Journal of Infection 90 (2025) 106427

Contents lists available at ScienceDirect

Journal of Infection

journal homepage: www.elsevier.com/locate/jinf

Protozoa and Protozoal Diseases

Metabarcoding captures genetic diversity and links cases in outbreaks of cryptosporidiosis in New Zealand

Paul Ogbuigwe [1], Patrick J. Biggs, Juan Carlos Garcia-Ramirez, Matthew A. Knox, Anthony Pita, Niluka Velathanthiri, Nigel P. French, David T.S. Hayman *

Massey University, Palmerston North, Manawatu-Wanganui, New Zealand

ARTICLE INFO

Article history:
Accepted 21 January 2025
Available online 29 January 2025

Keywords:
Cryptosporidiosis
Metabarcoding
Next generation sequencing
Molecular Epidemiology

SUMMARY

Cryptosporidiosis is a disease caused by the parasite *Cryptosporidium*. Globally, it is a leading cause of diarrhoea and a notifiable disease in New Zealand. Molecular analyses of *Cryptosporidium* isolated from notified cases do not always provide support for epidemiological links between individuals. We hypothesised this could be due to undetected diversity and the use of consensus Sanger sequence analyses. Here, we analysed 105 *Cryptosporidium* samples from outbreaks and sporadic cases occurring between 2010 and 2018 in New Zealand using both Next-Generation Sequencing (NGS) and Sanger sequencing of the glycoprotein 60 (*gp60*) locus. NGS metabarcoding at the *gp60* locus uncovered significant intra- and inter-sample genotypic diversity in outbreaks and identified subtypes shared by epidemiologically linked cases, along with rare subtypes, suggesting it may be a useful tool for epidemiological investigations.

Introduction

Cryptosporidiosis is a disease characterised by acute, watery diarrhoea affecting approximately 7.6% of the world's population.[1,2] The disease is caused by the protozoan parasite *Cryptosporidium*, which infects the epithelial cells of the gastrointestinal tract to cause the disease. *Cryptosporidium* has a broad host range, including humans, domestic animals and wildlife.[3–5] In healthy humans, the disease is usually self-limiting with an incubation period of 4 – 28 days, and acute infection lasting 6 – 7 days.[6] However, the disease can be fatal in immunocompromised humans, children and other young animals. Cryptosporidiosis was responsible for approximately 12.1% of diarrhoea-associated deaths globally in children under 5 years old.[7]

There are currently 44 recognised species of *Cryptosporidium*, but the majority of human infections are caused by *C. hominis* and *C. parvum*.[8] However, about 20 other species have been identified in human infections. *Cryptosporidium hominis* causes the majority of anthroponotic cases in humans, although it has also been found in a variety of animal hosts, including equine and non-human primate species.[9–11] Subtype designations such as IfA12G1R5 code for a combination of allele variants and frequencies of repeat regions in an approximately 700 bp long locus of the *gp60* gene, which codes for a 60 kDa surface glycoprotein. Over 10 subtype families have been identified in *C. hominis*, with the virulent subtype IbA10G2 being the variant most commonly found in infected individuals across all socioeconomic settings.[12,13] *C. parvum* has a wider host range, including humans, companion animals, livestock and wildlife.[1] More than 26 subtype families of *C. parvum* have been identified so far in all hosts.[13–15]

In New Zealand, cryptosporidiosis has been a notifiable disease since 1996.[16,17] Between 2010 and 2018, the period of this study, the rate of cryptosporidiosis per 100,000 population in New Zealand has averaged 20.9/100,000, with a high of 33/100,000 in 2018 and a low of 13/100,000 in 2014 (Fig. S1.). Initially, cryptosporidiosis was diagnosed microscopically, but this proved inefficient for characterisation due to the morphological similarity of the environmental stage of the parasite, the oocyst, between species. The advent of molecular typing technologies such as PCR allowed for better characterisation of the parasite and understanding of the epidemiology and population genetics of the parasite.[8,13] Different PCR primers have different specificities. *Cryptosporidium* species not detected by the gp60 gene PCR include those that do not possess the gp60 gene or have significant genetic variations in the gene, such as *C. felis, C. canis, C. viatorum, C. muris* and *C. andersoni*.[18] However, six species of *Cryptosporidium* have been identified infecting humans in New Zealand (*C. parvum*, *C. hominis*, *C. cuniculus*, *C. erinacei*, *C. meleagridis*,

* Corresponding author.
E-mail address: d.t.s.hayman@massey.ac.nz (D.T.S. Hayman).
[1] **Permanent address:** Te Whatu Ora Health New Zealand Waikato, Hamilton, Waikato, New Zealand.

https://doi.org/10.1016/j.jinf.2025.106427

and *C. tyzzeri*) using the *gp60* gene.[12,19,20] *Cryptosporidium parvum* is the most common species (59%) identified, followed by *C. hominis*.[20] The predominance of *C. parvum* over *C. hominis* increased with Covid-19 restrictions in the country, which reduced *C. hominis* transmission,[21] but cases of *C. hominis* have recently resurged (per. obs.).

