"Investigation of genetic variation within *Cryptosporidium hominis* for epidemiological purposes"

Contract number DWI 70/2/199

Final report to DWI

by

Dr Rachel Chalmers[1], Dr Colin Jackson[2], Dr Kristin Elwin[1], Dr Stephen Hadfield[1] and Professor Paul Hunter[3]

[1]UK Cryptosporidium Reference Unit, NPHS Microbiology Swansea
Singleton Hospital, Swansea, SA2 8QA

[2]Biological Sciences, University of Wales, Aberystwyth (now at Institute of Grassland and Environmental Research, Aberystwyth AB23 3EB)

[3]School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, NR4 7TJ

March 2007

CONTENTS **Page**

Executive summary 1

Background and Introduction 2

Aim 5

Objectives 5

SECTION 1. INVESTIGATION OF GP60 GENE 6

METHOD 6
Sample sets and selection 6
Prior DNA extraction and species identification 6
Bi-directional sequencing of the GP60 gene 7
DNA sequence analysis and GP60 allele identification 7

RESULTS 8
GP60 PCR EQA, sensitivity and typability 8
GP60 allele families and subtypes by sample set 8

DISCUSSION 14

SECTION 2: INVESTIGATION OF SSCP 15

METHOD 15
SSCP ITS-2 15
SSCP GP60 15

RESULTS: SSCP ITS-2 16
Optimisation 16
Evaluation 16

RESULTS: SSCP GP60 16
Optimisation 16
Evaluation 16
Library construction 17

DISCUSSION 17

SECTION 3 **17**

CONCLUSIONS **17**

Acknowledgements **17**

REFERENCES **18**

Appendix 1 **21**

Appendix 2 **27**

Appendix 3 **30**

EXECUTIVE SUMMARY

Cryptosporidiosis is a diarrhoeal disease of humans and young animals caused by the protozoan parasite *Cryptosporidium*. Illness in neonatal animals is caused by *Cryptosporidium parvum* and both this and a human-adapted species, *Cryptosporidium hominis*, cause cryptosporidiosis in people. Each species causes approximately half of the reported cases in England and Wales, but the distribution by person, time and place (the epidemiology) differs, as do the risk factors for infection. Further investigation of variants within *C. parvum* has revealed more about the natural history and epidemiology of this species, but genetic variation within *C. hominis* is poorly understood and the distribution of variants unknown. In a previous study, we found that 90% of *C. hominis* isolates from apparently un-linked (sporadic) cases in England and Wales were indistinguishable. A study of the population genetics of *C. hominis* in the north east of Scotland showed isolates were genetically very similar, indicating that the parasite population structure was almost clonal, while in other countries more variation has been observed.

In this project, to investigate whether other typing methods for *C. hominis* isolates could be useful for epidemiological purposes in England and Wales, more genetic loci were studied and any relationship between *C. hominis* variants and epidemiological factors was investigated. We identified sporadic and outbreak cases from the national collection of *Cryptosporidium* oocysts maintained at the UK Cryptosporidium Reference Unit in Swansea. In the first instance, we read the DNA sequence of part of the parasite genome by looking at a large part of the GP60 gene. This is a highly variable gene, and many "families" have been identified in both *C. parvum* and *C. hominis*. The gene also contains variable numbers and forms of a repeating sequence of nucleotides, microsatellite DNA, which can be used to identify variation within these families.
Cases in a case control study were investigated for relationships between exposure and GP60 results. Family Ib subtype A10G2 was the most common in sporadic cases, representing over 90% all isolates. However, people with non-IbA10G2 isolates were statistically more likely to have returned from non-European destinations than people with IbA10G2.

Isolates from two drinking waterborne outbreaks that occurred in the Autumn of 2005 were compared with sporadic cases and exposure. In the outbreak in north west Wales, *C. hominis* isolates were exclusively IbA10G2. This was also the predominant type locally in cases during the six years prior to the outbreak. In the outbreak in south east England, IbAl0G2 also predominated but two other *C. hominis* families were also present, although there was no difference in the results of the epidemiological analysis of the outbreak when these cases were excluded from the analysis.

We also investigated the development of an alternative typing method for *C. hominis* and *C. parvum* to DNA sequence analysis, which is costly and time consuming. We investigated single strand conformation polymorphisms (SSCP) at two genetic loci. Investigation of the ITS-2 region showed more variation between sample runs than between samples. However, SSCP on the GP60 gene appeared to be much more reliable, and had the advantage of providing direct comparison with the DNA sequence based analysis.

To conclude, this study:

1. tells us that indigenous *C. hominis* in the UK shows little genetic variation in the GP60 gene and supports previous findings at other loci
2. indicates little change in *C. hominis* over the time investigated
3. suggests wider global transmission may be subject to host-related or social factors
4. an international database of *Cryptosporidium* variants with standardised nomenclature would assist in interpretation of studies elsewhere
5. Further method development is required for rapid methods for epidemiological purposes.

BACKGROUND AND INTRODUCTION

Since January 2000, primary diagnostic laboratories throughout England and Wales have been sending *Cryptosporidium* isolates to the UK Cryptosporidium Reference Unit (CRU) to support the national collection of *Cryptosporidium* oocysts (Anon 2002). This now comprises over 14000 isolates, each of which is represented in the collection by stored oocyst suspensions, DNA, and patient data, and characterised at least to the species/genotype level using polymerase chain reaction restriction fragment length polymorphism analysis (PCR-RFLP). Molecular epidemiology based on a prior study in England and on the national collection has shown the spectrum and prevalence of *Cryptosporidium* species/genotypes in human disease. It has also demonstrated geographic and temporal differences between the two predominant species, *Cryptosporidium parvum* and *Cryptosporidium hominis* (McLaughlin *et al*., 2000; Chalmers *et al*., in preparation). The collection has also been used to support a case control study of sporadic cryptosporidiosis in Wales and the north west of England, in which separate risk factors for *C. hominis* and *C. parvum* were identified (Hunter *et al*., 2004a). For *C. hominis* these were contact with another person with diarrhoea, changing nappies, and travel abroad in the two weeks prior to illness. *C. hominis* has been associated with non-intestinal symptoms up to two months after acute disease (Hunter *et al*., 2004b).

An international trial for the comparison of typing methods for application to epidemiological investigations has shown parity between methods for species/genotype identification and intra species variation (Chalmers *et al*., 2005). Further discriminatory methods have been applied to investigate genetic variation within *C. parvum* and *C. hominis*, and were reviewed by Robinson (2006). At the CRU, multi-locus fragment typing (MLFT) of three microsatellite DNA markers (ML1, ML2 and gp15/45) identified greater variation in *C. parvum* than *C. hominis* isolates from the sporadic case control study, with 90% *C. hominis* isolates having the same MLFT (Anon, 2005; Hunter *et al*., 2007). There may be inherently less variation within *C. hominis* populations than within *C. parvum*, as has been suggested from a study in Scotland (Mallon *et al*., 2003) but not borne out in Malawi (Peng *et al*., 2003) and India (Gatei *et al*., 2006). Equally, other genetic targets may be more applicable for identifying genetic variation within *C. hominis* and for epidemiological investigations of this species.

It is apparent from prior research and from surveillance data that the spring peak in cryptosporidiosis reported in England and Wales, attributable to *C. parvum*, has declined in recent years (Nichols *et al*., 2006). The autumn peak, attributable to *C. hominis*, has not declined and during 2005 was particularly marked and sustained, an observation partly but not wholly explained by recognised outbreaks involving drinking and recreational waters. *C. hominis* is clearly an important human pathogen and yet the relationship between isolates causing outbreaks and sporadic disease is unclear and requires further investigation.

While the prior investigations and an international trial have provided important information about the utility of various subtyping methods and the epidemiology of cryptosporidiosis caused by *C. parvum* subtypes, further investigation is required of *C. hominis* isolates (Anon, 2005; Chalmers *et al*., 2005). Application of molecular techniques for investigation of genetic variation within *C. hominis* will permit better understanding of the population structure of *C. hominis* and better epidemiological tools to provide information on the spread and control of disease. This project focuses on the applicability for the molecular epidemiology of *C. hominis* of commonly used typing tools.

Isolates from the national collection of *Cryptosporidium* oocysts have been examined previously for genomic variation at three microsatellite loci using fragment size analysis (Anon, 2005; Hunter *et al*., 2007), ML1, ML2 and gp15/45 (Table 1).

Table 1. Microsatellite markers previously used at the CRU for *Cryptosporidium* genotyping

Marker	Chromosome	Location (*C. parvum* Iowa)	Repeat Unit	Codes for	GenBank accession	Reference
ML1	III	Hypothetical protein (nt 906689 – 911434, plus strand)	GAG	Glutamic acid	AJ249582	Caccio *et al*., 2000
ML2	VI	Possible sporozoite cystein-rich protein (Segment 1/4)	GA	Non-coding	AJ308567	Caccio *et al*., 2001
"gp15/45"	VI	GP60 surface glycoprotein (nt 266434-267408, minus strand)	TCA	Serine	AF164490	Strong *et al*., 2000

We now looked in more detail at the GP60 gene by reading the DNA sequence from most of the gene. This gene codes for a 60 kDa precursor protein that is cleaved to form two zoite cell surface glycoproteins (gp15 and gp45), involved in host cell attachment and invasion, that are highly antigenic (Strong *et al*., 2000). It is a very variable gene, and many allelic families have been identified in both *C. parvum* and *C. hominis* (Table 2). The gene also contains a polyserine tract (variable numbers and forms of a repeating sequence of three nucleotides coding for the amino acid serine), frequently referred to as microsatellite gp15/45, which had been included in the previous study to identify subgenotypes. Here, we included DNA sequence analysis of an 850 bp region of the GP60 gene encompassing the polyserine tract and the hypervariable downstream region (Strong *et al*, 2000; Peng *et al*, 2001; Alves *et al*., 2003,), providing information from a large part of the gene. DNA sequencing of the microsatellite region, as opposed to fragment size anaylsis, provides data which might otherwise be missed since length may mask differences in sequence.

