HPRU GI Pump Priming Project 2022

Diversity within *Cryptosporidium parvum* for improved epidemiological understanding

Final report, January 2023, V3

Project team

Rachel Chalmers, Guy Robinson, Harriet Risby: national Cryptosporidium Reference Unit, Public Health Wales

Nastassya Chandra, Roberto Vivancos and Andrew Fox: UKHSA North West

Grace King, Robert Smith and Daniel Thomas: Communicable Disease Surveillance Centre, Public Health Wales

Noel McCarthy: University of Warwick and Trinity College Dublin.

Executive summary

Cryptosporidium parvum is an important zoonotic cause of gastroenteritis. During validation of a multilocus variable number of tandem repeats analysis (MLVA), a high proportion (79 %) of MLVA profiles were unique among sporadic cases, suggesting genetic clusters of *C. parvum* specimens might indicate unrecognised outbreaks. *C. parvum* specimens, 28th March to 31st July 2022, from the north west of England and Wales were tested by MLVA. MLVA profiles were used to identify genetic clusters of cases. Epidemiological / exposure data from routinely-collected case questionnaires were checked for risk factors and compared to other cases within genetic clusters.

A total of 213 specimens, 118 from north west England and 95 from Wales, were analysed and 161 MLVA profiles identified; 13 of these were mixed profiles. Of the remaining 148 simple profiles, 123 (83 %) were unique within this dataset. Seventy seven cases formed 25 clusters, ranging from 2 to 9 (mode = 2) cases.

The overall questionnaire response rate was 75% but varied; 77/118 (65%) in north west England and 82/95 (86%) in Wales. Links were identified by MLVA. For example, two cases had the same MLVA profile as those in a known outbreak in Wales; they were also found to have visited the open farm so were included in the outbreak. One cluster was newly discovered to be linked to an open farm, north west England. An incident management team was convened. Nine cases had visited the farm, where failings were identified and which STEC cases had also visited.

Outcomes and impact

We have found MLVA to be a non-disruptive addition to the Cryptosporidium Reference Unit's genotyping workflow, and MLVA can be successfully applied with a fast turnaround time.

Known outbreaks were characterised and additional, epidemiologically linked but previously missed cases were identified by MLVA, strengthening evidence in outbreak investigations. Unrecognised clusters of cases can be identified by MLVA for further investigation and public health action.

We conclude that systematic analysis of *C. parvum* specimens by MLVA should be applied for all of England and Wales, and this change in practice will be progressed from spring 2023 subject to developing mechanisms for cluster reporting to provide context to systematic data capture in surveillance systems.

Introduction

The protozoan parasite *Cryptosporidium* is an important cause of diarrhoea and associated gastrointestinal symptoms. Up to 6000 laboratory-confirmed cases are reported *per annum* in the UK. The parasite is spread via faeces, and is resistant to chlorine disinfection. Outbreaks of cryptosporidiosis can occur and they may have a high impact as they can be widespread or lead to intrusive interventions such as the closure of swimming pools and open farms, formal notices for improvement in food production, or notices to boil water for drinking. Outbreaks also present a high risk of onward community transmission of illness.

Individual cases and outbreaks of cryptosporidiosis are under-ascertained and under-reported, with the number of estimated cases is higher than those of *Salmonella*, and the true extent of outbreaks may not be known. One approach to improving outbreak identification and extent might be to apply multilocus genotyping on a large scale to *Cryptosporidium* specimens to spot clusters of cases caused by genetically-related isolates.

A suitable multilocus variable number of tandem repeats (VNTR) analysis (MLVA) scheme for subtyping the zoonotic species *Cryptosporidium parvum* was validated in 2021 at the national Cryptosporidium Reference Unit (Robinson *et al*., 2022). Historical samples from un-related, sporadic cases and from outbreaks were tested, as well as ones from livestock. In 2021 and early 2022, the MLVA scheme was piloted in two outbreaks in real-time, providing useful public health information to strengthen associations with a suspected food source in one instance (Gopfert *et al*., 2022) and refining epidemiological analysis in another. This provided the evidence for the utility of MLVA in outbreak investigations and its application for this purpose is now established.

It is also likely that there is public health benefit to be gained from mass application of MLVA to identify clusters and outbreaks that might otherwise be missed. Historical data from the validation phase indicated that epidemiologically-unrelated *C. parvum* isolates displayed high diversity (the majority of MLGs, 79%, were unique) indicating that clusters of MLGs identified by MLVA might indicate unrecognised outbreaks (Robinson *et al*., 2022). It is this approach that the HPRU GI pump priming project aimed to investigate.

Primary Objective:

Explore whether the genetic diversity of *C. parvum* MLGs identified by MLVA can identify epidemiological clusters for further investigation at a local and/or national level.

The hypothesis is that truly un-related isolates have greater diversity than epidemiologically-related ones.

Secondary Objectives:

1. inform whether a future sentinel surveillance structure for *C. parvum* is possible and useful in England and Wales

2. assess whether the diversity of *C. parvum* MLGs defined by MLVA indicate that pursuing the application of NGS-based methods such as DNA hybridisation (baits capture technique) would be fruitful.

Methods

Figure 1 provides an outline of the workflow for the project. The detailed methods are described below and in Appendix 1: Laboratory Methods.