The genetic diversity and population structure of *Cryptosporidium* in humans[13] can help to understand the epidemiology of the disease, informing public health, and enabling the attribution of specific subtypes to outbreaks. Genotyping helps to understand transmission, improve detection resolution, investigation and interpretation of outbreaks.[22] Previous studies investigating cases of cryptosporidiosis in New Zealand have utilised Sanger sequencing.[12,16,17] This presents limitations because such technologies do not capture the full genetic diversity of *Cryptosporidium* within a host. Intra-sample diversity may be observed in Sanger sequencing chromatograms, but dominant consensus sequences are typically reported, and the separate sequences are difficult to characterise.[23] Next-generation sequencing (NGS) technologies sequence millions of reads per sample compared to just one consensus read achieved with Sanger sequencing, and therefore can capture significantly more diversity within each sample.[24] Grinberg et al.[25] previously applied NGS to two *C. parvum* human isolates and showed extensive intra-host diversity in samples for which Sanger sequencing had identified a single subtype only.

Application of these NGS techniques in population-level studies may help identify emerging subtypes of *Cryptosporidium*. For instance, subtype IfA12G1R5 was previously thought to be a rare subtype of *C. hominis* and is now dominant in some countries. In the USA, IbA10G2 was the dominant *C. hominis* subtype reported from outbreaks,[13] but by 2007 IaA28R4 was more predominant and from 2013 IfA12G1R5 became the dominant *C. hominis* subtype in sporadic and outbreak cases.[26,27] In Australia, IfA12G1R5 became the most dominant subtype in cases of cryptosporidiosis in Western Australia during 2017.[28]

However, the uncritical use of NGS techniques has potential drawbacks. Repeat regions have high replication slippage rates, making it difficult to discern biological diversity from error.[23] In previous work, we used NGS metabarcoding to determine the rate of replication slippage in mock communities of synthesised *Cryptosporidium* DNA in clonal plasmid vectors.[29] This work indicated that slippage rates increase with the length of the repeat region. The *gp60* gene has varying lengths of repeat regions used to classify *Cryptosporidium* subtypes, but slippage contributes to error rates of up to 20%.[29,30]

In this study, 105 human samples from historic cases of cryptosporidiosis occurring between 2010 and 2018 in New Zealand were analysed. Eighty-six samples were from outbreaks and 19 from sporadic cases. The aim was to utilise NGS to gain a better understanding of the genetic diversity within and between samples to help identify *Cryptosporidium* subtypes shared between samples from epidemiologically linked cases and the presence of emerging subtypes. This will provide a better understanding of the epidemiology of cryptosporidiosis in New Zealand.

Methods

Sampling

The Protozoa Research Unit (PRU) at the Hopkirk Institute, Palmerston North, New Zealand, receives anonymous human faecal samples diagnosed as positive by accredited diagnostic laboratories from routine surveillance of sporadic cases and outbreaks of cryptosporidiosis in New Zealand. All samples included in this study, both from outbreaks and routine surveillance, were from patients diagnosed with cryptosporidiosis. A list of the samples from the outbreaks that occurred in New Zealand between 2010 and 2018 that have been reported to the unit can be found in Table S1.

DNA purification and Sanger sequencing

Genomic DNA was extracted from faecal samples that had been stored at 4 °C using a Quick-DNA Faecal/Soil Microbe Kit (Zymo Research, Irvine, California, United States). The procedure required the use of a bead-beater (Tissue Lyser II, Qiagen) at 30 Hz for 5 min to disrupt the cysts. The purified DNA was stored at −20 °C prior to further processing. Nested PCR amplification of an approximately 700 bp fragment of the glycoprotein 60 (*gp60*) gene as described previously,[12] followed by sequencing of the amplification products using Big Dye Terminator version 3.1 reagents and an ABI 3730XL automated DNA sequencer (Applied Biosystems, Foster City, California, USA) was used to characterise each sample at the Massey Genome Service (Massey University, Palmerston North, New Zealand).

Next-Generation Sequencing

An approximately 400 bp fragment of the glycoprotein 60 (*gp60*) gene was amplified by nested PCR using a previously established PCR programme and set of primers.[31] The external primers were modified to contain MiSeq™ adapter sequences on the 5′ end according to standard protocols (Illumina Inc., 2013). Agarose gel electrophoresis was used to verify the presence of fragments of the correct size from all the PCR reactions. A blank containing deionised H_2O was used as a negative control, and the positive control was DNA from a sample that had been verified by PCR and Sanger sequencing as containing *Cryptosporidium* DNA.