Sequence analysis of the GP60 gene has been used for investigation of *Cryptosporidium* isolates from Uganda (Akiyoshi *et al*., 2006), India (Peng *et al*., 2001; Gatei *et al*., 2006), Canada (Trotz-Williams *et al*., 2006), USA (Strong *et al*., 2000; Zhou *et al*., 2003; Feltus *et al*., 2006), Australia (Chalmers *et al*., 2005), Malawi (Peng *et al*., 2003), Japan (Wu *et al*., 2005), Portugal (Peng *et al*., 2001; Alves *et al*., 2003), South Africa (Leav *et al*., 2002), China, Guatemala, Kenya, and Slovenia (Peng *et al*., 2001). Some researchers identified the isolates to the allelic family level (Ia through to Ie in *C. hominis*, IIa through to IIf in *C. parvum*) while others used further nomenclature to identify allelic subtypes according to Sulaiman *et al*. (2005) or Gatei *et al*. (2006) (Table 2), described in Methods. In some instances, the types are not named but the DNA sequences placed on the freely accessible database for any DNA sequences, Genbank http://www.ncbi.nlm.nih.gov/

Table 2. Nomenclature and geographical distribution of published *Cryptosporidium parvum* and *Cryptosporidium hominis* GP60 allelic subtypes

Location	GP60 allelic subtype	Reference
Kuwait	IbA9G3 IbA10G2 IdA14 IeA11G3T3 IIaA15G1R1 IIaA15G2R1 IIcA5G3 IIdA18G1 IIdA20G1a, b, c and d IIfA6	Sulaiman *et al*. (2005)[1]
India	IaA12G1 IaA12G1R1 IaA19R3 IaA21R3 IaA22R3 IbA9G3 IdA15G1 IdA15G29 IdA16 IeA11G3T3	Gatei *et al*. (2006)[2]
USA	IIaA15G2R1 IIaA15G2R2 IIaA16G1R1 IIaA16G2R1 IIaA16G3R2 IIaA17G2R1 IIaA17G2R2 IIaA17G4R2 IIaA18G2R1	Feltus *et al*. (2006)[1]
Canada	IaA19R3 IaA23R4 IdA19 IeA11G3T3 IIaA13G2R1 IIaA15G2R1 IIaA15G2R2 IIaA16G1R1 IIaA16G2R1 IIaA16G3R1 IIaA17G2R1 IIaA18G3R1	Trotz-Williams *et al*. (2006)[1,2]
Australia UK	IbA9G3 IbA10G2 IeA12G3T3 IIaA15G2 IIaA17G1 IIaA18G3 IIaA20G3	Chalmers *et al*. (2005)[1]
Portugal	IaA19R3 IbA10G2 IdA15 IeA11G3T3 IfA14G1 IIaA15G2R1 IIaA16G2R1 IIbA14 IIcA5G3a IIcA5G3b IIdA17G1 IIdA19G1 IIdA12G1 IIdA22G1	Alves *et al*. (2006)[1]

[1]Identified using the nomenclature system of Sulaiman *et al*. (2005)
[2]Identified using the nomenclature system of Gatei *et al*. (2006).

Candidate isolates for this study include those from confirmed outbreaks, compared with those from sporadic indigenous cases and from patients who have travelled abroad. GP60 analysis was then used as a benchmark against which single strand conformation polymorphism (SSCP) analysis of PCR amplified products from the ITS-2 region (Gasser *et al*., 2004) and the GP60 gene (Wu *et al*., 2003) were compared. SSCP was evaluated for use as a rapid, high throughput method of subtyping *C. hominis* and *C. parvum* isolates. SSCP of the internal transcribed region 2 (ITS-2) has been used by Gasser *et al*. (2004) for subtyping of *C. hominis* and *C. parvum* isolates. The group performed non-radioactive analysis of ITS-2 PCR fragments using a thermoregulated gel apparatus (SEA2000, Elchrom scientific). Subtyping of *C. hominis* and *C. parvum* isolates by PCR of the GP60 region followed by restriction enzyme digestion and SSCP analysis of the products was previously described by Wu *et al.* (2003). The group described good discrimination of both species among 41 human and animal *C. parvum* isolates. In this project, the GP60 PCR-RFLP-SSCP was adapted and investigated for use as a rapid cheaper alternative to sequencing using the SEA 2000 apparatus (Elchrom Scientific).

AIM

Use molecular techniques to compare the prevalence, genetic variation and clustering within sporadic cases of *C. hominis*, and contrast with outbreak cases in north Wales and the south of England during autumn 2005.

OBJECTIVES

Objective 1
Identify cases and archived DNA from up to 250 *C. hominis* isolates from the national collection of *Cryptosporidium* isolates from the case control study in Wales and north west England, from primary cases in a drinking waterborne outbreak in the south east and cases from a drinking waterborne outbreak in north Wales outbreak in 2005, and cases prior to that outbreak were included. EQA was provided by *C. hominis* DNA previously included in the international trial.

Objective 2
Undertake DNA sequence analysis of PCR products from the GP60 gene to provide benchmark data.

Objective 3
Undertake PCR-SSCP ITS-2 to investigate the prevalence and epidemiological distribution of *C. hominis* subtypes and undertake PCR-RFLP-SSCP GP60 of a subset of 50 isolates to investigate utility as a proxy test for GP60 sequencing.

Objective 4
Apply subtype data to the epidemiological database (national collection and sporadic case control study data) and analyse the relationship between subtypes and exposures.

Objective 5
To provide a final report on the applicability of molecular epidemiological tools for investigation of cases, clusters and outbreaks of disease caused by *C. hominis* in England and Wales.

SECTION 1. INVESTIGATION OF GP60 GENE

METHOD

Sample sets and selection

A total of 286 *C. hominis* isolates were investigated at the GP60 gene. In addition, EQA samples were included in each batch process and analysis, selected from archived *C. hominis* DNA previously characterised in other laboratories during the international trial (Chalmers *et al*., 2005). The sample sets were as follows:

i. Sporadic cases: DNA from 115 sporadic cases previously part of the case control study (Hunter *et al*, 2004) were examined to explore any associations between exposure and GP60 alleles. The data set included indigenous cases and those returning from abroad. Single variable analysis was performed in SPSS. Discriminatory power was calculated by the Hunter Gaston index (Hunter and Gaston, 1988).

ii. Drinking waterborne outbreak 1: DNA from all 76 primary *C. hominis* cases reported to the Hampshire and Isle of Wight Health Protection Unit (Portsmouth office) from an outbreak on the south east coast in the autumn of 2005 was identified in the CRU archive. These samples were examined to explore whether there were any associations between exposure and being a primary case with *C. hominis* GP60 alleles. Statistical analysis was undertaken by Neville Verlander, HPA Centre for Infections Statistics Unit, Colindale on behalf of the outbreak control team (Neira-Munoz and Smith, 2006). Single variable analysis was performed in EpiData Analysis and multivariable analysis in STATA. Those variables from the single variable analysis with odds ratios bigger than one and p-values of 0.2 or less were included in the multivariable model. Sex was included in the multivariable model regardless of its significance.

iii. **Drinking waterborne outbreak 2:** DNA from 71 out of a total of 231 cases linked to the *C. hominis* outbreak in north west Wales in the autumn of 2005 were identified in the CRU archive. Selection was made on the following criteria from case questionnaire data provided by the North Wales Health Protection Team: all cases reporting recent foreign travel were excluded; all cases where another member of the household was also a case were excluded; every other remaining case in the database was chosen. These samples were examined to compare this outbreak with others and with sporadic cases.

iv. Historical cases prior to outbreak 2: DNA from 24 out of 48 cases from the local laboratory serving the population submitted to the CRU since January 2000 were selected by identifying every alternate case in the CRU archive. These samples were examined to investigate local changes in predominant alleles over time.

Prior DNA extraction and species identification

Oocysts were separated by flotation from faecal debris using saturated NaCl solution and centrifugation for 8 min at 1100 *x* g (Ryley *et al*., 1976). The floated material containing the oocysts was washed with de-ionised oocyst-free water, the oocysts resuspended in 1 ml de-ionised, oocyst-free water and stored at +4 ^{o}C prior to use. To extract DNA, 200 µl oocyst suspension was incubated at 100 ºC for 60 min and DNA extracted using proteinase K digestion in lysis buffer at 56 ºC and a spin-column filtration technique (QIAamp DNA mini kit, Qiagen). DNA extracts were stored at -20^{o}C prior to use.

To identify *Cryptosporidium* species, the *Cryptosporidium* oocyst wall protein (COWP) gene was investigated using PCR-RFLP (Spano *et al*., 1997). Briefly, primers cry-15 and cry-9 were

used to amplify a 550 bp region of the COWP gene, which was then subjected to restriction endonuclease digestion by *Rsa*I. The digestion products were separated by agarose (3% w/v) gel electrophoresis, visualised using Sybr Green I (1x solution) and recorded using a digital imaging system (AlphaImager, Alpha Innotech). Product sizes were confirmed by comparison with a DNA molecular weight standard marker (Invitrogen).

Bi-directional sequencing of the GP60 gene

C. hominis isolates were analysed by DNA sequence analysis of the GP60 gene using a nested PCR protocol (Alves *et al*., 2003). The primary PCR primers were AL3531 (5'-ATA GTC TCC GCT GTA TTC-3') and AL3535 (5'-GGA AGG AAC GAT GTA TCT-3') . The secondary PCR primers were AL3532 (5'- TCC GCT GTA TTC TCA GCC-3') (GP60X2F) and AL3534 (5'-GCA GAG GAA CCA GCA TC-3') (GP60X2R), producing final fragments of 800 to 850 bp. Each PCR sample contained 1X PCR buffer (Qiagen, Crawley, UK), (containing 1.5mM $MgCl_2$), 1X Qiagen Q solution, 200 µM of each deoxynucleoside triphosphate, 300 nM of the forward and reverse primers, 2 units of Qiagen HotStar *Taq* polymerase, and 2 µl of DNA template (for primary PCR) or 2 µl of primary PCR product (for secondary PCR) in a 100 µl reaction. The PCR program for both the primary and secondary PCR comprised 40 cycles of denaturation at 94°C for 45 s, annealing at 52°C for 45 s, and extension at 72°C for 60 s, with an initial denaturation at 95°C for 15 min and a final extension at 72°C for 10 min. All PCR products were purified with the Qiaquick PCR purification kit (Qiagen, Crawley, UK) and submitted to GeneService, Cambridge for sequencing. This was done in both directions using forward and reverse primers GP60X2F and GP60X2R and the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, UK.) on an ABI3730 automated sequencer (Applied Biosystems, UK.).