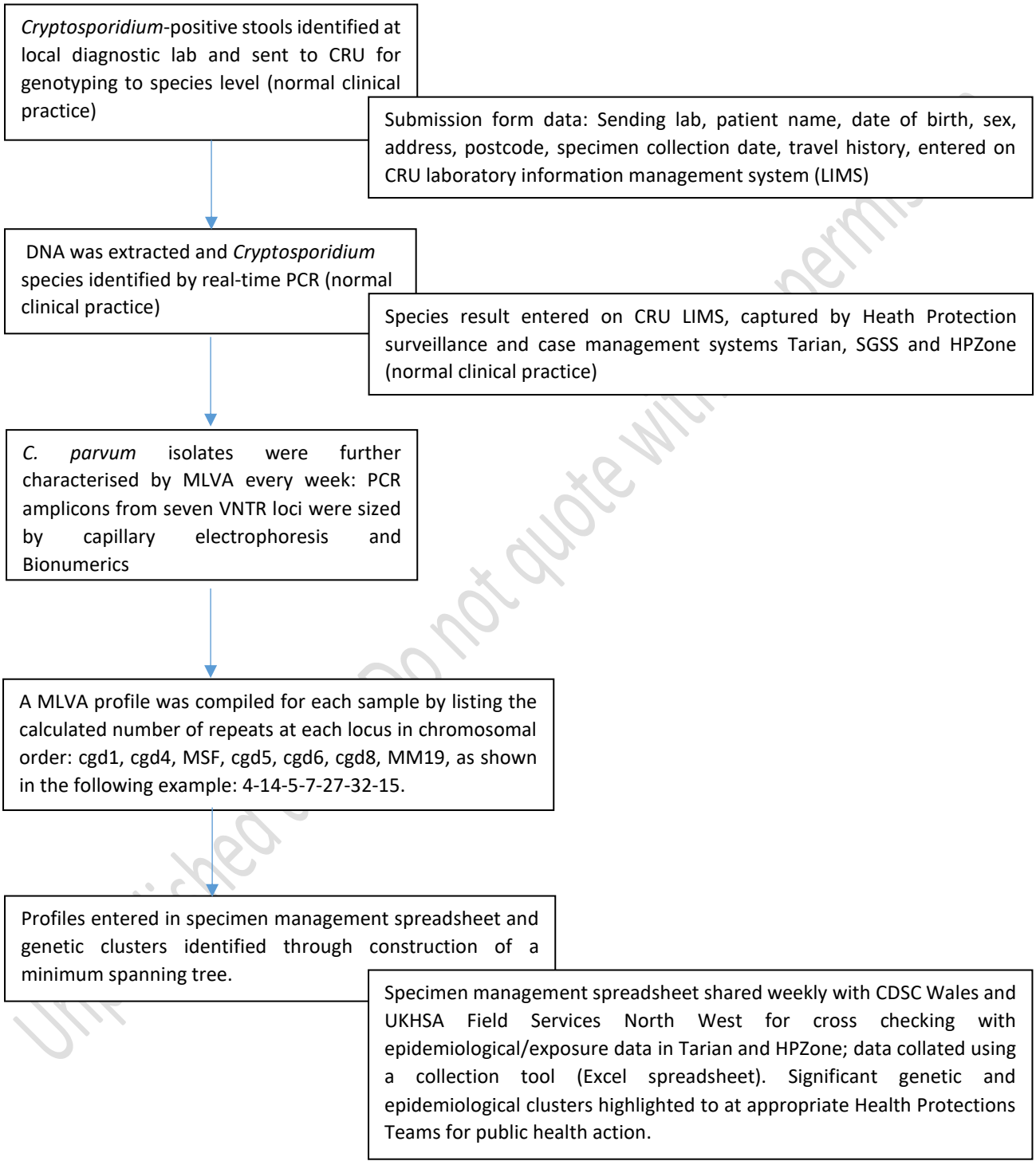

Figure 1. *Cryptosporidium parvum* diversity project workflow

Data transfer and exposure/epidemiological data capture

Excel spreadsheets containing patient identifiable data (name, specimen reference number), *Cryptosporidium* species and MLVA profiles were shared securely every week with the UKHSA Field Services north west senior epidemiologist (Nastassya Chandra, north west England data only) and Public Health Wales epidemiologist (Grace King, Wales data only). Genetic clusters of cases, defined by identical MLVA profiles at all seven loci, were highlighted.

A data capture tool was developed in Excel to record information across common variables (Table 1) from all routinely administered gastrointestinal-illness / *Cryptosporidium* case questionnaires in the north west of England and Wales from HPZone and Tarian respectively. Efforts were made to take account of the differing questions administered. Any missing questionnaires were chased after 7 days.

Table 1. The documentation and classification criteria in the data capture tool for each case.

Variable	**Explanation**
Case / sample identifier	For the purposes of populating the data extraction tool, the original laboratory number and, in the north west HPZone number and in Wales the Tarian number will be used as the unique identifier
Diagnostic laboratory	Name of reporting lab
Cryptosporidium species	Only those confirmed as *C. parvum* are of interest to this study
Region	Wales / Cheshire and Merseyside / Cumbria and Lancashire / Greater Manchester
Local authority	Lower tier local authority of residence, based on mapping from resident postcode
Resident postcode	Resident postcode
Questionnaire uploaded	Added a comment for all cases whether a questionnaire was uploaded to Tarian or HPZone
i-log or outbreak name	To identify cases considered as part of incident/outbreak
Age	Age at time of specimen date. Calculated from reported date of birth.
Sex	Male or Female
Ethnicity	Based on ethnic group categories in the 2021 census of England and Wales
Household case	To determine if case is likely a primary or secondary case Anyone in household ill with GI symptoms in the 14 days prior to case's illness? If yes, details of house household contacts. Contact with anyone outside household who was ill with GI symptoms in the 14 days prior to case's illness?
Sexual contact with someone who was ill or with GI symptoms	To determine if secondary case from sexual contact.
A man who has sex with other men (GBMSM)	Men who have sex with men are at high risk for sexual transmission of enteric pathogens. This was captured in England questionnaires only.
Onset date	Date of symptom(s) onset (estimated or exact)
Symptoms	Yes/No options for the following symptoms: abdominal pain, blood in stools, diarrhoea, fever, vomiting, nausea. Other symptoms – to describe.
Duration / still ill	Duration of illness (days) and still ill or not at time of questionnaire
Hospitalisation	Was the case admitted to hospital for this illness, if yes, which hospital. Questionnaires from Wales did not capture this question, however sometimes captured in Tarian.

Venue of work / daycare / education – address and postcode	Suggestion: classify as linked if same venues identified. Name and address of work Attendance at nursery, playgroup, school or college – name and address of education setting attended
Venue of food consumption	Suggestion: classify as linked if same venues identified. Venues food or drink consumed away from the home in the 14 days before illness.
Venue of visits/ activities	Suggestion: classify as linked if same activities identified
Food history	Limited detail about food consumption. Wales questionnaire captured information on seafood, salad, fruit, milk products from farm, pasteurised/unpasteurised dairy products, ready made sandwiches.
Raw milk consumption / source	
Pasteurised milk source	
Water supply / consumption.	Suggestion: classify as linked if same private water supply identified Private, water from a river, stream or spring, bottled water, disruption to mains supply.
Swimming and other recreational activities	Suggestion: classify as linked if same swimming pool identified Use of swimming, spa or paddling pools with details. Participating in freshwater activities, seawater activities, outdoor pursuits or field sports.
Animal contact and farm visits or living	Contact with pets at home or outside of the household – dogs, cats, birds, reptiles, rodents, other. Did any of the pets have diarrhoea before you became ill? Does your work involve contact with animals or faeces? Do you live on a farm or small holding? Did you visit a farm/stable/zoo in the 14 days before illness – provide details? Follow up questions: Did you handle or touch animals at the farm/stables/zoo – cattle, poultry, pigs, sheep, rodents, others. Were their handwashing facilities available? Suggestion: classify as linked if same farm/stable/zoo identified
Travel overseas or in UK	Suggestion: classify as linked if same country or location in UK identified Travel overseas in the 14 days before becoming ill? Date left and returned, country attended, hotel/place of residence abroad. Did you spend time away from home bit with the UK 14 days before becoming ill? Date left and returned – places cited.
MLVA profile	Classify as linked if sample MLVA profile identified
Comments	Any relevant comments from questionnaire

Inclusion Criteria

C. parvum DNA from stool samples sent from cases diagnosed by laboratories in Wales and the north West of England, received at the CRU from 28th March to 31st July and exposure information from these cases where a questionnaire has been completed. Mixed species included if *C. parvum* is present.

Exclusion Criteria

Cryptosporidium DNA with single species other than *C. parvum*; insufficient DNA for MLVA; outside study area (as indicated by submitting lab or resident LA); lack of minimum demographic data in surveillance records: sex, date of birth, onset date, postcode; samples received by CRU before 28th March 2022 or after 31st July 2022.

Laboratory Data Analysis

Specimen submission rates were calculate from laboratory notifications.

MLVA typability was defined as VNTR alleles identified at all seven loci.

Complex MLVA profiles (those with >1 allele at any locus) were excluded from statistical and graphical analyses where their inclusion would have necessitated the construction of hypothetical profiles.

The variability of the VNTRs was described and Simpson's indices of diversity for MLVA was calculated according to Hunter and Gaston, 1988. The Hunter Gaston Diversity Index (HGDI) was calculated for each locus individually and for the combined scheme for all non-outbreak specimens, using alleles from a single representative of each profile identified in outbreak specimens. Null alleles were excluded. Alleles from specimens with mixed profiles were included in the calculation of HGDI for each locus but were excluded for the seven-locus scheme.