The PCR products for all 105 samples were cleaned according to Illumina recommended protocols.[32] The DNA concentration of each sample was measured using NanoDrop (Thermo Fisher Scientific, Waltham, Massachusetts, United States). The products were cleaned-up and diluted to a concentration of 5 ng/μl according to the Illumina protocol, then delivered to the Massey Genome Service (Massey University, Palmerston North, New Zealand) for library preparation and amplicon-based sequencing. Sequencing was carried out on an Illumina MiSeq™ using 500-cycle V2 chemistry, according to the manufacturer's recommendations, producing 2 × 250 base paired-end reads. Due to the potential uneven representation of bases at each cycle with amplicon sequencing, an Illumina PhiX control library was loaded onto the Illumina MiSeq™ run at 20% volume, to even out the base composition and prevent biases in the initial few cycles that otherwise would result in base-calling errors.

gp60 database and nomenclature

The samples received at PRU are analysed through PCR and Sanger sequencing at the *gp60* locus using an in-house reference database including more than 139 unique *gp60* sequences from *C. hominis* and *C. parvum*; most previously submitted to GenBank[12] with accession numbers KY123918–KY124121 and MT265681-MT265802,[20] and available *gp60* sequences from GenBank using Geneious v.10.2.6.[33]

Within each subtype family (e.g., IIa), subtypes are classified by the number of trinucleotide repeats (TCA or TCG) coding for the amino acid serine and whether they have one or two copies (R) of an ACATCA sequence immediately after the trinucleotide repeats, i.e., subtype IIaA19G4R1, indicates that the subtype belongs to allele family IIa, there are 19 copies of the TCA repeat (A19), and four copies of the TCG repeat (G4) followed by one copy of ACATCA (R1).[13] We include these repeats when found. Ia and If can have different copies of a 15-bp repetitive sequence (AAGACGGTGGTAAGG or AAGAAGGCAAAGAAG, respectively), but these may have variations

and in our analyses of If sequences require multiple mismatches, so we do not include them. We do not differentiate sequences further based on SNPs in nonrepeat regions.[34]

Sequence processing and analysis

The Illumina reads for the 105 samples involved in this study were analysed inside the Quantitative Insights Into Microbial Ecology 2 (QIIME 2) environment.[35] The DADA2 methodology[36] was used to filter and trim the forward and reverse sequence reads, dereplicate them, calculate and plot error rates, merge paired reads and construct a sequence table, and remove chimeras. After sequence processing, there were 1341 unique sequences. Species and subtype classifications were carried out manually for all reads used below. Shannon and Simpson's diversity indices were used to assess intra-sample alpha diversity. Analysis of the processed sequences was conducted using the phyloseq R library.[37] We then simply looked at the numbers and percentages of sequences within case samples and among the cases.

Results

All 105 samples were previously analysed by PCR at the *gp60* locus and found to be positive for the presence of *Cryptosporidium*, and all negative controls were negative and positive controls positive. There were 89 samples for which the subtype according to Sanger sequencing was available. After NGS and Sanger sequencing analysis of the same locus (*gp60*), both Sanger sequencing and NGS identified 5 *Cryptosporidium* species: *C. hominis*, *C. parvum*, *C. cuniculus*, *C. tyzzeri* and *C. erinacei* (Fig. 1).

The most common dominant species identified in the samples by Sanger sequencing and NGS were *C. hominis* (62/89, 70% and 78/105, 74% respectively), followed by *C. parvum* (20/89, 22% and 21/105, 20% respectively) (Table S2). Seventy-four (83.1%) samples had the same dominant subtype family according to both NGS and Sanger sequencing, while 14 samples (15.7%) had different dominant subtype families, dependent on the sequencing method used (Table S2), including four (4.5%) species differences (Table S2; Fig. 1). The Sanger data for most of the outbreaks showed one dominant subtype could be used to link the cases epidemiologically, yet there were outbreaks in which multiple dominant subtypes were present according to the Sanger and NGS data (Table S2, Fig. S2).

NGS produced a total of 6314,333 sequences, of which 5316,709 (84%) were 1% or higher of the different sequences (Table S3). The most abundant sequence was a *C. hominis* subtype IbA10G2 which represented 29% (1801,158) of the total sequences (Table S3). Through NGS there is evidence of intra-sample diversity in most faecal samples analysed (Fig. 2). This diversity varies by outbreak and individual. The within-individual diversity and within-outbreak diversity can be seen in the raw data, which includes low numbers of sequences. For example, the 7 cases in the Auckland 2010 outbreak and 22 cases in the Hawke's Bay 2013 outbreak have similar within-sample Shannon and Simpson's diversity among the cases from the outbreaks, but with the 7 cases in Hawke's Bay 2013 having overall higher within sample diversity indices (Figs. S3, S4). There is also substantial diversity within subtype families, particularly Ib, Ig and IIa. The most common subtype of *C. hominis* found in cases was IbA10G2 (78.1%), followed by IgA17 (64.8%), IgA16 (57.1%), IgA20 (54.3%) and IfA12G1 (17.1%) (Table S4). IIaA18G3R1 (48.6%) was the most common *C. parvum* subtype among cases, followed by IIaA17G3R1 (33.3%), IIaA19G4R1 (18.1%) and IIdA17G1 (12.4%).