DNA sequence analysis and GP60 allele identification (see Figure 1, page 12)

The sequences were received in FASTA format with accompanying dye traces (ABI files) from GeneService in Cambridge, UK. The forward sequences were edited at the 5' and 3' termini to remove all low quality data. All forward sequences were then edited to all begin at the same position, 15 bases before the first triplet of the microsatellite. The reverse sequences were amended using the reverse complement option of the BioEdit programme, and edited to remove low quality data. The 5' sequences of the reverse-complemented data were adjusted so that they began at the same position upstream of the microsatellite as did the forward sequences.

All good quality forward/reverse sequences were assembled into an input file for the ClustalX alignment programme. The alignment was run, and isolates with identical microsatellite and GP60 profiles recorded. Ambiguities between pairs of forward/reverse sequences were resolved by consulting the sequence trace files. Reverse sequences which did not extend as far as the microsatellite were excluded from the alignment, as were sequences with poor quality read data or excessive mis-called bases.

To facilitate rapid identification of defined *C. hominis* alleles, a separate ClustalX input file was assembled using sequence from the GenBank database representing one high quality, extensive example each of *C. hominis* allele family, Ia to Ig. Each sequence representing these five alleles had approximately 900 bases of sequence from the GP60 gene. One data file from isolates identified in the first alignment as having unique profiles were then added to the input file, and the alignment run. The majority of unique isolate types aligned unambiguously with one of the five *C. hominis* allele families Ia to Ie. Where this was not the case, the nucleotide sequences were accessed and aligned using the Basic Local Alignment Sequence (BLAST) tool (US National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) against all the sequences on GenBank. A small number seemed to be variants or subtypes of existing

types, particularly Ia and Id. All sequences which could not be assigned a definitive Ia-Ie classification were sent to Dr Lihua Xiao at CDC, Atlanta, for further investigation.

Once the initial Ia to Ig allele family had been defined, the microsatellite triplet codons were categorised, and a full genotype profile produced. Within each allele family there exists multiple subtypes, varying from each other according to the number of trinucleotide repeats (TCA, TCG or TCT) which code for the amino acid serine (Sulaiman *et al*., 2005). Further, the presence and number of repeats of the sequence AA(A/G)ACGGTGGTAAGG after the microsatellite region (repeat R1, R2...) provides another element to the nomenclature (Gatei *et al*., 2006; Lihua Xiao, personal communication). This is the basis for the designation of the complete subtype family name (Figure 1).

Subtypes were assigned to each isolate and recorded on a single central spreadsheet to prevent results from being overlooked or duplicated. Sequencing data were archived using a BioNumerics database (www.applied-maths.com) to permit rapid identification of isolates, create a permanent archive of profiles and to produce a phylogenetic tree for illustrative purposes (Figure 2). Briefly, an initial pairwise alignment similarity matrix of consensus sequences was created using the un-weighted pair group method with arithmetic mean (UPGMA), and multiple alignment performed. Clustering was based on the neighbour-joining algorithm with the Kimura 2 parameter correction. The *Cryptosporidium meleagridis* GP60 sequence reported by Sulaiman *et al.* (2005), accession number AF401499 was used as an out-group for the tree. Bootstrap analysis using 1000 replicates was performed.

RESULTS

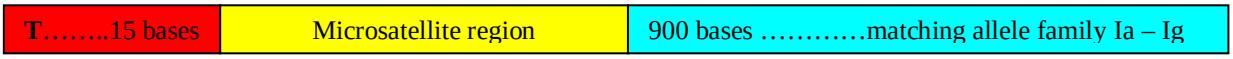
GP60 PCR EQA, test sensitivity and typability

Six of the twelve previously tested EQA samples amplified with the GP60 primers first time and a further two amplified at a second attempt. Four samples did not amplify, although re-testing the DNA with the routine COWP PCR showed that the DNA had not degraded. All of the eight GP60 positive samples had the same alleles demonstrated previously. These were then included in each GP60 PCR batch as control samples.

Of the 286 isolates investigated (Appendix 1), 254 were typable and 30 were not typable either because they did not amplify in the PCR (28 isolates) or because the sequencing results were equivocal (two isolates). Therefore the overall typability was 89.4%.

Of the 254 typable isolates, 250 matched sequences previously reported on GenBank. Four isolates had no close homologues on the GenBank database, matching at less than 96% identity. These were named according to the nomenclature procedure demonstrated in Figure 1 and where the quality of the data permitted, submitted to Genbank. The distribution of alleles is shown in Table 3.

The first part of the subtype name is the allele family and is one character from a series Ia through to Ig. This classification is based on variations in the approximately 900 bases of DNA sequence downstream of the microsatellite region.



Examples of all the gene sequences of allele families Ia-Ig are contained in the public database GenBank. A defined common start point for the final alignment is identified from a T base 15 bases before the start of the microsatellite region. The sequence downstream of the microsatellite is compared to each individual allele family using a computerised alignment programme.

This microsatellite region contains repeats of the triplets TCA and TCG (and sometimes TCT) and provides the subtype nomenclature that appears after the allele family name:

The number of 'TCA' triplet repeats in the microsatellite is counted e.g.:

TTTCTGTTGAGAGCTCATCATCATCATCATCATCATCGTCATCATCGTCAACAAC…
1 2 3 4 5 6 7 8 9 10

'TCA'= 10

Then the number of 'TCG' triplet repeats in the microsatellite is counted e.g.:

TTTCTGTTGAGAGCTCATCATCATCATCATCATCATCGTCATCATCGTCAACAAC…
1 2

'TCG' = 2

The number of TCA triplets makes up the 'A' component of the genotype; the number of 'TCG' triplets the 'G' component. The example above is therefore subtype A10G2

In a similar manner, the third triplet base occasionally showed transversion from 'A' to 'T', making the triplet 'TCT'. The number of 'TCT triplets is designated the 'T' parameter, and is recorded after the 'G' component, e.g. A9G2T1 etc.

The R repeat is located downstream of the serine-encoding microsatellite, and this parameter ('R') records the number of contiguous copies of a short, degenerate repeat sequence (AA(A/G)ACGGTGGTAAGG). If this repeat is present, it is recorded after the microsatellite designation, e.g. A9G2T1R3

The above example matched exactly with the allele family Ib sequence from the database, and is subtype A10G2. Hence the full genotype for this isolate is IbA10G2**.**

Figure 1. Outline of procedure to assign GP60 allele family and subtypes to *C. hominis* isolates

Table 3. Distribution of GP60 subtypes between sample sets and comparison with published data

	Project sample set				Interpretation
GP60 family and subtype alleles	**Sporadic cases from case control study**	**Drinking waterborne outbreak 1**	**Drinking waterborne outbreak 2**	**Historical samples prior to outbreak 2**	**Percentage sequence similarity, GenBank accession number and reference**
IaA12R3	1				No UK homologues. Nepal 99.9% AY167595 (Wu *et al*., 2003)
IaA18R2				1	No UK homologues. Uganda 99.8% AY873783 (Akiyoshi *et al*., 2006)
IaA22R1	1				Not previously reported on GenBank.
IaA23R4	1				No UK homologues. USA 99.8% AF164504 (Strong *et al*., 2001)
Presumptive IaA25	1				Not previously reported on GenBank
IaA30R3	1				No UK homologues. Source not stated 100% AF403170 (Peng *et al*., 2001) Uganda 99.9% XM_663000 (TU502) (Xu *et al*., 2004)
IbA9G2	1				No UK homologues. 99.6% identity with: Japan AY167596 (Wu *et al*., 2003) Portugal AY166807 (Alves *et al*., 2003) South Africa AF440626, AF440628, AF440637, AF440640 (Leave *et al*., 2002) Source not stated AF374344, AF374348, AF374352 (Ong *et al*., unpublished) USA AF528762 (Sturbaum *et al*., 2003); AF164498, AF164499 (Strong *et al*., 2001)
IbA10G2	92	63	63	20	Leicestershire (CRU unpublished data) Northern Ireland (Glaberman *et al*., 2002) W3950: 100% identity with isolates from: France AY702626 (Cohen *et al*., 2006) South Africa AF440626 (Leave *et al*., 2002) USA AF164499 (Strong *et al*., 2001)
IdA15G1		1			No UK homologues. Bangladesh 100% AY700388 (Hira *et al*., unpublished)
IeA11G3T3		6		1	East Sussex (CRU unpublished data) W11639: 99.4% identity with isolates from: Japan AY167593 (Wu *et al*., 2003) Portugal AY166808 (Alves *et al*., 2003) Kuwait AY738184 (Sulaiman *et al*., 2005) Peru AF528760 (Sturbaum et al., 2003) Bangladesh AY700389 (Hira *et al*., unpublished)
IfA12G1	1				No UK homologues. South Africa 100% AF440639 (Leave *et al*., 2002)
IgA24	2				Not previously reported on GenBank but matched that reported by Glaberman *et al*. (2002) in Northern Ireland (Dr Lihua Xiao, personal communication) Sequences submitted to GenBank (accession numbers EF214734 and EF214735).
NEGATIVE	13	6	8	2	
Equivocal	1				
TOTAL	115	76	71	24	

GP60 allele families and subtypes by sample set

Sporadic cases: Of 115 *C. hominis* isolates, 14 were not typable as 13 did not amplify and one gave an equivocal reaction. Therefore the sample set typability is 87.8%.

Of the typable isolates, there were nine identified types (Table 3), 92 of 101 typable isolates fell into a single type IbA10G2. Each of the other identified types contained only a single isolate member except for IgA24 which contained two. This gives a discriminatory power of 0.171, which is very low.

When those isolates which had the IbA10G2 allele were compared with those having other alleles, a history of recent foreign travel was more common for the other alleles. However, this was not statistically significant p=0.1374 (Two-tailed Fishers exact test), OR=3.01 (95% CI 0.59 to 16.20) (Table 4).

Table 4. Foreign travel and GP60 alleles in sporadic cases

Foreign Travel	IbA10G2	Other alleles	TOTAL
No	65	4	69
Yes	27	5	32
TOTAL	92	9	101

All five of the other types with a reported travel history had travelled outside of Europe, three to Pakistan (alleles IaA12R3, IaA22R1 and IaA30R3), one to Kenya (allele IaA25) and one to New Zealand (IgA24) (Table 5). Only three of the IbA10G2 cases were known to have travelled outside of Europe (Tunisia and Turkey).