Where patterns were identified, the relationship between individual loci and the frequency of the alleles identified was explored by Chi square.

To investigate the efficiency of the typing scheme, the maximum number of MLVA profiles that were generated from the best combination of one to seven loci were inferred from Accurate Marker Choice for Accession Identification and Discrimination (AMaCAID) analysis using Model 1 in R (version 4.2.1) (Caroli *et al*., 2011).

To identify clusters of specimens with the same MLVA profiles, a minimum spanning tree (MST) for all those with simple profiles was constructed using Bionumerics software (version 7.6, Applied Maths, Belgium). Specimens were assigned as "non-cluster", known outbreaks or clusters. Genetic clusters were defined as two or more cases with an identical MLVA profile and were highlighted.

Epidemiological Analysis

Cases were located in time by specimen date and location by postcode of residence.

Questionnaire response rates were calculated by geographical region Wales and North West England, and by lower tier local authority.

We categorised the study population into genetic cluster or non-cluster based on the MLVA profile and described them in spatially and temporally variation (epidemiological curve and map), demographics, symptoms and, where available, exposure and risk factors and subsequent epidemiological links.

To understand the number of possible cases that may not have been sent for genotyping at the CRU, we compared the number of *Cryptosporidium* cases in the Second Generation Surveillance System (SGSS) and Tarian with the number of specimens referred to the CRU.

Results and Discussion

A total of 213 PCR-confirmed *C. parvum* specimens, from 213 cryptosporidiosis cases, were tested by MLVA during the study, 118 from north west England and 95 from Wales. Overall typability (where all seven loci were positive) was 173/213 (81 %). The details of the MLVA alleles are in Appendix 2: Laboratory Results.

The MLVA scheme revealed a total of 161 MLVA profiles; 13 of these were mixed. Of the remaining 148 simple profiles, 123 (83 %) were unique within this dataset. The minimum-spanning tree based on the 148 simple MLVA profiles yielded 25 clusters of 77 cases, each cluster ranging from 2 to 9 (mode = 2) cases (Figure 1).

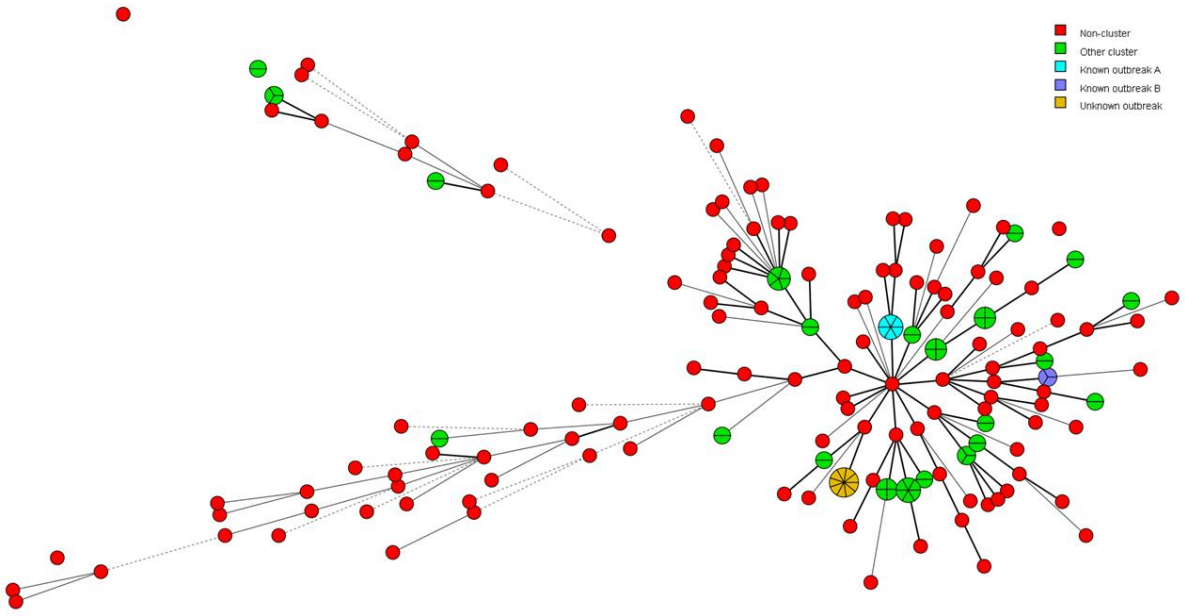

Figure 1: Minimum spanning tree of MLVA of all simple profiles, displaying non-cluster cases, outbreak cluster cases, and other clusters of cases.

Each MLVA profile is indicated by one node or branch tip displayed as circles and connected by branches. The branch line style indicates the number of loci that differ between MLVA profiles: thick solid line = 1 difference, thin solid line = 2 differences, dashed line = 3 differences, no line = 4+ differences. The branch length is not indicative of genetic distance. The wedges in the circles indicate the number of specimens with that particular MLVA profile.

The overall questionnaire response rate for confirmed *C. parvum* cases was 155/213 (73%) of those in Tarian/HPZone but varied between the study areas and across lower tier local authorities (Appendix 3, table i); the rate was higher in Wales (82/95, 86 %) than north west England (73/118, 62 %), and a response did not necessarily include complete data. Four cases in the north west of England were not identified in HPZone. Questionnaires also differed in design even within region, which we attempted to take account of in the data capture tool but harmonisation of questionnaires would be useful for achieving a standard approach to epidemiological investigations.

The background descriptive epidemiology is further presented in Appendix 3: Descriptive epidemiology results. Here we focus on the outbreaks identified.

MLVA identified a previously unknown outbreak

The largest cluster of nine cases in the MST with the MLVA profile 4-12-5-7-27-37-16 was further investigated by the local Health Protection Team and was discovered to indicate a previously unknown outbreak linked to an open farm.

Although the construction of hypothetical profiles was avoided and mixed profiles were therefore not included in the MST, observation of the alleles present was useful, especially in this previously unknown outbreak. A total of 11 cases identified during the study period clustered either by simple profiles or by alleles in common with the main MLVA profiles, and further investigations by the Incident Management Team identified that eight of these (including one of those with a mixed profile) had visited the farm (Table 2). Taking account of mixed profiles was therefore shown to be useful in practice as they may indicate an outbreak strain.

The finding of cases with the outbreak profile that had not visited the farm is indicative that additional reservoirs of *C. parvum* or missing links between cases may exist.

Table 2. The farm premise exposures of cases with the previously unknown outbreak MLVA profile

Case	Visited the open farm?	Specimen date	MLVA profile
1	Yes	24/04/2022	4-12-5-7-27-37-16
2	Yes	29/04/2022	4-12-5-7-27-37-16
3	Yes	20/05/2022	4-12-5-7-27-37-16
4	Yes	20/05/2022	4-12-5-7-27-37-16
5	Yes	21/04/2022	4-12-5-7-27-37-16
6	Yes	25/04/2022	4-12-5-7-27-37-16
7	No	26/04/2022	4-12-5-7-27-12/37-16
8	No	28/04/2022	4-12-5-7-27-37-16
9	Yes	21/04/2022	4-12-5-7-27-37-16
10	Yes	20/06/2022	4-12/13-5-7-27-31/37-16
11	Not known (did not respond to questionnaire)	26/07/2022	4-12-5-7-27-37-16
12	Yes	24/04/2022	4-12-5-7-27-37-16

Scrutiny of this farm as a "context" in HPZone also identified shiga-toxin producing *Escherichia coli* (STEC) cases linked to the premises. Environmental inspections revealed failures at the farm and actions were taken to drive improvements.