However, this diversity might have included PCR and sequencing error (e.g. single nucleotide variations) and slippage (e.g., trinucleotide repeats)[29] (Figures S5-S7). Yet the diversity can be seen at the species level (Fig. 2A) and when subtypes are aggregated to remove diversity caused by sequence differences that are not serine

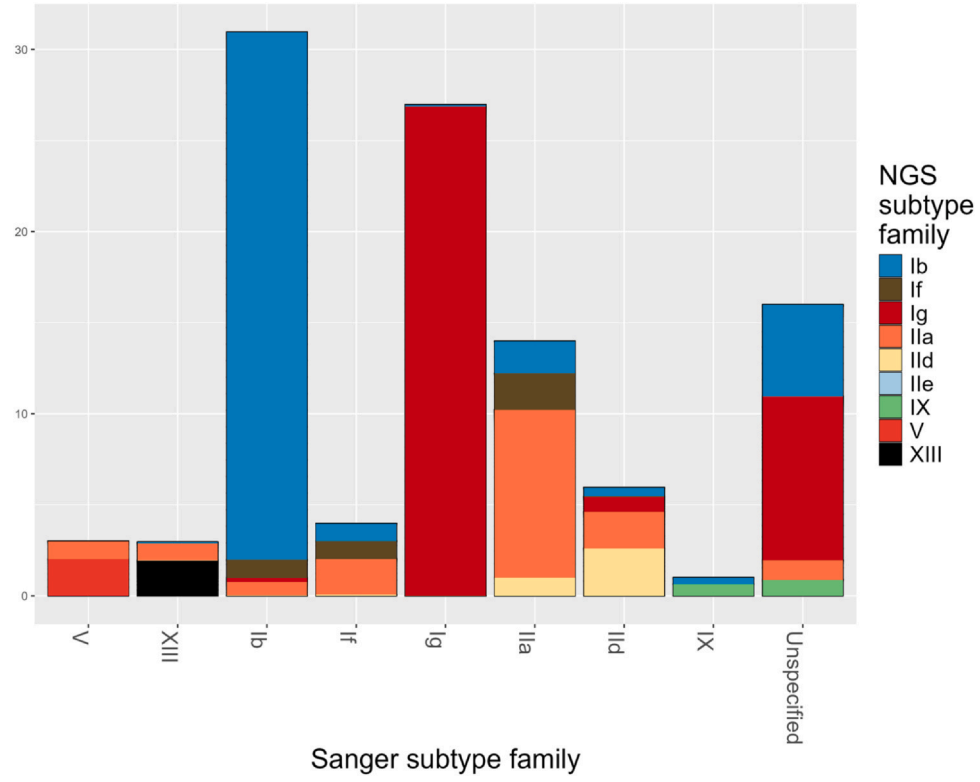

Fig. 1. The correspondence between Sanger sequencing subtype family identified and the most abundant sequences from NGS from the same samples. The subtype family identified by Sanger sequencing for all cases in this study is on the x-axis and on the y-axis the number of subtype family sequences from the top 300 identified among all samples using NGS. *Cryptosporidium hominis* is subtype family I, *C. parvum* II, *C. cuniculus* V, *C. erinacei* XIII, and *C. tyzzeri* IX. Further details are in Fig. S2.