Table 5. Destination of foreign travel and GP60 alleles in sporadic cases

Destination	IbA10G2	Other alleles
BALEARIC ISLANDS	3	
CANARY ISLANDS	5	
CYPRUS	3	
FRANCE	4	
GREECE	2	
KENYA		1 (IaA25)
Not Known	1	
NEW ZEALAND		1 (IgA24)
PAKISTAN		3 (IaA12R3, IaA22R1 and IaA30R3)
SPAIN	6	
TUNISIA	2	
TURKEY	1	
Grand Total	27	5

Four cases that had not travelled outside Europe had non-IbA10G2 alleles. Apart from one IgA24, these were different (one IaA23R4 and one IfA12G1) from the alleles of cases that had travelled outside Europe. The relationship between travel outside of Europe and GP60 alleles was highly statistically significant p=0.00008 (Fishers exact test), OR=37.08 (95% CI 4.76 to 303.65) (Table 6). No other epidemiological associations were present.

Table 6. Travel outside Europe and GP60 alleles in sporadic cases

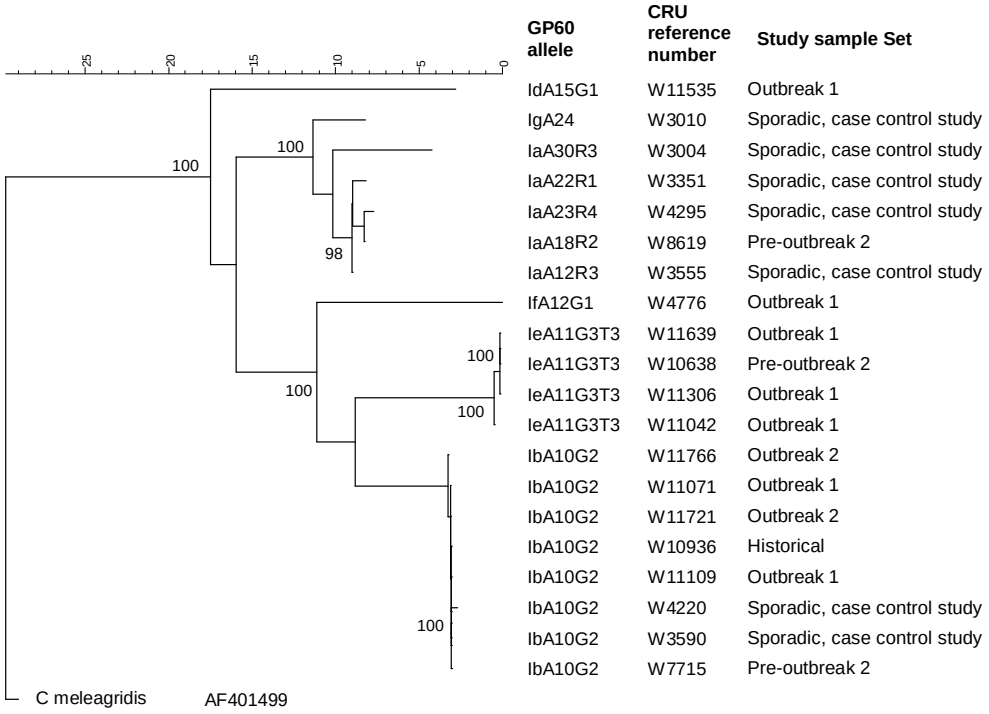
Travel out of Europe	IbA10G2	Other alleles	TOTAL
No	88	4	92
Yes	3	5	8
TOTAL	91	9	100

Drinking waterborne outbreak 1 (south east England): Of 76 isolates, six were not typable. Therefore the sample set typablility is 92%. The typable isolates belonged predominantly to two previously described alleles: 63 were IbA10G2, six were IeA11G3T3 and one was IdA15G1. These data were provided to Neville Verlander, HPA Centre for infections Statistics Unit, Colindale. Two sets of analyses were performed, one using all seventy six cases and the other restricted the cases to those with the *C. hominis* IbA10G2 allele only. Multivariable analysis for all 76 cases showed that there was a dose-response relationship between number of glasses of unboiled tap water and being a case of cryptosporidiosis, with an odds ratio of 1.29 per glass (95% CI: 1.12 to 1.48). Results were similar when restricting the cases to allele 1bA10G2; the odds ratio per glass of unboiled tap water being 1.35 (95% CI: 1.16 to 1.57).

Drinking waterborne outbreak 2 (north west Wales): Of 71 isolates, eight were not typable. Therefore the sample set typability is 88.7%. The 63 typable isolates all belonged to the IbA10G2 subtype family.

North west Wales isolates prior to outbreak 2: Of 24 isolates, two were not typable. Therefore the sample set typability is 91.7%. The 22 typable isolates were predominantly IbA10G2 (20 isolates), one was IaA18R2 and one was IeA11G3T3.

Comparision of isolates: the neighbour-joining tree illustrating clustering of GP60 sequences for representative isolates from the different study groups confirms that the sequences clustered into discrete groups reflecting the allelic families.



GP60 allele	CRU reference number	Study sample Set
IdA15G1	W11535	Outbreak 1
IgA24	W3010	Sporadic, case control study
IaA30R3	W3004	Sporadic, case control study
IaA22R1	W3351	Sporadic, case control study
IaA23R4	W4295	Sporadic, case control study
IaA18R2	W8619	Pre-outbreak 2
IaA12R3	W3555	Sporadic, case control study
IfA12G1	W4776	Outbreak 1
IeA11G3T3	W11639	Outbreak 1
IeA11G3T3	W10638	Pre-outbreak 2
IeA11G3T3	W11306	Outbreak 1
IeA11G3T3	W11042	Outbreak 1
IbA10G2	W11766	Outbreak 2
IbA10G2	W11071	Outbreak 1
IbA10G2	W11721	Outbreak 2
IbA10G2	W10936	Historical
IbA10G2	W11109	Outbreak 1
IbA10G2	W4220	Sporadic, case control study
IbA10G2	W3590	Sporadic, case control study
IbA10G2	W7715	Pre-outbreak 2

Figure 2. Neighbour-joining tree illustrating clustering of GP60 sequences for representative alleles. Scale represents sequence distance (%). Bootstrap values above 50% are shown.

DISCUSSION

The most common GP60 *C. hominis* allele, IbA10G2, represented over 90% of all isolates. This subtype has previously been detected in the UK in sporadic cases in Leicestershire (CRU unpublished data) and outbreak cases from Northern Ireland (Glaberman *et al*., 2002). The allelic family Ib has been recognised as a major cause of waterborne outbreaks of *C. hominis* in the UK (Glaberman *et al*., 2002), US – Milwaukee outbreak (Zhou *et al*., 2003) and France (Cohen *et al*., 2006) and in sporadic cases in Australia (Chalmers *et al*., 2005). The conclusions of Cohen *et al*. (2006) that Ib is the predominant allele associated with waterborne outbreaks can be explained by data from this study by the hypothesis that this is simply because this is the most common allele in human cryptosporidiosis (and therefore the most common *C. hominis* allele contaminating water from human sewage). During the environmental survey undertaken as part of the outbreak investigations in north west Wales, *C. hominis* IbA10G2 was detected in sewage influent and effluent and in raw and treated water (Outbreak Report: North West Wales, 2005). IbA10G2 also predominates in the Netherlands (Peter Wielinga, personal communication) and since non-IbA10G2 infections appear to be associated with travel to non-European destinations, it may be hypothesised that this is the predominant allele circulating in Europe.

However, by itself, GP60 sequence typing has very low discriminatory power (Hunter-Gaston discriminatory power of sporadic cases = 0.171). None the less, the evidence from this and other studies (Mallon *et al*., 2003; Hunter *et al*., 2007) is that *C. hominis* appears to be highly conserved in the UK. Investigation of *C. hominis* isolates at a further seven mini and microsatellite markers located on different chromosomes is currently underway in collaboration with Professor Giovanni Widmer and Dr Sultan Tanriverdi, Tufts University, MA. Preliminary results indicate that these and previous findings are supported. However, it must be borne in mind that the picture may change over time. During this project we tested isolates collected each year since 2000 from north west Wales, which showed very little variation although longer more extensive studies are required to demonstrate stability over time.

Allele IeA11G3T3 was found in both north west Wales and in the outbreak in the south east where it has was also detected in separate outbreak cases from East Sussex (Cryptosporidium Reference Unit unpublished data), linked to swimming pool use. The allele family Ie has been found worldwide in sporadic cases of cryptosporidiosis.

Of the less common subtypes identified, previous identifications are generally reported from further afield. However, it is interesting to note that of the four cases with non-IbA10G2 alleles that did not report foreign travel, one had the IgA24 allele which matched an isolate from Northern Ireland (Glaberman *et al*., 2002) and may well circulate in the UK, one (IbA9G2) appears to be widespread, while the remaining two, IaA23R4 and IfA12G1 matched isolates from USA and South Africa respectively (Strong *et al*., 2001; Leave *et al*., 2002).

One of the Pakistani isolates, IaA12R3 has been isolated previously from a patient from Nepal (Wu *et al*., 2003), while another, IaA30R3, has been isolated from Ugandan children (Akiyoshi *et al*., 2006). A previously unreported allele (IaA22R1) found in this study has no closely related sequences on GenBank.

Only one isolate out of 24 from the historical set from north west Wales had not been found in the UK: IaA18R2 has been reported previously from Ugandan children (Akiyoshi *et al*., 2006). Likewise, IdA15G1 was found in one sample from the outbreak from the south east and has been found previously in Bangladeshi children (Hira *et al*., unpublished).

The sensitivity of the GP60 PCR affects typability of the method which, over all was 89.5% (inter-set range 87.8% to 92.0%). A quarter of the original panel of 12 EQA samples failed to amplify with the GP60 primers, although DNA was shown not to be degraded in storage. Typability of this method is also dependant on obtaining good quality sequence data. While this method has been applied to environmental and water samples (where the numbers of oocysts present are lower than in clinical samples) the sensitivity of the test is not as great as that applied for species/genotype identification.

Care was taken during this study to follow the emerging consensus for naming of new GP60 subtypes. The system described by Sulaiman *et al*. (2005) portrays the "R" repeat as consisting of an "ACATCA" directly after the microsatellite region. However, the same group in a later paper used R to describe the longer repeat described in Fig.1 in *C. hominis* isolates but did not explain the change of system (Gatei *et al*., 2006). Several studies have implemented the nomenclature scheme for studies of isolates from widely distributed *Cryptosporidium* isolates (Table 1). For there to be universally accepted nomenclature there is a need for clarification of the system. All newly identified subtypes should be named according to the accepted system and their sequences along with source data submitted to GenBank.