MLVA identified additional cases in outbreak A

One cluster of six cases in the MST was part of a previously identified outbreak (outbreak A), also linked to an open farm; these cases had the MLVA profile 4-12-5-7-27-28-16. Two of the cases had not previously been identified as part of the outbreak and MLVA enabled their inclusion in the investigation by the local Health Protection Team. One case that displayed in the MST as a non-cluster case had the MLVA profile 4-14-5-7-27-30-16, which differed from the main outbreak at two loci, cdg4 and cgd8, but had also visited the farm. The finding of various profiles among cases linked to farm premises can occur, possibly due to the variety on animals present and / or duration of outbreaks, or the case may have acquired their infection elsewhere.

MLVA strengthened the microbiological evidence in outbreak B

One other cluster, of three cases in the MST, was also a previously known small outbreak (outbreak B) where the identification of a single MLVA profile 4-14-5-7-27-31-17 in all three cases strengthened the evidence for the common exposure considered by the Health Protection Team as the source of the outbreak.

Outcomes and impact

We have found MLVA to be a non-disruptive addition to the Cryptosporidium Reference Unit's genotyping workflow; the *Cryptosporidium* DNA has already been extracted from stool specimens for species identification, and MLVA can be successfully applied with a fast turnaround time.

Known outbreaks were characterised and additional, epidemiologically linked but previously missed cases were identified by MLVA, strengthening evidence in outbreak investigations.

Unrecognised clusters of cases can be identified by MLVA for further investigation and public health action; thus we have satisfied our primary objective.

We conclude that systematic analysis of *C. parvum* specimens by MLVA should be applied for all of England and Wales, and this change in practice will be progressed from spring 2023 subject to developing mechanisms for cluster reporting to provide context to systematic data capture in SGSS.

The diversity of *C. parvum* found indicates that pursuing the application of NGS-based methods such as DNA hybridisation (baits capture technique) would be fruitful. A cost-benefit analysis would be warranted.

We have identified a need for the harmonisation of case exposure questionnaires across regions.

Two papers are in preparation:

1. A laboratory methods paper describing the "Application of a new multilocus variable number tandem repeat analysis (MLVA) scheme for the seasonal investigation of *Cryptosporidium parvum* cases in Wales and the north west of England, spring 2022".

2. An epidemiological paper describing the "Investigation of genetic clusters of *Cryptosporidium parvum* to develop improved epidemiological understanding: Wales and North West England, March to June 2022".

Ongoing and future work

We continue to encourage diagnostic microbiology laboratories to send *Cryptosporidium*-positive stools for genotyping, at least to identify species for unbiased surveillance. The genotyping service is free at the point of use.

We are developing mechanisms for cluster reporting to provide context to systematic data capture in SGSS.

Including mixed profiles is useful in practice as they may indicate an outbreak strain. As data accrues it will be important to develop probability based approaches to formalise how to interpret mixed profiles in including or isolates within putative clusters.

The MLVA scheme is being harmonised in Europe and a common database of MLVA profiles established in https://pubMLST.org

We are working with the UKHSA gastrointestinal network to harmonise questionnaires.

Discussions are underway with the Animal and Plant Health Agency to establish the routine genotyping and MLVA analysis of *C. parvum* from livestock especially during the spring/early summer.

Acknowledgements

Rahma Mohammed, Cryptosporidium Reference Unit, for specimen preparation and DNA extraction.
Kristin Elwin, Cryptosporidium Reference Unit for *Cryptosporidium* species identification.
Arthur Morris, Cardiff University for *Cryptosporidium parvum* genome analysis.
Diagnostic microbiology laboratories for continuing to send *Cryptosporidium*-positive stools for genotyping.
Health Protection Teams and Environmental Health Departments for following up cases.

This work is funded by the National Institute for Health and Care Research (NIHR) Health Protection Research Unit in Gastrointestinal Infections, a partnership between the UK Health Security Agency, the University of Liverpool and the University of Warwick. The views expressed are those of the authors and not necessarily those of the NIHR, the UK Health Security Agency or the Department of Health and Social Care.

Appendix 1: Laboratory Methods

MLVA Scheme

Seven *C. parvum* VNTR loci were investigated, described previously by Robinson *et al*. (2022): cgd1_470_1429 (cdg1); cgd4_2350_796 (cdg4); cgd5_10_310 (MSF); cgd5_4490_2941 (cdg5); cgd6_4290_9811 (cdg6); cgd8_4440_NC_506 (cdg8); cgd8_4840_6355 (MM19).

Specimens, DNA Extraction & PCR

DNA was extracted from *Cryptosporidium*-positive stools submitted from local diagnostic microbiology laboratories, using the QIAamp Fast DNA Stool Kit (Qiagen, Germany) and *C. parvum* was confirmed by real-time PCR (Robinson *et al*., 2020). All *C. parvum* specimens identified between 28th March 2022 and 31st July 2022 from laboratories in the north west of England and Wales were subtyped by MLVA. Laboratories in these regions were informed by email of the current study prior to its initiation to encourage stool submission to the CRU.

Two multiplex PCRs for the seven VNTR markers, set up as a four-plex targetting cgd1, cgd4, cgd8 and MM19 and a three-plex targetting MSF, cgd5 and cgd6, were performed with the Type-it Microsatellite PCR kit (Qiagen, Germany) as described previously (Robinson et al., 2022).

Fragment Analysis & Subtyping

VNTR PCR products were vortexed (MixMate: Eppendorf, Germany), centrifuged (5430: Eppendorf, Germany) briefly, and diluted 10-fold with HiDi Formamide (Applied Biosystems, U.S.A.). The diluted PCR products were vortexed and centrifuged, and 2 µl added to a master mix containing 12 µl HiDi Formamide and 0.5 µl GeneScan 600 Liz dye Size Standard 2.0 (Applied Biosystems, U.S.A.) in MicroAmp Optical 96-Well Reaction Plates (Applied Biosystems, U.S.A).

The plates were briefly vortexed and centrifuged, plate septa (Thermofisher, U.S.A) applied and loaded into the SeqStudio Genetic Analyser (Applied Biosystems/Thermofisher, U.S.A) and fragment analysis run using the following plate settings in SeqStudio Plate Manager (Thermofisher, U.S.A.): 1) Size Standard: GS600_LIZ_(60-400); 2) Dye Set: G5 (D2-33); 3) Run Module 1: FragAnalysis.

Raw .fsa files were imported into Bionumerics software (version 7.6, Applied Maths, Belgium), which we have found to be particularly beneficial for MLVA analysis as it allows the loci with longer repeat units (e.g. MSF is a 12 bp repeat and cgd4 a 15 bp repeat) to be analysed readily, and also allows for simultaneous visualisation of alleles in individual or multiple channels. The MLVA management function was used to construct bins to capture peaks and determine VNTR alleles by the calculated number of repeats. The bin ranges were initially created from the validation panel of 259 samples described previously using the mean band sizes for peaks from the same allele ± 0.7 bp (Robinson et al., 2022). New alleles were confirmed by sequencing as described previously (Robinson et al., 2022).

A MLVA profile was compiled for each sample by listing the calculated number of repeats at each locus in chromosomal order (cgd1, cgd4, MSF, cgd5, cgd6, cgd8, MM19), expressed as shown in the following example: 4-14-5-7-27-32-15.

