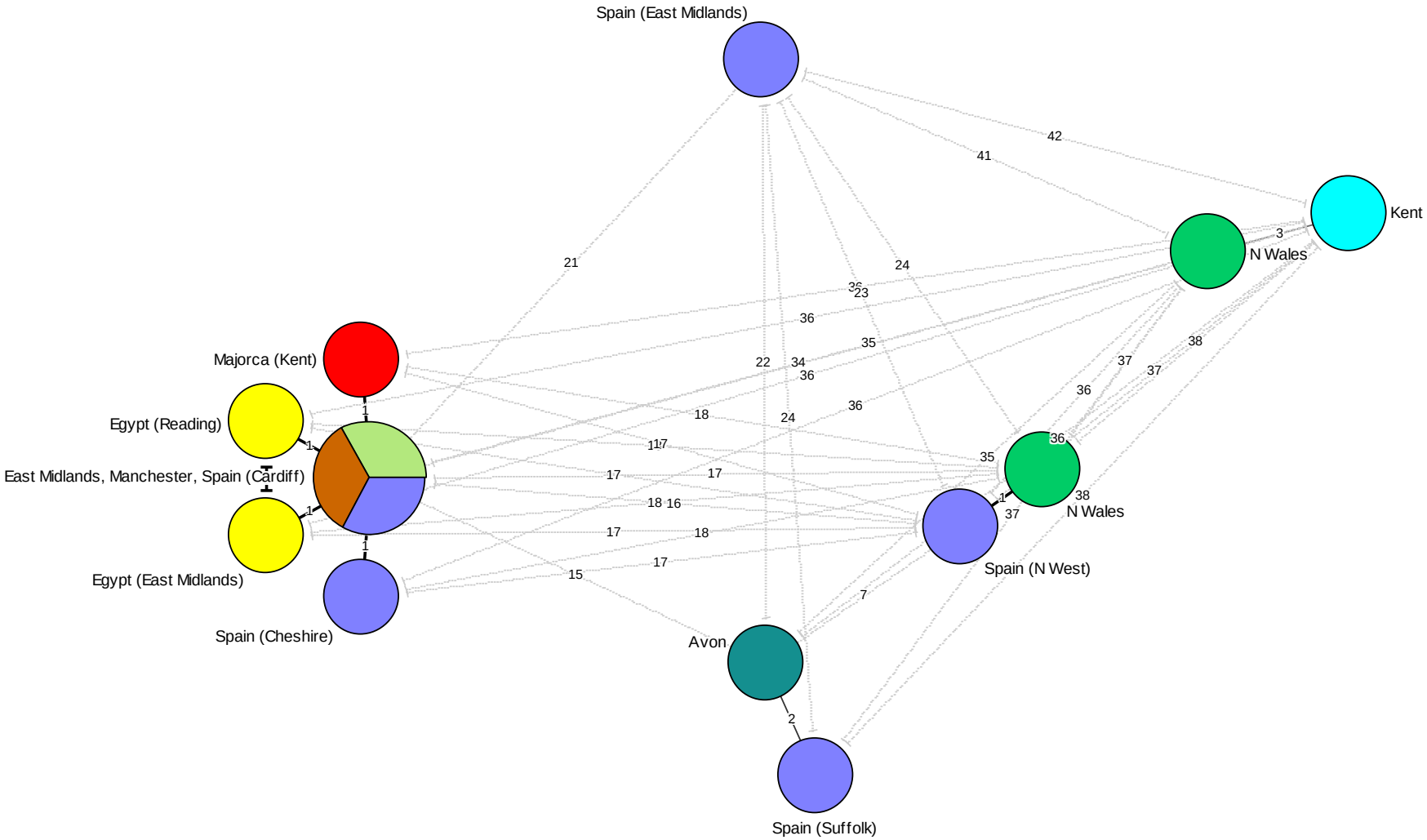


Figure 3. A minimum spanning tree of the SNPs across the genomes of 14 *C. hominis* IbA10G2 samples during the 2015 excessive summer increase. Colours and labels represent foreign travel or local region if no foreign travel reported. Branch numbers reflect the number of variant loci.

Bioinformatic analysis of *Cyclospora cayetanensis* genomes

During 2015 and 2016 there were widespread outbreaks of *Cyclospora* in the UK that were linked with travel to the Yucatan peninsula in Mexico. In 2015, subtyping was attempted at the CDC in Atlanta using their MLST method with limited success. In 2016, this option was not available to us and we decided to develop a method to enable whole genome sequencing from routinely submitted clinical samples, usually with limited material available and low numbers of parasites.

Methods and Materials

At the CRU we successfully purified and extracted *Cyclospora* DNA by salt floating the samples twice, sterilising the oocysts by bleach treatment, freeze-thawing and isolating the DNA by QIAamp DNA mini kit (Qiagen). *Cyclospora* DNA was boosted prior to NGS by multiple displacement amplification (MDA) using the REPLI-g Midi kit (Qiagen). Prior to the STSM we generated raw NGS data following DNA library preparation using a Nextera XT library preparation kit and sequencing on an Illumina MiSeq platform with a 2 x 250 bp V2 Illumina kit. This STSM allowed me access to the bioinformatician, Dr Robert Söderlund at SVA, who used my raw reads to assemble them *de novo* with SPAdes into a new draft *Cyclospora* genome.