The GP60 sequencing method is applicable to epidemiological studies for *Cryptosporidium*, but indigenous UK *C. hominis* isolates show very little variation even when the extent of heterogeneity is explored using multiple markers. This is currently being further substantiated in a collaborative study with Tufts University.

SECTION 2: INVESTIGATION OF SSCP

METHOD

SSCP ITS-2

ITS-2 amplification was performed as described by Gasser *et al*. (2004). PCR products were run on 3% Phorecus agarose gels to evaluate the success of PCR. SSCP of PCR-positive samples was performed as described by Gasser *et al*. (2004), except that they were run for 15 hr instead of 14 hr and were stained with SYBRGold (Sigma) for 1.5 hr and de-stained for 45 min. Gel images were taken using an AlphaImager system (Alpha Innotech) and banding patterns analysed using BioNumerics software (Applied Maths).

SSCP GP60

GP60 SSCP was performed according to a method adapted from that described by Wu *et al*. (2003). GP60 amplification was performed according to the nested PCR method modified from that described by Alves *et al*. (2003) for use at CRU using primary PCR primers AL3531 and AL3535, and secondary primers AL3532 and AL3534. PCR products were digested with *Rsa*I and fragments analysed by SSCP on GMA gels using the thermoregulated SEA 2000 apparatus (Elchrom). SSCP was performed in an air-conditioned (set point 19°C) Swansea University laboratory to improve temperature control. Gels were stained with SYBRGold for 1.5 hr and de-stained for 45 min. Gel images were taken using an AlphaImager system and banding patterns analysed using BioNumerics software (Applied Maths). The sample W11402 was run in three lanes across each gel to allow normalisation within and between gels using BioNumerics. Calculation of band pattern similarity was performed using the densitometric curve-based Pearson similarity coefficient in BioNumerics. Dendograms were produced using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method in BioNumerics.

Evaluation panel samples (10 *C. hominis*, 10 *C. parvum*) were chosen to represent the subtype families available as well as examples of closely related families. Library samples were chosen to represent the different GP60 sequence subtypes identified to date at the CRU.

RESULTS: SSCP ITS-2

Optimisation

SSCP of the ITS-2 region was performed as described by Gasser *et al.* (2004), therefore little optimisation was required for standardisation. SSCP was performed in an air conditioned laboratory (set at 19°C) and gels stained as described above.

Evaluation

Initial evaluation was performed by testing samples previously tested by Dr Robin Gasser, as well as selected samples from the drinking water-associated outbreaks described above. Foreign traveller samples previously tested by Dr Robin Gasser at the University of Melbourne were used as a library of different subtypes. Banding patterns were found to be very similar for many of the different subtypes. This, in conjunction with a significant degree of inter- and intra-gel variation in band positions made identification of subtypes very difficult. It was not possible to develop a reliable procedure for comparison of patterns from different gel runs using BioNumerics, making the method difficult to evaluate fully. In summary the method was found to be of possible value for investigation of variability within discrete sample sets but not reproducible enough for large-scale subtyping investigations. ITS-2 SSCP was therefore deemed unsuitable for investigation of the prevalence and epidemiological distribution of *C. hominis* subtypes in the study samples.

RESULTS: SSCP GP60

Optimisation

Preliminary experiments (data not shown) were performed as described by Wu *et al*. (2003) with the modifications of replacing the external primers with the Alves *et al*. (2003) GP60 PCR (AL3531 and AL3535) since the original Wu *et al*. (2003) reverse primer contains a mismatched base, and of using the Elchrom SSCP apparatus.. However, band intensity was very variable hampering profile analysis. Increasing the number of PCR cycles provided no improvement.

The method was further developed by nesting the PCR using the Alves *et al*. (2003) GP60 PCR internal primers (AL3532 and AL3534). Consistently strong bands were observed and profiles readily observed. This was therefore adopted as the preferred method. Digestion of the PCR product with the restriction enzyme *Rsa*I was found to be essential to create fragments of the correct length for SSCP analysis.

Evaluation

Inter-assay reproducibility was evaluated by running evaluation samples on separate SSCP runs. Inter-run reproducibility was assessed by comparing patterns produced after analysis of samples for the first time with patterns produced in subsequent runs. Similarity ranged from 87.8-98.2% (mean, 95.2%).

Cluster analysis was performed on the evaluation panel (Appendix 2). Overall there was a trend for repeats to cluster with their counterparts and for the clusters to follow the sequence-based subtyping result. There were several exceptions, with some of the more closely related sequence subtypes clustering together (IIaA15G1R1, IIaA17G1R1, IIaA18G1R1, IIaA21G1R1) and IIdA17G1 also clustering with these subtypes. If a cut-off of 86% similarity is applied to the cluster analysis all of the isolates cluster with counterparts from the same sequence subtype

family apart from one of three OZ16 patterns and three of eight W8670 patterns which cluster within the IIa group. Both of these samples showed large inter-run pattern differences suggesting the possibility of an experimental error or the presence of mixed subtypes. When these were removed from the analysis the isolates clustered according to the sequence data (Appendix 3).

Library construction

The library panel was analysed using the same clustering method. All but four isolates gave patterns, including nine that had previously given poor sequencing results. The overall trend was for isolates to cluster appropriately.

DISCUSSION

Reading the DNA of the parasite genome (sequencing) provides a high level of data. For investigation of population structures, multiple loci are required and this is expensive and time consuming. Alternative methods to type isolates within species for epidemiological purposes may be more appropriate once the population structure is understood. The GP60 PCR-RFLP-SSCP is one method that may have the potential to identify isolates to the family level within species. The creation of a library allows profiles from samples of unknown subtype to be assessed for their similarity to previously identified patterns, and identification improved by running appropriate numbers and types of standards in each gel. GP60 PCR-RFLP-SSCP offers a potentially rapid method for subtyping large numbers of isolates which would be particularly useful for identifying unusual subtypes within a background of a prevailing subtype. However, further standardisation is required but it does provide supporting evidence especially for isolates that are difficult to sequence due to generation of poor sequence traces.

It is worthy of note at this point that DNA sequence analysis is the most widely used analytical method for *Cryptosporidium* typing. Most development has gone into multilocus sequence typing of limited numbers of samples for research purposes. It is desirable to further develop, in collaborative work, rapid tools for epidemiological purposes. These should not be limited to the methods described here but take advantage of developing technologies.

SECTION 3
CONCLUSIONS

- *C. hominis* is highly conserved in indigenous UK cases.
- Alleles other than IbA10G2 are linked to recent foreign travel outside Europe.
- Predominance of Ib allelic family and more specifically IbA10G2 in European drinking waterborne outbreaks reflects its high prevalence in sporadic cases.
- Indications are that that allele IbA10G2 has been stable in the UK population for some time but it is not possible to predict that this will remain so.
- To encourage consistency in naming new GP60 alleles (and indeed any profiles) there is a need for a publicly accessible international database for naming and storing submitted sequences. This would also provide a basis on which to develop a future multi-locus database and support further method development.

ACKNOWLEDGEMENTS

This project was funded by Defra with contract management by DWI. Any views expressed are not necessarily those of Defra.

Dr Lihua Xiao of the Centers for Disease Control and Prevention (USA) for his assistance with identifying unusual GP60 sequences.

REFERENCES

Akiyoshi DE, Tumwine JK, Bakeera-Kitaka S. and Tzipori S. Subtype analysis of *Cryptosporidium* isolates from children in Uganda. Journal of Parasitology 2006: 92; 1097-1100.

Alves M, Xiao L. Sulaiman I, Lal AA. Matos O and Antunes F. Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. Journal of Clinical Microbiology 2003; 41: 2744-2747.

Alves M, Xiao L, Antunes F, Matos O. Distribution of *Cryptosporidium* subtypes in humans and domestic and wild ruminants in Portugal. Parasitology Research 2006; 99: 287-292.

Anon. The development of a national collection for oocysts of *Cryptosporidium*. Final Report to DEFRA: Drinking Water Inspectorate. Foundation for Water Research, Marlow, Bucks, UK 2002.

Anon. Investigation of *Cryptosporidium* clinical isolates and analysis with epidemiological data. Final report to Drinking Water Inspectorate 2005.

Anon. Outbreak of cryptosporidiosis in north west Wales, 2005. Report of the outbreak control team. Anglesey County Council, Anglesey Local Health Board, Centre for Research Ino Environment and Health, Dwr Cymru, Environment Agency Wales, Gwynedd County Council, Gwynedd Local Health Board, National Public Health Service for Wales. 2005

Cacciò S, Homan W, Camilli R, Traldi G, Kortbeek T, Pozio E. A microsatellite marker reveals population heterogeneity within human and animal genotypes of *Cryptosporidium parvum. Parasitology* 2000; 120: 237-244.

Cacciò S, Spano F, Pozio E. Large sequence variation at two microsatellite loci among zoonotic (genotype C) isolates of *Cryptosporidium parvum*. *International Journal for Parasitology* 2001; 31: 1082-1086.

Chalmers RM, Ferguson C, Cacciò SM, Gasser RB, El-Osta YGA, Heijnen L, Xiao L, Elwin K, Hadfield S, Sinclair M, Stevens M. Direct comparison of selected methods for genetic categorisation of *Cryptosporidium parvum* and *Cryptosporidium hominis* species. International Journal for Parasitology 2005; 35: 397-410 .

Cohen S, Dalle F, Gallay A, Di Palma M, Bonnin A, Ward HD. Identification of Cpgp40/15 Type Ib as the predominant allele in isolates of *Cryptosporidium* spp. From a waterborne outbreak of gastroenteritis in South Burgandy, France. Journal of Clinical Microbiology 2006: 44; 589-591.

Feltus DC, Giddings CW, Schneck BL, Monson T, Warshauer D, McEvoy JM. Evidence Supporting Zoonotic Transmission of *Cryptosporidium* spp. in Wisconsin. Journal of Clinical Microbiology. 2006; 44: 303-8.

Gasser RB, Abs El-Osta Y, Prepens S, Chalmers RM. An improved “cold SSCP” for the genotypic and subgenotypic characterisation of *Cryptosporidium*. Molecular and Cellular Probes 2004; 18: 329-332.