Samples that contained one true peak at each locus were described as having simple profiles, and those with >1 true peaks at any locus were defined as mixed profiles. Mixed MLVA profiles were recorded with a forward slash between the alleles observed at each locus, with the lowest copy number displayed first regardless of which was stronger, for example cgd4 here has two alleles at cgd 4: 4-13/14-5-7-27-32-15. Hypothetical profiles were not created from mixed profiles.

If no true peak was visible at a locus, or if peaks were <150 RFU, the DNA (approximately 90 µl) was desiccated with a Concentrator Plus (Eppendorf, Germany) at 45°C for 45 minutes, reconstituted in 20 µl nuclease free water and re-tested. If no allele was identified on re-test, then a null allele (Ø) was recorded. Patterns in the occurrence of null alleles were explored, and sequence variation within the locus investigated using genome data from isolates previously demonstrating similar null alleles (Isolates UKP97, UKP99, UKP127, UKP128 and UKP129 from BioProject PRJEB15112).

Quality Control

PCR positive (*C. parvum* DNA) and no template controls (nuclease free water) were included in every PCR batch.

The *C. parvum* DNA also acted as a reference sample of known MLVA profile. A GeneScan 600 Liz size standard (v2.0, ThermoFisher) was included to ensure accurate allele sizing.

In addition to the PCR positive DNA, two *C. parvum* DNA samples were repeated as quality control and produced the same results each time.

Appendix 2: Laboratory Results

A total of 234 *Cryptosporidium*-positive stool specimens were received for genotyping from primary diagnostic laboratories in Wales (where all seven laboratories sent 101 specimens) and the north west of England (where 15/16 laboratories sent 133 specimens) during the study period (Table i). Specimens from Macclesfield were not included in this study as they are sent to the Royal Stoke Hospital in the West Midlands for testing.

The proportion of case specimens identified through the study compared to routine Tarian notifications (Wales) and SGSS (NW England) (data extracted August 2022) was 94 % (95/101) of Wales cases and 89 % (118/213) of NW England cases. 214 were C. parvum. Although not all laboratories submitted specimens (Table i), the submission rate was high and resulted in 94 % and 89 % of all laboratory confirmed C. parvum cases in Wales and NW England respectively had specimens referred to the reference unit, were identified as C. parvum, went through MLVA and were included in the study. One specimen was unfortunately overlooked during identification for MLVA.

Table i. Referral and detection of *C. parvum* in *Cryptosporidium*-positive stools for genotyping

Region	Primary diagnostic microbiology laboratory	Specimens containing *C. parvum* / total number submitted (%)
North West England	Royal Blackburn Hospital	9/10 (90)
	Victoria Hospital (Blackpool)	13/13 (100)
	Royal Bolton Hospital	2/3 (67)
	Micropath (Wirral)	13/17 (76)

	Cumberland Infirmary Carlisle	21/21 (100)
	Furness General Hospital	4/4 (100)
	Royal Lancaster Infirmary	7/7 (100)
	Royal Liverpool Hospital	4/6 (67)
	Royal Oldham Hospital	3/5 (60)
	UKHSA North West, Manchester Royal Infirmary	5/9 (56)
	Royal Preston Hospital	8/8 (100)
	Salford Royal Hospital	11/11 (100)
	Stepping Hill Hospital, Stockport	0/0
	Warrington Hospital	4/4 (100)
	Whiston Hospital	15/15 (100)
Wales	PHW Microbiology Aberystwyth	12/13 (92)
	PHW Microbiology Cardiff	10/11 (91)
	Royal Glamorgan Hospital	7/8 (88)
	Royal Gwent Hospital	17/17 (100)
	PHW Microbiology Rhyl	24/25 (96)
	PHW Microbiology Swansea	23/25 (92)
	Withybush Hospital Haverfordwest	2/2 (100)

Overall MLVA typability (where all seven loci were positive) was 173/213 (81.2 %) (Table ii).

Further to those alleles identified previously (Robinson *et al*., 2022), 19 new alleles were confirmed by Sanger sequencing and new bins were generated with the confirmed band sizes. The most variable locus was cgd8 with 29 alleles and the least variable was cgd1 with four alleles (Table ii).

Table ii. Variability of the VNTRs in 213 *C. parvum* specimens

Locus	Range of calculated number of repeats	No. of alleles	No. of specimens with null alleles	No. of typable specimens (% typability)	Hunter-Gaston discrimination index
cgd1	4-7	4	40	173 (81.2)	0.24
cgd4	11-17	7	4	209 (98.1)	0.62
MSF	2-10	8	1	212 (99.5)	0.44
cgd5	3-18	12	1	212 (99.5)	0.64
cgd6	13-31	6	1	212 (99.5)	0.55
cgd8	4-54	29	2	211 (99.1)	0.93
MM19	4-40	27	6	207 (97.2)	0.81
7-locus scheme			40	173 (81.2)	0.99

The VNTR loci displayed variable discriminatory power as seen by the HGDI (Table ii). HGDI is a function of the number of alleles detected at each locus and their frequency of detection (Figure i). For example, more alleles were detected in MSF than cgd4 and cgd6, the HGDI was lower as one allele accounted for 74 % of the total in MSF (Figure i). Although only two more alleles were detected at cgd8 than at MM19, the HGDI was much higher due to the distribution of allele frequency (Figure i).