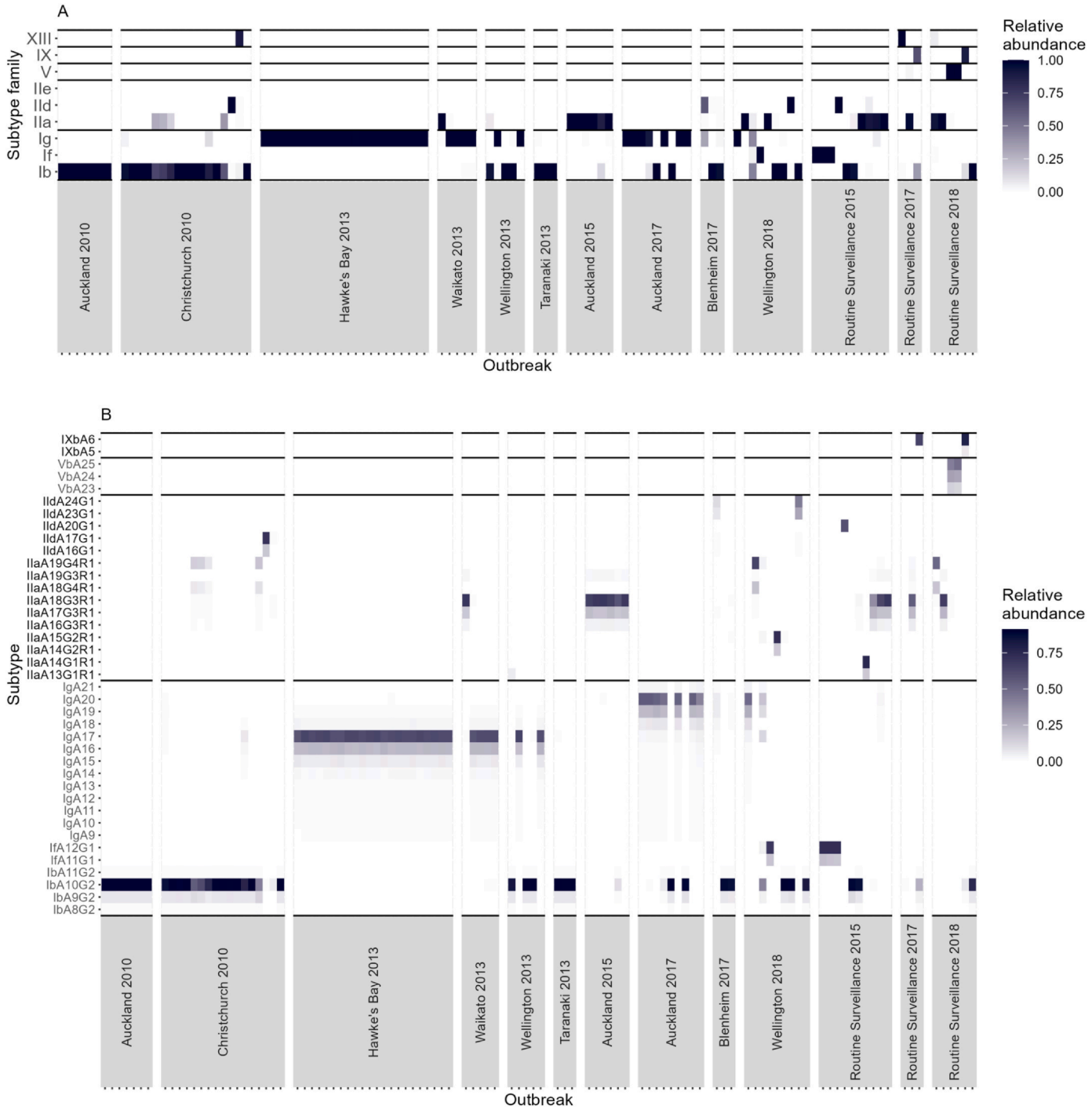

Fig. 2. The relative abundance of the most common sequences present in samples from all studied cryptosporidiosis outbreaks. The multiple subtypes present in each sample are displayed on the y-axis and sequences with differences (e.g. non-subtype related single nucleotide polymorphisms [SNPs]) are aggregated by: A) subtype family for the top 300 sequences and, B) subtype for the top 50 sequences. Subtypes for the top 300 sequences are in Fig. S5 and disaggregated data for the top 50 are in Figs. S6 and S7 to show the subtype and SNP diversity in greater detail. *Cryptosporidium hominis* is subtype family I, *C. parvum* II, *C. cuniculus* V, *C. erinacei* XIII, and *C. tyzzeri* IX. Subtypes are identified by family (e.g., IIa), the number of trinucleotide repeats (i.e., TCA or TCG) coding for serine and the number (R) of ACATCA sequence copies (e.g., R1 or R2) immediately after the trinucleotide repeats.

repeats (Fig. 2B) using the most common and abundant sequences, whether the top 50 (~1.5% of 3545; Fig. 2B) sequences for the subtype or top 300 (~8.5% of 3545; Fig. 2A) sequences for subtype.

Some outbreaks are characterised by single subtype families with identical subtype profiles, such as the Hawke's Bay 2013 outbreak (Fig. 2). Others have greater subtype variation, and this may add uncertainty to epidemiological investigations. In the outbreak that occurred in Christchurch in 2010, of the 17 samples investigated, 14 were classified by Sanger sequencing (Table S2). Of those 14, 11 shared *C. hominis* subtype family Ib, and one each of *C. parvum* IIa, *C. parvum* IId, and *C. erinacei*. One sample was classified as IIa by Sanger sequencing but showed IbA10G2 as an abundant sequence read by NGS, along with IIaA19G4R1, suggesting coinfection with at least two species. Two samples classified as IId and *C. erinacei* respectively both included this subtype and species according to NGS (Figure S2). However, analysis shows that *C. hominis* IbA10G2 was present in all

the samples from that outbreak (Figures S5) and was dominant in most cases (Figs. S2 and 2), indicating a shared subtype for these epidemiologically linked cases not evidence from Sanger sequencing. In addition, several of the samples from that outbreak shared the same variant of *C. parvum* IId.