Gatei W, Das P, Dutta P, Sen A, Cama V, Lal AA, Xiao L. Multilocus sequence typing and genetic structure of *Cryptosporidium hominis* from children in Kolkata, India. Infection, Genetics and Evolution 2006 E-publication ahead of print (accessed 13/12/06).

Glaberman S, Moore JE, Lowery CJ, Chalmers RM, Sulaiman I, Elwin K, Rooney PJ, Millar BC, Dooley JSG, Lal AA, Xiao L. Three drinking-water-associated cryptosporidiosis outbreaks, Northern Ireland. Emerging Infectious Diseases 2002; 8: 631-633.

Goh S, Reacher M, Casemore DP, Verlander NQ, Chalmers RM, Knowles M, Williams J, Richards S. Risk factors for sporadic cryptosporidiosis in North Cumbria, North West of England 1996 to 2000. Emerging Infectious Diseases 2004; 9: 1005-1017.

Hunter PR, Hughes LS, Woodhouse S, Syed Q, Verlander N, Chalmers RM and members of the project steering committee. Case-control study of sporadic cryptosporidiosis with genotyping. Emerging Infectious Diseases 2004a; 10: 1241-1249.

Hunter PR, Hughes S, Woodhouse S, Raj N, Syed Q, Chalmers RM, Verlander NQ, Goodacre J. Health sequelae of human cryptosporidiosis in immunocompetent patients. Clinical Infectious Diseases 2004b; 39: 504-510.

Lauritsen JM. (Editor) EpiData Data Entry, Data Management and basic Statistical Analysis System. Odense Denmark, EpiData Association, 2000-2005.

Mallon ME, MacLeod A, Wastling JM, Smith H, Tait A. Multilocus genotyping of *Cryptosporidium parvum* type 2: population genetics and sub-structuring. Infection, Genetics and Evolution 2003; 3: 207.

McLauchlin J, Amar C, Pedraza-Diaz S, Nichols GL. Molecular epidemiological analysis of *Cryptosporidium* spp. in the United Kingdom: results of genotyping *Cryptosporidium* spp. in 1,705 fecal samples from humans and 105 fecal samples from livestock animals. Journal of Clinical Microbiology 2000; 11; 3984-3990.

Neira-Munoz E and Smith J on behalf of the outbreak control team. Report on cryptosporidiosis in south east Hampshire September-November 2005. St Mary's Hospital, Finchdean House, Milton Road, Portsmouth 2006.

Peng MM, Meshnick SR, Cunliffe NA, Thindwa BD, Hart CA, Broadhead RL, Xiao L. Molecular epidemiology of cryptosporidiosis in children in Malawi. Journal of Eukaryotic Microbiology 2003: 50 (suppl): 557-559.

Peng MM, Matos O, Gatei W, Das P, Stantic-Pavlinic M, Bern C, Sulaiman IM, Glaberman S, Lal AA, Xiao L. A comparison of *Cryptosporidium* subgenotypes from several geographic regions. Journal of Eukaryotic Microbiology 2001; 28S-31S.

Statacorp. 2003. *Stata Statistical Software: Release 8.2*. College Station, TX: Stata corporation.

Strong WB, Gut J, Nelson RG. Cloning and sequence analysis of a highly polymorphic *Cryptosporidium parvum* gene encoding a 60-kilodalton glycoprotein and characterization of its 15- and 45-kilodalton zoite surface antigen products. Infection and Immunity 2000; 68: 4117-4134.

Sturbaum GD, Jost BH and Sterling CR. Nucleotide changes within three *Cryptosporidium parvum* surface protein encoding genes differentiate genotype I from genotype II isolates. Molecular and Biochemical Parasitology 2003: 128; 87-90.

Sulaiman IM, Hira PR, Zhou L, Al-Ali FM, Al-Shelahi FA, Shweiki HM, Iqbal J, Khalid N, Xiao L. Unique endemicity of cryptosporidiosis in children in Kuwait. Journal of Clinical Microbiology 2005; 43: 2805-2809.

Trotz-Williams LA, Martin DS, Gatei W, Cama V, Peregrine AS, Martin SW, Nydam DV, Jamieson F, Xiao L. Genotype and subtype analyses of *Cryptosporidium* isolates from dairy calves and humans in Ontario. Parasitology Research. 2006; 99:346-52.

Wu Z, Nagano I, Booonmars T, Nakada T and Takahashi Y. Intraspecies polymorphism of *Cryptosporidium parvum* revealed by PCR-restriction fragment length polymorphism (RFLP) and RFLP-single strand conformational polymorphism analysis. Applied and Environmental Microbiology 2003; 69: 4720-4726.

Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA. The genome of *Cryptosporidium hominis* TU502. Nature 2004; 431: 1107-1112.

Zhou L, Singh A, Jiang J, Xiao L. Molecular surveillance of *Cryptosporidium* spp. in raw waste water in Milwaukee: implications for understanding outbreak occurrence and transmission dynamics. Journal of Clinical Microbiology 2003; 41: 5254-5257.

APPENDIX 1. Results of GP60 sequence analysis of isolates used in this study.

CRU reference number	Sample set	GP60 allele
W4776	Sporadic; case control study	IfA12G1
W4771	Sporadic; case control study	IbA10G2
W4762	Sporadic; case control study	IbA10G2
W4733	Sporadic; case control study	IbA10G2
W4725	Sporadic; case control study	IbA10G2
W4691	Sporadic; case control study	IbA10G2
W4690	Sporadic; case control study	IbA10G2
W4687	Sporadic; case control study	IbA10G2
W4686	Sporadic; case control study	IbA10G2
W4672	Sporadic; case control study	IbA10G2
W4670	Sporadic; case control study	IbA10G2
W4665	Sporadic; case control study	IbA10G2
W4663	Sporadic; case control study	IbA10G2
W4660	Sporadic; case control study	IbA10G2
W4655	Sporadic; case control study	IbA10G2
W4641	Sporadic; case control study	PCR negative
W4635	Sporadic; case control study	IbA10G2
W4623	Sporadic; case control study	IbA10G2
W4621	Sporadic; case control study	PCR negative
W4620	Sporadic; case control study	IbA10G2
W4618	Sporadic; case control study	IbA10G2
W4604	Sporadic; case control study	IbA10G2
W4591	Sporadic; case control study	IbA10G2
W4583	Sporadic; case control study	IbA10G2
W4580	Sporadic; case control study	IbA10G2
W4569	Sporadic; case control study	IbA10G2
W4548	Sporadic; case control study	IbA10G2
W4540	Sporadic; case control study	IbA10G2
W4539	Sporadic; case control study	PCR negative
W4526	Sporadic; case control study	IbA10G2
W4525	Sporadic; case control study	IbA10G2
W4512	Sporadic; case control study	IbA10G2
W4511	Sporadic; case control study	IbA10G2
W4498	Sporadic; case control study	IbA10G2
W4430	Sporadic; case control study	IbA10G2
W4412	Sporadic; case control study	IbA10G2
W4410	Sporadic; case control study	IbA10G2
W4384	Sporadic; case control study	IbA10G2
W4375	Sporadic; case control study	IbA10G2
W4331	Sporadic; case control study	IbA10G2
W4312	Sporadic; case control study	IbA10G2
W4310	Sporadic; case control study	IbA10G2
W4301	Sporadic; case control study	IbA10G2
W4295	Sporadic; case control study	IaA23R4

W4293	Sporadic; case control study	IbA10G2
W4282	Sporadic; case control study	IbA9G2
W4273	Sporadic; case control study	IbA10G2
W4259	Sporadic; case control study	IbA10G2
W4258	Sporadic; case control study	IbA10G2
W4252	Sporadic; case control study	IbA10G2
W4250	Sporadic; case control study	IbA10G2
W4239	Sporadic; case control study	IbA10G2
W4238	Sporadic; case control study	IbA10G2
W4223	Sporadic; case control study	IbA10G2
W4222	Sporadic; case control study	IbA10G2
W4220	Sporadic; case control study	IbA10G2
W4206	Sporadic; case control study	IbA10G2
W4196	Sporadic; case control study	PCR negative
W4185	Sporadic; case control study	IbA10G2
W4142	Sporadic; case control study	IbA10G2
W4141	Sporadic; case control study	IbA10G2
W4140	Sporadic; case control study	IbA10G2
W4121	Sporadic; case control study	IbA10G2
W4118	Sporadic; case control study	IbA10G2
W4117	Sporadic; case control study	IbA10G2
W4107	Sporadic; case control study	IbA10G2
W4088	Sporadic; case control study	IbA10G2
W4048	Sporadic; case control study	IbA10G2
W4034	Sporadic; case control study	IbA10G2
W4016	Sporadic; case control study	IbA10G2
W3999	Sporadic; case control study	IbA10G2
W3996	Sporadic; case control study	IbA10G2
W3957	Sporadic; case control study	IgA24
W3948	Sporadic; case control study	PCR negative
W3889	Sporadic; case control study	IbA10G2
W3888	Sporadic; case control study	PCR negative
W3887	Sporadic; case control study	IbA10G2
W3871	Sporadic; case control study	IbA10G2
W3870	Sporadic; case control study	IbA10G2
W3812	Sporadic; case control study	IbA10G2
W3804	Sporadic; case control study	PCR negative
W3802	Sporadic; case control study	PCR negative
W3801	Sporadic; case control study	IbA10G2
W3796	Sporadic; case control study	IbA10G2
W3765	Sporadic; case control study	IbA10G2
W3690	Sporadic; case control study	IbA10G2
W3676	Sporadic; case control study	IbA10G2
W3670	Sporadic; case control study	PCR negative
W3590	Sporadic; case control study	IbA10G2
W3555	Sporadic; case control study	IaA12R3
W3554	Sporadic; case control study	IbA10G2
W3542	Sporadic; case control study	PCR negative
W3504	Sporadic; case control study	IbA10G2