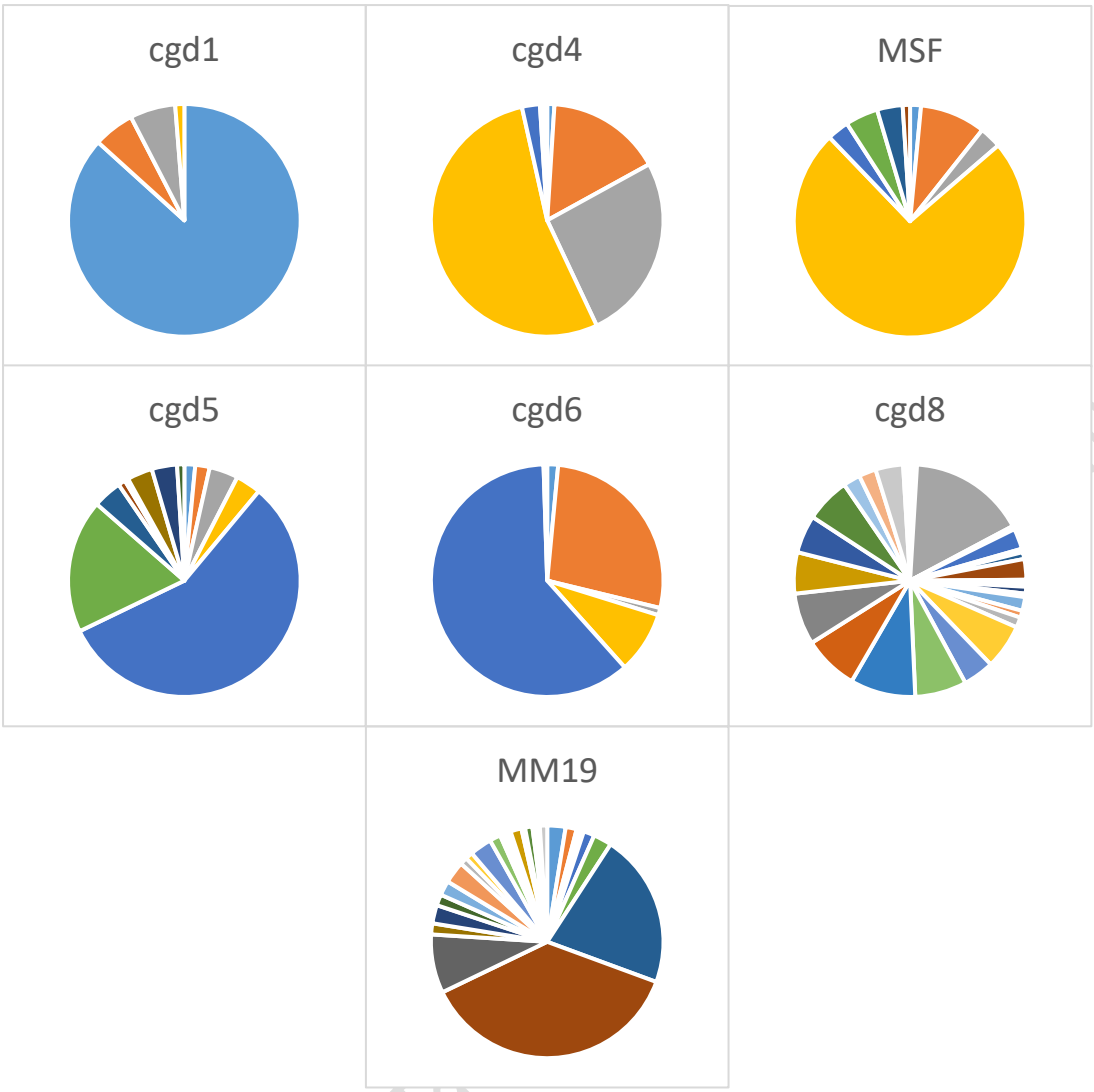

Figure i. The frequency with which alleles were detected at individual loci, with each segment reflecting the proportion of each allele found at that locus. *The colours are simply to make the segments clear within each pie.*

AMaCAID analysis revealed that the maximum of 140 MLGs, achieved by analysing all seven loci, could also be achieved by omitting MSF (Figure ii). After also omitting cgd1 this number reduced by three to 137. Omitting other loci in optimal combinations reduced the maximum MLGs more greatly (Figure ii). There may be an argument for re-designing the seven-locus scheme with just five loci, cgd4, cgd5, cgd6, cgd8 and MM19, in which case re-design of the cgd1 primers would not be required.

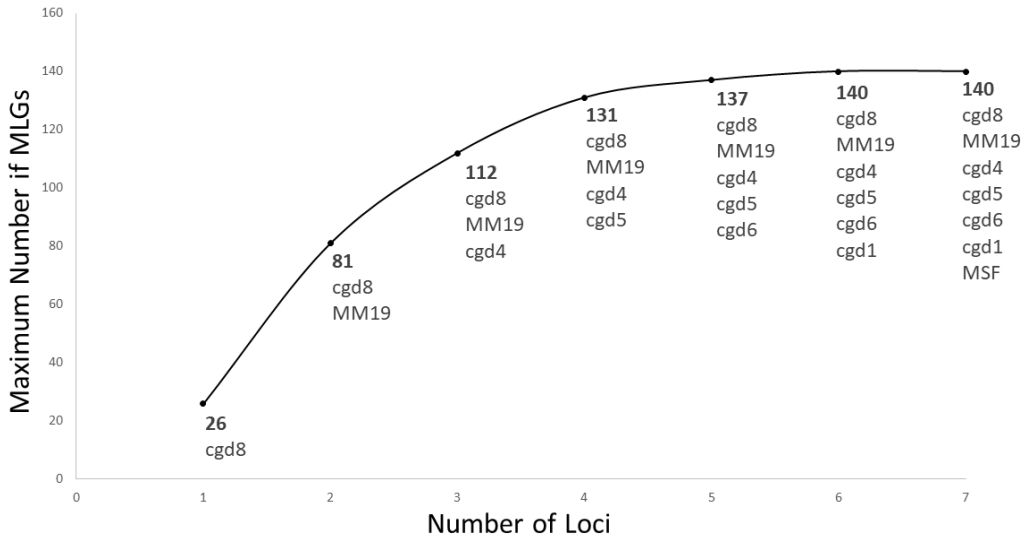

Figure ii. AMaCAID analysis of the maximum number of genotypes discriminated by marker combos. Maximum number of Multi-Locus Genotypes (MLGs) discriminated as a function of the number of loci used.

A total of 40 specimens had 49 null alleles where peaks were below the threshold of 150 RFU; 33 specimens had null alleles only in 1 locus, 3 at 2 loci, 3 at 3 loci and 1 at all seven loci (Figure iii). In addition to being least variable, cgd1 also had the highest number of null alleles (Table ii). All 40 of these specimens from the total 213 (18.8%) had a null allele at the cgd1 locus. Null alleles occurred at other loci, but at a much lower frequency (e.g. MM19 = 3%, cgd4 = 2%, cgd8 = 1%).

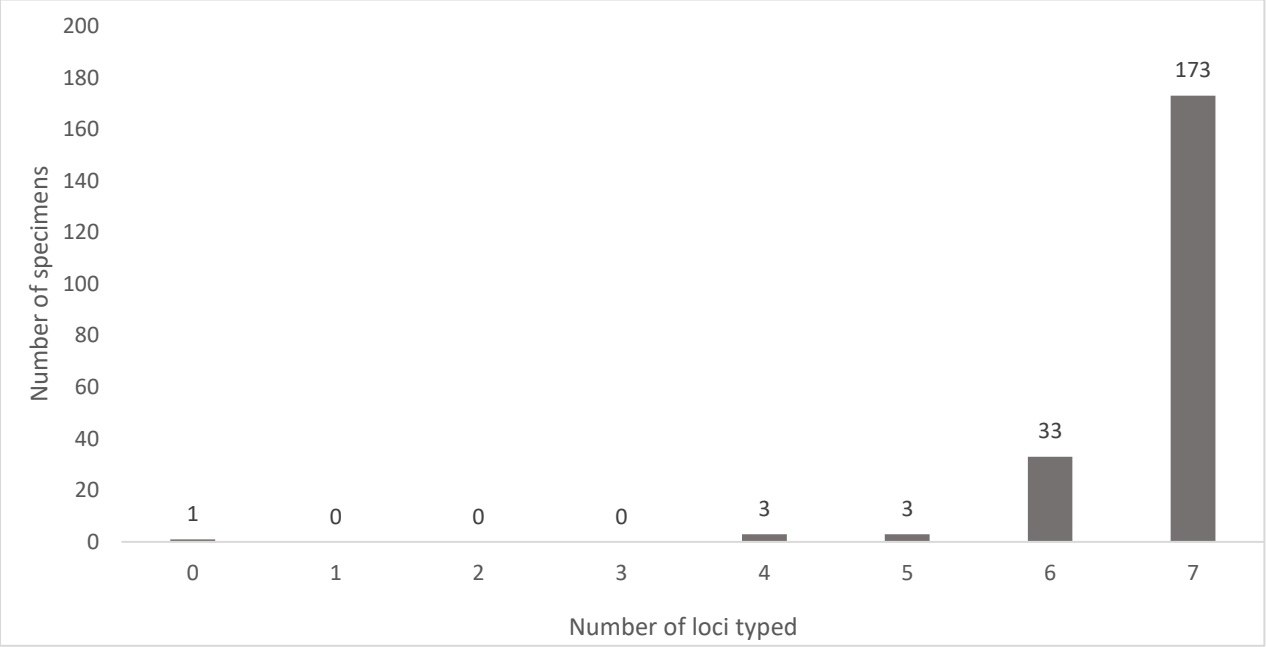

Figure iii: Frequency distribution of specimens by the number of loci typed within the seven-locus scheme.