Results and Discussion

We were able to generate a draft *Cyclospora* genome of 56 megabases in 24,474 contigs showing good quality parameters with a mean length of 2285bp and an N50 value of 6955. This is larger than the reference genome (45 megabases) so some of the contigs may be artefacts from contaminants. However, there was not sufficient time to do any further analysis of this genome, but the next step I will take is to compare it to the *Cyclospora* reference genome (C. cayetanensis CDC:HCNY16:01, accession LIGJ00000000) to see the quality of coverage that we were able to achieve from the challenging nature of *Cyclospora* samples that are generally received. This will hopefully act as a proof of principle for being able to compare *Cyclospora* genomes during future outbreak investigations.

Standardisation of *Cryptosporidium* gp60 nomenclature

The gp60 genotyping method is currently the gold standard and used globally. However, the complex nature of genotyping this gene has resulted in many using slight variations of the nomenclature resulting in confusion and results that cannot be properly compared. Additionally, the discovery of new genotypes by different groups at the same time has resulted in names being duplicated, further adding to the confusion.

As a member of the scientific advisory committee for the online genome resource CryptoDB, I raised a query with the group whether we could provide a gp60 genotyping tool that keeps a record of known genotypes, allocates new ones and allows a consistent use of nomenclature. One of the discussions we had was the programming for automated typing. I was aware of a piece of software

developed at FoHM for this purpose (currently unpublished) and this STSM allowed me to meet with its developers Marianne Lebbad, Erik Alm and Jessica Besser to discuss the potential use of the software and its integration and development on the CryptoDB platform. During the visit, Erik demonstrated the software and we all discussed this potential collaboration between CryptoDB and FoHM. The meeting was positive and I will be reporting back to the committee at the next meeting, hopefully resulting in the two groups collaborating on setting this up extremely useful resource.

Conclusions and future collaboration

This STSM in Sweden has successfully enabled initial comparisons between the two methods that our groups are developing, with both showing promise and probably usefulness in slightly different applications. While more extensive evaluation is still required, the VNTR method appears to be a good method for rapid and cost-effective genotyping in outbreak investigations. The SNP approach allows phlyogenetic and other sequence-based analyses to be performed and may be more useful in longer-term surveillance and population genetic studies. For the purposes of technology transfer, both groups were able to learn more about each other's approaches and share protocols. Following the additional testing that is required, a discussion paper is planned for publication.

Prior to the STSM, both groups had attended the COST expert workshop to discuss the future of genotyping for surveillance and outbreak investigations. At the meeting, the benefits of fragment sizing and sequencing were discussed but no real data was then available to back up the opinions. This STSM has enabled us to start generating the experimental data required to make those next decisions.

The availability of bioinformaticians to assist with *in silico* analysis of *Cryptosporidium* and *Cyclospora* genomes was invaluable and will allow us to further develop new genotyping methods. Likewise, discussions on the potential for an automated, web-based gp60 typing tool will hopefully result in a more standardised approach to use of this globally important method.

Some of the data generated as part of this STSM is planned to be included in some of the presentations being given later this month at the 6th International *Giardia* and *Cryptosporidium* Conference in Havana, Cuba.

In addition to the sections described in this report, we also held several other meetings during the week where other technologies (including routine typing algorithms and methods, typing from water samples, DNA fishing and other genome work) used at the CRU, SVA and FoHM were discussed and shared, with the seeds of several ideas being sown for future collaborations.

References

Chalmers R, Caccio S. (2016) Towards a consensus on genotyping schemes for surveillance and outbreak investigations of *Cryptosporidium*, Berlin, June 2016. Euro Surveill. 2016;21(37):pii=30338.

Chalmers RM, Robinson G, Hotchkiss E, Alexander C, May S, Gilray J, Connelly L, Hadfield SJ. (2016a) Suitability of loci for multiple-locus variable-number of tandem-repeats analysis of *Cryptosporidium parvum* for inter-laboratory surveillance and outbreak investigations. Parasitology 2: 1-11.

Chalmers R, Nichols G, Fina L, Williams C, Robinson G, Elwin K, McKeown P, Beser J, Kortbeek T, Fuentes I, Espinosa L (2016b) Report of the microbiological investigation of *Cryptosporidium*, late summer-early autumn 2015: support for multi-country outbreak assessment. Final report to ECDC, ID 4751.

Elwin K, Hadfield SJ, Robinson G, Crouch N, Chalmers RM. (2012) *Cryptosporidium viatorum* n. sp. (Apicomplexa: Cryptosporidiidae) among travellers returning to the United Kingdom from the Indian Subcontinent. International Journal for Parasitology 42: 675-682.

Pérez-Cordón G, Robinson G, Nader J, Chalmers RM. (2016) Discovery of new variable number tandem repeat loci in multiple *Cryptosporidium parvum* genomes for the surveillance and investigation of outbreaks of cryptosporidiosis. Experimental Parasitology 169: 119-128.

Acknowledgements

I would like to acknowledge the following people who helped generate the data presented in this report during the week I spent in Sweden. Firstly I would like to share my appreciation to Dr Karin Troell for hosting me at SVA and giving me the fantastic opportunities and stimulating scientific conversations that we had. I would also like to thank Dr Jessica Beser and Dr Romanico Arrighi at FoHM, Stockholm for their wonderful hospitality and discussions; Dr Olov Svartström, Marianne Lebbad and Erik Alm, also at FoHM for sharing their technical knowhow; Dr Robert Söderlund at SVA for his bioinformatic skills in assembling the Cyclospora genome; Dr Björn Hallström (ex-SVA) for his valuable assistance in assembling and analysing our *C. hominis* genomes to excavate the information that we requested; Arthur Morris (Aberystwyth University) for his help looking at the draft *Cyclospora* genome. Finally, thank you to the COST FBP group for funding and Prof. Rachel Chalmers (Public Health Wales) for giving the time available to be able to participate in this STSM.