W3499	Sporadic; case control study	IbA10G2
W3441	Sporadic; case control study	IbA10G2
W3423	Sporadic; case control study	PCR negative
W3407	Sporadic; case control study	PCR negative
W3391	Sporadic; case control study	PCR negative
W3351	Sporadic; case control study	IaA22R1
W3334	Sporadic; case control study	Presumptive IaA25
W3229	Sporadic; case control study	IbA10G2
W3227	Sporadic; case control study	IbA10G2
W3116	Sporadic; case control study	IbA10G2
W3114	Sporadic; case control study	IbA10G2
W3090	Sporadic; case control study	IbA10G2
W3015	Sporadic; case control study	IbA10G2
W3010	Sporadic; case control study	IgA24
W3009	Sporadic; case control study	IbA10G2
W3004	Sporadic; case control study	IaA30R3
W3000	Sporadic; case control study	IbA10G2
W2998	Sporadic; case control study	IbA10G2
W2955	Sporadic; case control study	IbA10G2
W2945	Sporadic; case control study	IbA10G2
W2696	Sporadic; case control study	IbA10G2
W2588	Sporadic; case control study	equivocal
W11875	Outbreak 1: south east England	IbA10G2
W11845	Outbreak 1: south east England	IbA10G2
W11807	Outbreak 1: south east England	IbA10G2
W11743	Outbreak 1: south east England	IbA10G2
W11726	Outbreak 1: south east England	IbA10G2
W11698	Outbreak 1: south east England	IbA10G2
W11677	Outbreak 1: south east England	IbA10G2
W11648	Outbreak 1: south east England	IbA10G2
W11647	Outbreak 1: south east England	IbA10G2
W11646	Outbreak 1: south east England	IeA11G3T3
W11640	Outbreak 1: south east England	IbA10G2
W11639	Outbreak 1: south east England	IeA11G3T3
W11569	Outbreak 1: south east England	IbA10G2
W11568	Outbreak 1: south east England	IbA10G2
W11543	Outbreak 1: south east England	IbA10G2
W11535	Outbreak 1: south east England	IdA15G1
W11532	Outbreak 1: south east England	IbA10G2
W11531	Outbreak 1: south east England	IbA10G2
W11528	Outbreak 1: south east England	IbA10G2
W11500	Outbreak 1: south east England	IbA10G2
W11499	Outbreak 1: south east England	IbA10G2
W11489	Outbreak 1: south east England	PCR negative
W11468	Outbreak 1: south east England	IbA10G2
W11467	Outbreak 1: south east England	IbA10G2
W11466	Outbreak 1: south east England	IbA10G2
W11465	Outbreak 1: south east England	I bA10G2
W11463	Outbreak 1: south east England	IbA10G2
W11441	Outbreak 1: south east England	PCR negative
W11425	Outbreak 1: south east England	IbA10G2
W11424	Outbreak 1: south east England	IbA10G2

W11381	Outbreak 1: south east England	PCR negative
W11380	Outbreak 1: south east England	PCR negative
W11378	Outbreak 1: south east England	IbA10G2
W11374	Outbreak 1: south east England	IbA10G2
W11371	Outbreak 1: south east England	PCR negative
W11365	Outbreak 1: south east England	IbA10G2
W11364	Outbreak 1: south east England	IbA10G2
W11362	Outbreak 1: south east England	IbA10G2
W11361	Outbreak 1: south east England	IbA10G2
W11354	Outbreak 1: south east England	IbA10G2
W11353	Outbreak 1: south east England	IbA10G2
W11351	Outbreak 1: south east England	IbA10G2
W11349	Outbreak 1: south east England	IbA10G2
W11308	Outbreak 1: south east England	IbA10G2
W11307	Outbreak 1: south east England	IbA10G2
W11306	Outbreak 1: south east England	IeA11G3T3
W11305	Outbreak 1: south east England	IbA10G2
W11304	Outbreak 1: south east England	IbA10G2
W11267	Outbreak 1: south east England	PCR negative
W11265	Outbreak 1: south east England	IbA10G2
W11241	Outbreak 1: south east England	IbA10G2
W11238	Outbreak 1: south east England	IbA10G2
W11237	Outbreak 1: south east England	IbA10G2
W11236	Outbreak 1: south east England	IbA10G2
W11235	Outbreak 1: south east England	IbA10G2
W11234	Outbreak 1: south east England	IbA10G2
W11233	Outbreak 1: south east England	IbA10G2
W11231	Outbreak 1: south east England	IeA11G3T3
W11213	Outbreak 1: south east England	IbA10G2
W11212	Outbreak 1: south east England	IeA11G3T3
W11185	Outbreak 1: south east England	IbA10G2
W11173	Outbreak 1: south east England	IbA10G2
W11155	Outbreak 1: south east England	IbA10G2
W11153	Outbreak 1: south east England	IbA10G2
W11130	Outbreak 1: south east England	IbA10G2
W11129	Outbreak 1: south east England	IbA10G2
W11128	Outbreak 1: south east England	IbA10G2
W11110	Outbreak 1: south east England	IbA10G2
W11109	Outbreak 1: south east England	IbA10G2
W11108	Outbreak 1: south east England	IbA10G2
W11107	Outbreak 1: south east England	IbA10G2
W11106	Outbreak 1: south east England	IbA10G2
W11090	Outbreak 1: south east England	IbA10G2
W11071	Outbreak 1: south east England	IbA10G2
W11042	Outbreak 1: south east England	IeA11G3T3
W10876	Outbreak 1: south east England	IbA10G2
W12340	Outbreak 2: north west Wales	IbA10G2
W12284	Outbreak 2: north west Wales	IbA10G2
W12251	Outbreak 2: north west Wales	IbA10G2
W12248	Outbreak 2: north west Wales	IbA10G2
W12245	Outbreak 2: north west Wales	IbA10G2
W12227	Outbreak 2: north west Wales	IbA10G2
W12198	Outbreak 2: north west Wales	IbA10G2
W12197	Outbreak 2: north west Wales	IbA10G2

W12182	Outbreak 2: north west Wales	IbA10G2
W12174	Outbreak 2: north west Wales	IbA10G2
W12173	Outbreak 2: north west Wales	IbA10G2
W12172	Outbreak 2: north west Wales	IbA10G2
W12160	Outbreak 2: north west Wales	IbA10G2
W12159	Outbreak 2: north west Wales	IbA10G2
W12156	Outbreak 2: north west Wales	IbA10G2
W12153	Outbreak 2: north west Wales	IbA10G2
W12141	Outbreak 2: north west Wales	IbA10G2
W12139	Outbreak 2: north west Wales	IbA10G2
W12135	Outbreak 2: north west Wales	PCR negative
W12133	Outbreak 2: north west Wales	IbA10G2
W12110	Outbreak 2: north west Wales	IbA10G2
W12077	Outbreak 2: north west Wales	IbA10G2
W12072	Outbreak 2: north west Wales	PCR negative
W12071	Outbreak 2: north west Wales	IbA10G2
W12057	Outbreak 2: north west Wales	PCR negative
W12055	Outbreak 2: north west Wales	IbA10G2
W12050	Outbreak 2: north west Wales	IbA10G2
W12023	Outbreak 2: north west Wales	IbA10G2
W12022	Outbreak 2: north west Wales	PCR negative
W12018	Outbreak 2: north west Wales	IbA10G2
W12014	Outbreak 2: north west Wales	PCR negative
W12009	Outbreak 2: north west Wales	IbA10G2
W11989	Outbreak 2: north west Wales	IbA10G2
W11986	Outbreak 2: north west Wales	IbA10G2
W11976	Outbreak 2: north west Wales	IbA10G2
W11964	Outbreak 2: north west Wales	IbA10G2
W11963	Outbreak 2: north west Wales	IbA10G2
W11962	Outbreak 2: north west Wales	IbA10G2
W11933	Outbreak 2: north west Wales	IbA10G2
W11932	Outbreak 2: north west Wales	IbA10G2
W11931	Outbreak 2: north west Wales	IbA10G2
W11912	Outbreak 2: north west Wales	PCR negative
W11865	Outbreak 2: north west Wales	IbA10G2
W11863	Outbreak 2: north west Wales	IbA10G2
W11862	Outbreak 2: north west Wales	IbA10G2
W11843	Outbreak 2: north west Wales	IbA10G2
W11830	Outbreak 2: north west Wales	IbA10G2
W11829	Outbreak 2: north west Wales	IbA10G2
W11802	Outbreak 2: north west Wales	IbA10G2
W11799	Outbreak 2: north west Wales	IbA10G2
W11766	Outbreak 2: north west Wales	IbA10G2
W11745	Outbreak 2: north west Wales	IbA10G2
W11722	Outbreak 2: north west Wales	IbA10G2
W11721	Outbreak 2: north west Wales	IbA10G2
W11709	Outbreak 2: north west Wales	IbA10G2
W11708	Outbreak 2: north west Wales	PCR negative
W11695	Outbreak 2: north west Wales	IbA10G2
W11694	Outbreak 2: north west Wales	IbA10G2
W11684	Outbreak 2: north west Wales	IbA10G2
W11683	Outbreak 2: north west Wales	IbA10G2
W11664	Outbreak 2: north west Wales	IbA10G2
W11657	Outbreak 2: north west Wales	IbA10G2

W11611	Outbreak 2: north west Wales	IbA10G2
W11609	Outbreak 2: north west Wales	IbA10G2
W11594	Outbreak 2: north west Wales	IbA10G2
W11575	Outbreak 2: north west Wales	IbA10G2
W11524	Outbreak 2: north west Wales	IbA10G2
W11252	Outbreak 2: north west Wales	IbA10G2
W11126	Outbreak 2: north west Wales	PCR negative
W11036	North West Wales Historical: 2005	IbA10G2
W10936	North West Wales Historical: 2005	IbA10G2
W10837	North West Wales Historical: 2005	IbA10G2
W10824	North West Wales Historical: 2005	IbA10G2
W10638	North West Wales Historical: 2005	IeA11G3T3
W9953	North West Wales Historical: 2005	IbA10G2
W9905	North West Wales Historical: 2005	IbA10G2
W9904	North West Wales Historical: 2005	IbA10G2
W9903	North West Wales Historical: 2005	IbA10G2
W9847	North West Wales Historical: 2005	IbA10G2
W9472	North West Wales Historical: 2005	IbA10G2
W9117	North West Wales Historical: 2005	PCR negative
W8619	North West Wales Historical: 2005	IaA18
W7717	North West Wales Historical: 2005	IbA10G2
W7715	North West Wales Historical: 2005	IbA10G2
W7420	North West Wales Historical: 2005	IbA10G2
W6871	North West Wales Historical: 2005	PCR negative
W6001	North West Wales Historical: 2005	IbA10G2
W5726	North West Wales Historical: 2005	IbA10G2
W4819	North West Wales Historical: 2005	IbA10G2
W4107	North West Wales Historical: 2005	IbA10G2
W2690	North West Wales Historical: 2005	IbA10G2
W2397	North West Wales Historical: 2005	IbA10G2
W1941	North West Wales Historical: 2005	IbA10G2
W1421	North West Wales Historical: 2005	IbA10G2
W1149	North West Wales Historical: 2005	IbA10G2

APPENDIX 2. Inter- and intra-run Pearson similarity values (%) for panel samples tested by SSCP of the GP60 gene.