The 9 samples from the Wellington 2018 outbreak were classified by Sanger sequence analysis (Table S2): 3 as C. *hominis* Ib, 1 *C. hominis* Ig, 1 *C. hominis* If, 2 *C. parvum* IIa, and 2 *C. parvum* IId, with no common subtype shared among all the samples. The NGS classification showed a more complex pattern revealing greater diversity, but with *C. hominis* Ig present in a higher proportion of samples and *C. hominis* IbA10G2 detected in several samples at high abundance (Fig. 2) and all samples at lower abundance (Figure S6). The raw NGS abundance data for the rest of the outbreaks is shown in Figures S6 and S7. All species and subtypes identified with accession numbers or the sequences for novel sequences are provided in Table S5. The code and abundance data for reproducibility are provided at https://github.com/dtsh2/cryptosporidium_metabarcoding.

Discussion

We conducted a comparative analysis of samples from past sporadic cases and outbreaks of cryptosporidiosis in New Zealand to connect epidemiologically linked cases and identify emerging or previously unidentified variants of *Cryptosporidium* that could have public health significance. NGS data identified the presence of the major species (*C. hominis*, *C. parvum*, *C. cuniculus*, *C. tyzzeri* and *C. erinacei*) that have previously been found in New Zealand. The two most abundant subtypes identified in this study by NGS were *C. hominis* IbA10G2 (78.1%) and IgA17 (64.8%) suggesting that anthroponotic transmission plays a considerable role in outbreaks and the spread of *Cryptosporidium* in New Zealand[29] despite its substantial livestock industry. Similar to previous studies in New Zealand,[12,20,30] IIaA18G3R1 was identified as the dominant subtype of *C. parvum* (48.6%).

The ability of NGS to sequence multiple reads in each sample allows to uncover the hidden genetic diversity contained within hosts which can serve as a method for the early identification of emerging variants of a pathogen before they become dominant. *Cryptosporidium hominis* IfA12G1 (15.2%) and IgA20 (54.3%) were identified as emerging subtypes based on the number of samples present. This is of importance because previous studies have documented the rise of IfA12G1R5 in the USA and Australia,[27,28] and the same could be occurring in New Zealand. The mechanisms that lead to such replacement are not clear.[38] Similarly, IgA20 was identified as the subtype of *Cryptosporidium* responsible for an outbreak of cryptosporidiosis in the United States in 2015.[13]

NGS data showed that 14 of 89 samples had different dominant subtypes as those classified by Sanger sequence analysis. In all 14 samples, the subtype identified through the Sanger method was still present in the NGS approach, but at lower abundance. Further analysis showed evidence of high intra-species and subtype diversity observed from the NGS results. These results suggest that *Cryptosporidium* infections in humans frequently genetically diverse[39]; either because of ingestion of genetically diverse oocysts, genetically diverse sporozoites within oocysts (potential results of sexual reproduction), or both. Advances in in vitro manipulation of *Cryptosporidium* could provide a clearer picture of the source of this diversity in individuals in the future.[40,41] However, our results demonstrate *Cryptosporidium* infections in humans are frequently genetically diverse, and NGS is superior to consensus sequencing at capturing this diversity.

This diversity likely depends on both the diversity of the population of *Cryptosporidium* in the exposure source and the impacts of any subsequent human to human transmission.[30,38] For example, some sources in which the oocysts from an outbreak come from a single infected individual may be more likely to have low within and between sample diversity, whereas others may be more likely to lead to higher diversity in the exposure and within and between samples from cases. Often, however, the source is unclear. Among 11 New Zealand recently analysed outbreaks, including those here, only 6 had known sources (swimming pools,[3] raw milk[2] and a daycare facility[1]) and among those the Blenheim 2017 outbreak was linked to a swimming pool complex, yet included *C. hominis* subtype IbA10G2 and *C. parvum* subtype IIdA24G1 (Fig. 2;[30]).

The disagreement between Sanger and NGS results regarding the most abundant subtype families is more common in our routine surveillance samples than in outbreak samples. This is possibly due to outbreaks being more frequently caused by a single dominant strain or a closely related group of strains originating from a common source, leading to a more genetically homogenous pathogen population during outbreaks. During outbreaks, transmission events can impose genetic bottlenecks, reducing diversity, with only a subset of the pathogen population successfully infecting new hosts, further narrowing genetic variation. Together these factors together result in higher genetic diversity during routine surveillance compared to the reduced diversity seen in outbreak-specific situations, so leading to greater concordance between methods.

One of the problems hampering a full understanding of the epidemiology of cryptosporidiosis in New Zealand has been the inability to identify shared subtypes in epidemiologically linked cases. The outbreaks that occurred in Christchurch in 2010 and Wellington in 2018 had multiple dominant subtypes according to Sanger sequence analysis. By using NGS on the outbreak samples that occurred in Christchurch in 2010, *C. hominis* IbA10G2 was identified as a subtype present in all the samples. The samples from the outbreak in Wellington in 2018 were significantly more diverse, with 5 different dominant subtypes being identified within the 9 samples involved in that outbreak but once again *C. hominis* IbA10G2 was a subtype common to all the samples, even if at low abundance (Figure S6, S7). Further work applying NGS analysis to outbreaks will help our understanding of the epidemiology and transmission patterns of the parasite,[42] improving on *gp60* gene analyses, though there are still significant barriers to overcome before routine whole genome sequencing will be possible.[43]