The most commonly seen combination among the incompletely typed specimens was nine repeats at cgd8 and a null allele at cgd1 in 25/40 (63%), significantly more than other alleles at these loci (Mantel-Haenszel chi-squared 76.89, p=0.000) (Table iii).

Null alleles may indicate a weak sample, true null alleles or variation in the primer sequence sites.

The Ct values in the original genotyping real-time PCRs of the 40 specimens that contained null alleles (range 29.22-38.16, mean = 32.18, SD 3.15) were significantly higher than those of the 173 typable specimens (range 24.44 to 35.73, mean = 30.29, SD 3.05), t-value = 2.72, p= 0.007.

Five specimens that had null MLVA alleles at cgd1 but were typable at other loci (notably cgd8 with 5, 9 or 10 repeats), had been previously analysed by whole genome sequencing and sequence alignment revealed mismatches in the cgd1 MLVA PCR primer sites. The re-design of the primers for cgd1 might therefore be warranted.

Table iii: 2 x 2 table of specimens with cgd1 null allele or other alleles, versus cgd8-9 or other alleles at cgd8. Mixed profiles were not included in the calculations.

	cgd8 nine repeats	cgd8 other alleles	Total
cgd1 Ø allele	25	15	40
cgd1 other alleles	8	153	161
Total	33	168	201

Appendix 3: Descriptive epidemiology results

A total of 213 C. parvum isolates, recevied between 28 March and 29 July 2022, were further investigated by MLVA, with specimen dates between 23 March and 27 July 2022.

118/213 cases were resident across 38 lower tier local authorities in the North West of England and 95 cases resident across 22 lower tier local authorities in Wales (Table iv). In north west England, 62 cases were in the Cumbria and Lancashire (C&L) Health Protection Team (HPT) area, 36 in Cheshire and Merseyside (C&M) HPT and 20 in Greater Manchester (GM) HPT, and no resident postcode was captured for four cases – one from GM, one from C&L and two from C&M (Table iv and Figure iv).

Table iv: Case questionnaire response rate by lower tier local authority, n=213

Local authority	Number of cases	Number of case questionnaires received	Proportion of questionnaires received (%)
Allerdale	4	1	25%
Barrow	1	0	0%
Blackburn	2	2	100%
Blackburn with Darwen	1	1	100%
Blackpool	1	0	0%
Blaenau Gwent	4	4	100%
Bolton	2	2	100%
Bridgend	2	1	50%

Caerphilly	3	3	100%
Carmarthenshire	1	1	100%
Cardiff	6	6	100%
Carlisle	2	0	0%
Carmarthenshire	8	8	100%
Ceredigion	10	9	90%
Cheshire East Council	1	1	100%
Cheshire West and Chester	7	2	29%
Chorley	2	0	0%
Conwy	9	8	89%
Copeland	2	0	0%
Cumbria	13	13	100%
Denbighshire	3	2	67%
Durham	1	1	100%
Flintshire	2	1	50%
Fylde	5	0	0%
Gwynedd	5	4	80%
Halton	1	1	100%
Isle of Anglesey	2	2	100%
Knowsley	1	1	100%
Lancashire	10	10	100%
Liverpool	1	1	100%
Manchester	2	2	100%
Merthyr Tydfil	2	1	50%
Monmouthshire	4	4	100%
Neath Port Talbot	3	3	100%
Newport	3	3	100%
Oldham	1	1	100%
Pembrokeshire	3	1	33%
Pendle	1	0	0%
Powys	6	5	83%
Preston	1	1	100%
Rhondda Cynon Taff	5	5	100%
Ribble Valley	3	2	67%
Rochdale	1	0	0%
Rossendale	2	0	0%
Salford	5	5	100%
Sefton	7	5	71%
South Lakeland	6	3	50%
South Ribble	2	1	50%
St Helens	7	4	57%
Swansea	7	6	86%
Tameside	1	0	0%
Trafford	2	1	50%
Vale of Glamorgan	4	4	100%
Warrington	3	3	100%

Wigan	5	4	80%
Wirral	5	5	100%
Wrexham	3	1	33%
Wyre	3	0	0%
(blank)	4	0	0%
Total	**213**	**155**	**73%**

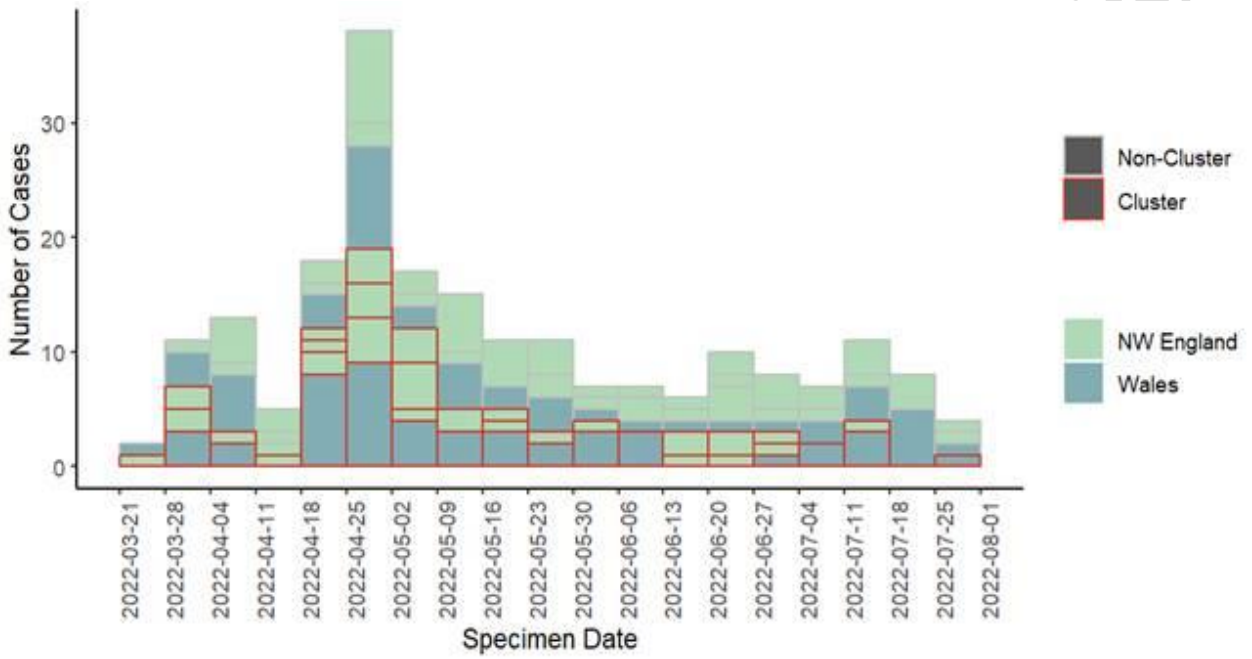

Figure iv: Temporal distribution of cases by specimen date within the study period, by region and cluster status, 28 March - 29 July 2022 (n=209*)

**4 cases with unknown specimen date*

There were more women than men in the study population, with 124 (58%) of all cases being reported as female, 88 (41%) as male and 1 with an unknown sex. Most cases were under 20 years old (114/213, 54%) with 34% (72/213) being aged between 0 and 9 years at the time of their specimen date. This age-sex profile is as expected for *Cryptosporidium* cases with a greater proprotion of young children and a peak in females of child-bearing and caring age (Figure v). There were 4 cases with no date of birth recorded.