A. Evaluation panel

CRU reference number	Study sample set/ Provenance	GP60 allele by sequencing	1st run*		2nd run*				3rd run*			4th run*		
			1L	2L	1L	2L	3L	1R	1L	2L	3L	1L	2L	3L
C. hominis														
W3004	Sporadic	Ia1A30R3	100		90.9							94.3	94.5	92.9
W3351	Sporadic	Ia1A22R1	100		92.3							87.7	91.1	89.7
W7656	National collection	IaA19G1R2	100		96.7	93.3			91.3	93.0				
W3227	Sporadic	1bA10G2	100		95.3							95.9	96.1	93.4
W12133	Outbreak 2	IbA10G2	100		96.3	98.4			96.3	96.2				
OZ 13	Chalmers *et al.* (2005)	IbA9G3	100		87.8							87.9	89.4	87.7
W11281	National collection	IeA11G3T3	100		93.1	96.9			97.3	96.6				
OZ 5	Chalmers *et al.* (2005)	IeA12G3T3	100		95.6							94.7	96.4	99.3
W4776	Sporadic	IfA12G1	100		96.9			96.5				97.1	98.5	98.2
W3010	Sporadic	IgA24	100		94.5									
C. parvum														
W7598	National collection	IIaA15G2R1	100		94.5	97.6			97.1	95.4		96.5		
W8824	National collection	IIaA17G1R1	100		96.5	98.3	98.4							
W7818	National collection	IIaA18G1R1	100	94.9	96.3	96.9	96.6		93.2	94.1	92.5			
W7649	National collection	IIaA17G1R1	100		98.2	98.3			97.7			97.0	96.7	
OZ 16	Chalmers *et al.* (2005)	IIaA20G3RI	100		Neg	Neg	Neg							
W8976	National collection	IIaA21G3R1	100	94.4	96.5	97.8	95.4							
W7266	National collection	IIcA5G3	100		95.9	93.6			90.2	94.0		87.6		
W7267	National collection	IIcA5G3	100		96.3	94.5			89.7	93.8		92.8	91.6	93.3
W8670	National collection	IIdA17G1	100	95.7	95.4	96.4	97.3							
W7597	National collection	IIeA7G1	100		98.2	98.8			95.0	96.1		97.6	97.1	97.5

*Numbers represent order on gel; L = left-hand gel; R = right-hand gel. Similarity values were calculated using the first run as a reference.

B. Library samples

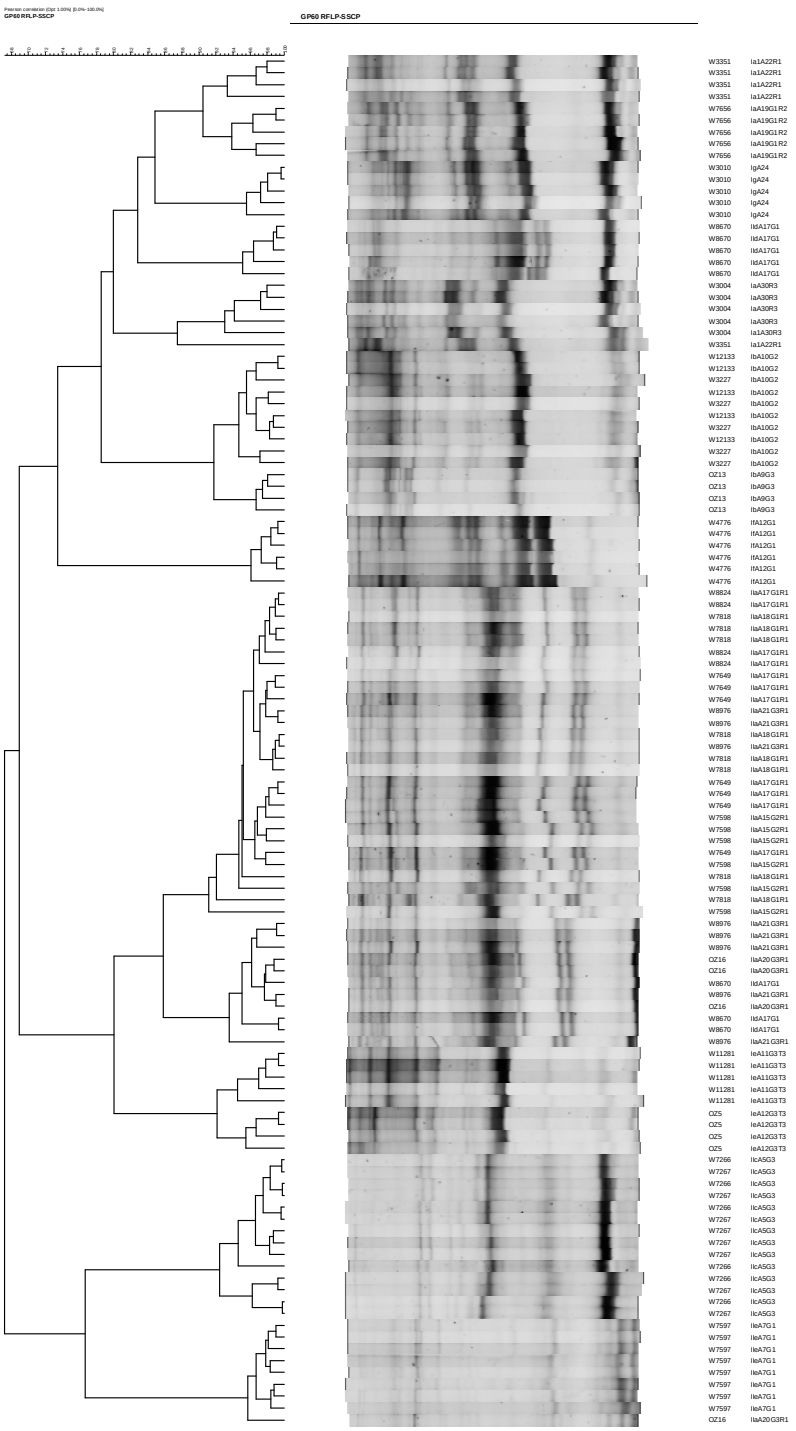
CRU reference number	Study sample set / Provenance	GP60 allele by sequencing	1st run*				2nd run*				3rd run*		4th run*
			1L	2L	3L	1R	1L	2L	3L	1R	1L	2L	1L
C. hominis													
W3334	Sporadic	IaA25	Neg										
W3555	Sporadic	Ia1A12R3	100				90.5						
W4295	Sporadic	IaA23R4	100				94.6						
W4282	Sporadic	IbA9G2	Neg				Neg						
W11009	Outbreak 1	IbA10G2	100				96.2	96.2			95.6	96.4	
W10988	National collection	IbA10G2	100				96.6				97.1	96.5	
W12141	Outbreak 2	IbA10G2	100				97.7				97.3		
W12156	Outbreak 2	IbA10G2	100				98.4				97.5		
W4384	Sporadic	1bA10G2	100				89.4						
W4034	Sporadic	1bA10G2	100				87.0						
W4304	Sporadic	IbA10G2	100										
W4591	Sporadic	1bA10G2	100										
W4583	Sporadic	1bA10G2	100										
W1400	Sporadic	1bA10G2	100				85.2						
W 7387	Sporadic	1bA10G2	100										
OZ 21	Chalmers *et al.* (2005)	IbA10G2	100				87.8						
OZ 24	Chalmers *et al.* (2005)	IbA10G2	Neg										
OZ 9	Chalmers *et al.* (2005)	IbA9G2T1/IbA9G3	100				88.8						
W4540	Sporadic	IbA8G4	100										
W11535	Outbreak 1	IdA15G1	100										
W11457	Brighton OB	IeA11G3T3	100	96.3	95.6								
OZ 8	Chalmers *et al.* (2005)	IeA12G3T3	100										
W11212	Outbreak 1	IeA11G3T3	100										
W11231	Outbreak 1	IeA11G3T3	100										
W11639	Outbreak 1	IeA11G3T3	100										
W11646	Outbreak 1	IeA11G3T3	100										
W10638	Historical	IeA11G3T3	100										
W11306	Outbreak 1	IeA11G3T3	100										
W2588	Sporadic	Equivocal A37/38	100				93.9						
W8619	Historical	IaA18	100			97.0							

W11843	Outbreak 2	Seq fail	100										
W11932	Outbreak 2	Seq fail	100										
W11933	Outbreak 2	Seq fail	100										
W11963	Outbreak 2	Seq fail	100										
W12174	Outbreak 2	Seq fail	100										
W3590	Sporadic	Seq fail	100										
W3676	Sporadic	Seq fail	100										
W4220	Sporadic	Seq fail	100										
W12174	Outbreak 2	Seq fail	100				98.5				96.7		96.3
C. parvum													
W6840	National collection	IIaA15G2R1	100										
W8966	National collection	IIaA15G2R1	Neg										
W7649	National collection	IIaA17G1R1	100										
W6826	National collection	IIaA17G1R1	100										
OZ25	Chalmers *et al.* (2005)	IIaA19G3RI IIaA20G3RI	100										
W7651	National collection	IIaA20G3R1	100	94.9									
W3957	Sporadic	IgA24	100				94.7			95.4			
W7215	National collection	IIcA5G3	100										
W7265	National collection	IIcA5G3	100				97.2	97.6	95.4				
W3619	National collection	IIdA25G2	100				96.3	96.8	97.6				
W7597	National collection	IIeA7G1	100				98.2	98.8			95.0	96.1	Neg

*Numbers represent order on gel; L = left-hand gel; R = right-hand gel. Similarity values were calculated using the first run as a reference.

APPENDIX 3. UPGMA cluster analysis of GP60 SSCP panel isolates based on Pearson similarity values.

A. Evaluation panel



B. Library panel isolates