We did not sequence negative templates and/or reagent controls. Sequencing negative controls and reagents are essential for understanding the level of cross- and environmental (including kit) contamination in a sequence run, including index hopping. However, index hopping is usually between 0.1 to 1% on the Illumina MiSeq platform.[44–46] The NGS sequencing in this study produced millions of reads and the low quality and abundance reads were excluded from the study. DADA2 reports fewer false positives than other methods.[36] In our DADA2 analysis only the top 50 and 300 sequences (< 10%) with high abundance across all samples were used to make inferences (Figs. 1 and 2). This makes it very unlikely to report false positive sequences or index hopping. Each outbreak showed different patterns and abundance of subtypes present, which suggest negligible levels of cross-contamination. PCR slippages could account for some genotype diversity observed in this study. However, as previously described by Zahedi et al., (2017)[31] in a similar study and our own work,[29,30] PCR slippage is unlikely to account for all the diversity observed in this study. For example, the outbreak in Christchurch 2010 has four samples showing the co-occurrence of *C. parvum* IIa and IId subtypes, which is not the result of slippage.[23]

Conclusions

Our study highlights the benefits of NGS analysis in the classification and characterisation of *Cryptosporidium* samples; capturing the broad genetic diversity present in individuals infected with the

parasite and allowing the identification of shared subtypes in epidemiologically linked cases. Our work showed the ability of this technique to identify emerging variants of *Cryptosporidium* and advances our understanding of the epidemiology of this parasite in New Zealand to help inform future public health decisions.

Ethical approval

Not applicable. Our protozoa genotype surveillance is out of scope and does not require Health and Disability Ethics Committees (HDEC) approval.

Funding

This work was supported by funds from MicroAquaTech, Massey University, Royal Society Te Aparangi grant RDF-MAU170 (DTSH), New Zealand Ministry of Health contract number 355766-02 and a Percival Carmine Chair in Epidemiology and Public Health (DTSH).

Author contributions

DH conceived and designed the study with advice and funding assistance from NF. PO, AP and NV conducted data gathering. PB, JG, PO, MK and DH performed analyses. PO wrote the original draft article and all coauthors contributed to editing.

Data availability

Sequence accession numbers and new sequences are in Appendix A Supporting information Table S5. Code is available on GitHub repository: https://github.com/dtsh2/cryptosporidium_metabarcoding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Ahmed Fayaz, Lynn Rogers and the technicians at mEpiLab for their assistance with this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jinf.2025.106427.