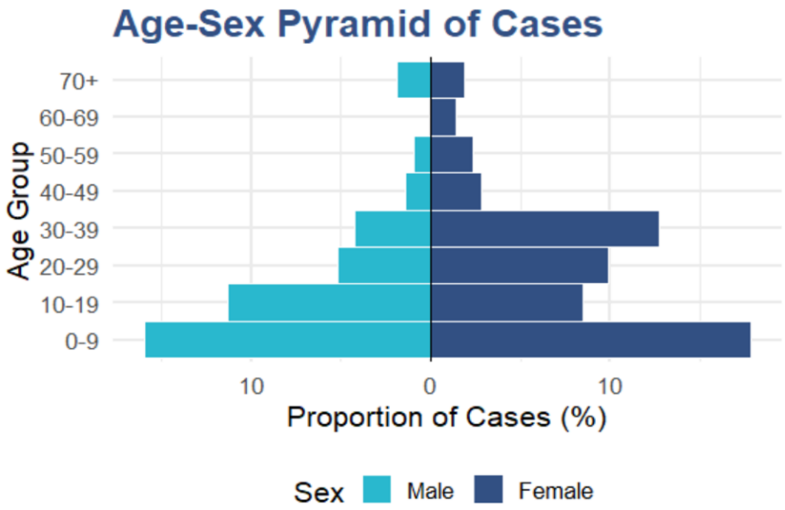

Figure v. Age-sex pyramid of *C. parvum* cases included in the study, 28 March - 29 July 2022 (n=209).

Date of symptom onset was captured for 144 cases, ranging between February 2022 and July 2022. Most cases had an onset date preceding a specimen date with a median time difference of 7 days (range 1 to 84 days). For 6 cases where a specimen was reportedly taken before an onset date there was a median time of 6 days (range 0 to 18 days); these dates are plausible (for example, specimens take for other purposes) but may be due to a reporting error when capturing onset dates.

Of 134 cases that completed a question about ongoing illness, 55 (41%) said that they were still experiencing symptoms at the time of completing the questionnaire. Median duration of illness was 11 days (range 2 to 28 days) for those not reporting symptoms at the time of questionnaire, and varied slightly between cluster cases (11 days, 6-21 days) and non-cluster cases (10 days, 2-28 days).

Where symptoms were specified, most cases reported diarrhoea and abdominal pain; this was similar across cluster and non-cluster cases. Blood in stools was the least frequently reported symptom (Table v).

Table v: Proportion of self-reported symptoms among all, cluster and non-cluster cases

Reported Symptom	All cases	Cluster cases	Non-cluster cases
Diarrhoea	95%	96%	94%
Abdominal pain	85%	81%	88%
Vomiting	63%	62%	64%
Nausea	54%	48%	62%
Fever	46%	46%	46%
Blood in stool	9%	5%	13%
Fever	46%	46%	46%

The reported exposures indicated that zoonotic exposure was important in the transmission *of C. parvum* during the spring/early summer, with 52% cases reporting living or visiting a farm (Table vi). Having a private water supply was reported by 7% of cases which is greater than the 1.2% of the population of England and 2.2% of the population of Wales served by a private water supply (Drinking Water Inspectorate, 2022a and b).

Table vi. Reported exposures of *C. parvum* cases from routinely collected questionnaires

Contact with pets	Live on or visited a farm	Eating out	Pool or spa	International travel	Private water supply
103/145 (71%)	82/159 (52%)	78/147 (53%)	50/141 (35%)	21/143 (15%)	9/131 (7%)

A total of 207 cluster and non-cluster cases were mapped by residence postcode; four postcodes were missing, and two were unable to be matched onto Office for National Statistics postcode data to get the longitude and latitude coordinates for the mapping. All the unmapped cases were from the north west England.

The mapping of cases, identified as cluster or non-cluster, showed that clusters were distributed throughout the study area, although the cluster cases showed a narrower geographical distribution than non-cluster cases (Figure v). No cluster cases were found in 19 lower tier local authorities, whereas there were 8 lower tier local authorities where only cluster cases were found (Figure v). Interestingly, only one cluster was seen in the upper tier area of Cumbria (Figure v).

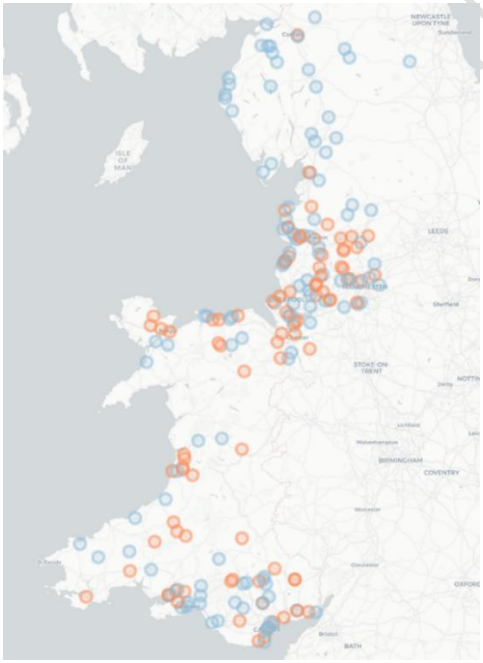

Figure v. Map of MLVA clusters of *C. parvum* cases (red) and non-cluster *C. parvum* cases (blue).
Each circle = one case

References

Caroli S, Santoni S, Ronfort J. AMaCAID: a useful tool for accurate marker choice for accession identification and discrimination. Molecular Ecology Resources 2011;11(4):733-8.

Drinking Water 2021a. Private water supplies in England. Drinking Water Inspectorate, London, 2022. Available online at https://dwi-content.s3.eu-west-2.amazonaws.com/wp-content/uploads/2022/07/22160713/E02750082_DWI_Chief-Inspectors-Private-Water_England.pdf (Last accessed 23/01/23)

Drinking Water 2021b. Private water supplies in Wales. Drinking Water Inspectorate, London, 2022. Available online at https://dwi-content.s3.eu-west-2.amazonaws.com/wp-content/uploads/2022/07/22160704/E02750083_DWI_Chief-Inspectors-Private-Water_Wales.pdf (Last accessed 12/01/22)

Gopfert A, Chalmers RM, Whittingham S, Wilson L, Van Hove M, Ferraro CF, Robinson G, Young N, Nozad B. An outbreak of Cryptosporidium parvum linked to pasteurised milk from a vending machine in England: a descriptive study, March 2021. Epidemiology and Infection 2022;150:e185.

Hunter PR, Gaston MA: Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. Journal of Clinical Microbiology 1988;26:2465-2466.

Robinson G, Pérez-Cordón G, Hamilton C, Katzer F, Connelly L, Alexander CL, Chalmers RM. Validation of a multilocus genotyping scheme for subtyping Cryptosporidium parvum for epidemiological purposes. Food and waterborne parasitology. 2022;27:e00151.

Robinson G, Elwin K, Chalmers RM. Cryptosporidium diagnostic assays: molecular detection. In Cryptosporidium 2020 (pp. 11-22). Humana, New York, NY.