References

1. Cacciò SM, Widmer G. *Cryptosporidium: parasite and disease.* Springer Science & Business Media; 2013.
2. Dong S, Yang Y, Wang Y, Yang D, Yang Y, Shi Y, et al. *Prevalence of Cryptosporidium infection in the global population: a systematic review and meta-analysis. Acta Parasitol* 2020;**65**:882–9.
3. Pumipuntu N, Piratae S. *Cryptosporidiosis: a zoonotic disease concern. Vet World* 2018;**11**(5):681.
4. Garcia-R JC, Hayman DT. *Origin of a major infectious disease in vertebrates: the timing of Cryptosporidium evolution and its hosts. Parasitology* 2016;**143**(13):1683–90.
5. Garcia-R JC, Pita AB, Velathanthiri N, Pas A, Hayman DT. *Mammal-related Cryptosporidium infections in endemic reptiles of New Zealand. Parasitol Res* 2023;**122**(5):1239–44.
6. Hunter PR, Hughes S, Woodhouse S, Nicholas R, Syed Q, Chalmers RM, et al. *Health sequelae of human cryptosporidiosis in immunocompetent patients. Clin Infect Dis* 2004;**39**(4):504–10.
7. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. *Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Infect Dis* 2017;**17**(9):909–48.
8. Feng Y, Ryan UM, Xiao L. *Genetic diversity and population structure of Cryptosporidium. Trends Parasitol* 2018;**34**(11):997–1011. https://doi.org/10.1016/j.pt.2018.07.009
9. Inácio SV, Widmer G, de Brito RLL, Zucatto AS, de Aquino MCC, Oliveira BCM, et al. *First description of Cryptosporidium hominis GP60 genotype IkA20G1 and Cryptosporidium parvum GP60 genotypes IIaA18G3R1 and IIaA15G2R1 in foals in Brazil. Vet Parasitol* 2017;**233**:48–51.
10. Parsons MB, Travis D, Lonsdorf EV, Lipende I, Roellig DMA, Kamenya S, et al. *Epidemiology and molecular characterization of Cryptosporidium spp. in humans, wild primates, and domesticated animals in the Greater Gombe Ecosystem, Tanzania. PLoS Negl Trop Dis* 2015;**9**(2):e0003529.
11. Widmer G, Köster PC, Carmena D. *Cryptosporidium hominis infections in non-human animal species: revisiting the concept of host specificity. Int J Parasitol* 2020;**50**(4):253–62.
12. Garcia–R JC, French N, Pita A, Velathanthiri N, Shrestha R, Hayman DT. *Local and global genetic diversity of protozoan parasites: spatial distribution of Cryptosporidium and Giardia genotypes. PLoS Negl Trop Dis* 2017;**11**(7):e0005736.
13. Xiao L, Feng Y. *Molecular epidemiologic tools for waterborne pathogens Cryptosporidium spp. and Giardia duodenalis. Food Waterborne Parasitol* 2017;**8**:14–32.
14. Bujila I, Troell K, Ögren J, Hansen A, Killander G, Agudelo L, et al. *Cryptosporidium species and subtypes identified in human domestic cases through the national microbiological surveillance programme in Sweden from 2018 to 2022. BMC Infect Dis* 2024;**24**(1):146. https://doi.org/10.1186/s12879-024-09049-x
15. Bujila I, Ohlson A, Hansen A, Agudelo L, Kühlmann-Berenzon S, Galanis I, et al. *Outbreak of the novel Cryptosporidium parvum IIγA11 linked to salad bars in Sweden, December 2023. Epidemiol Infect* 2024;**152**:e140. https://doi.org/10.1017/s0950268824001432
16. Learmonth J, Ionas G, Pita A, Cowie R. *Seasonal shift in Cryptosporidium parvum transmission cycles in New Zealand. J Eukaryot Microbiol* 2002(Supplement):34S–35SS.
17. Learmonth JJ, Ionas G, Ebbett KA, Kwan ES. *Genetic characterization and transmission cycles of Cryptosporidium species isolated from humans in New Zealand. Appl Environ Microbiol* 2004;**70**(7):3973–8.
18. Ryan U, Zahedi A, Feng Y, Xiao L. *An update on zoonotic Cryptosporidium species and genotypes in humans. Animals* 2021;**11**(11):1–22. https://doi.org/10.3390/ani11113307. 3307.
19. Garcia-R JC, Hayman DT. *Evolutionary processes in populations of Cryptosporidium inferred from gp60 sequence data. Parasitol Res* 2017;**116**(7):1855–61.
20. Garcia-R JC, Pita AB, Velathanthiri N, French NP, Hayman DT. *Species and genotypes causing human cryptosporidiosis in New Zealand. Parasitol Res* 2020;**119**:2317–26.
21. Knox MA, Garcia RJ, Ogbuigwe P, Pita A, Velathanthiri N, Hayman DTS. *Absence of Cryptosporidium hominis and dominance of zoonotic Cryptosporidium species in patients after Covid-19 restrictions in Auckland, New Zealand. Parasitology* 2021;**148**(11):1288–92. https://doi.org/10.1017/s0031182021000974
22. Morris A, Robinson G, Swain MT, Chalmers RM. *Direct sequencing of Cryptosporidium in stool samples for public health. Front Public Health* 2019;**7**:360.
23. Dettwiler I, Troell K, Robinson G, Chalmers RM, Basso W, Rentería-Solís ZM, et al. *TIDE analysis of Cryptosporidium infections by gp60 typing reveals obscured mixed infections. J Infect Dis* 2022;**225**(4):686–95.
24. Garcia-R JC, Ogbuigwe P, Pita AB, Velathanthiri N, Knox MA, Biggs PJ, et al. *First report of novel assemblages and mixed infections of Giardia duodenalis in human isolates from New Zealand. Acta Trop* 2021;**220**:105969.
25. Grinberg A, Biggs P, Dukkipati V, George T. *Extensive intra-host genetic diversity uncovered in Cryptosporidium parvum using Next Generation Sequencing. Infect Genet Evol* 2013;**15**:18–24.
26. Putignani L, Menichella D. *Global distribution, public health and clinical impact of the protozoan pathogen Cryptosporidium. Interdiscip Perspect Infect Dis* 2010;**2010**(1):753512.
27. Hlavsa MC, Roellig DM, Seabolt MH, Kahler AM, Murphy JL, McKitt TK, et al. *Using molecular characterization to support investigations of aquatic facility-associated outbreaks of Cryptosporidiosis – Alabama, Arizona, and Ohio, 2016. MMWR Morb Mortal Wkly Rep* 2017;**66**(19):493–7. https://doi.org/10.15585/mmwr.mm6619a2
28. Braima K, Zahedi A, Oskam C, Reid S, Pingault N, Xiao L, et al. *Retrospective analysis of Cryptosporidium species in Western Australian human populations (2015–2018), and emergence of the C. hominis IfA12G1R5 subtype. Infect Genet Evol* 2019;**73**:306–13.
29. Knox MA, Biggs PJ, Garcia-R JC, Hayman DT. *Quantifying replication slippage error in Cryptosporidium metabarcoding studies. J Infect Dis* 2024;**230**:e144–8.
30. Garcia-R JC, Hayman DT. *A review and analysis of cryptosporidiosis outbreaks in New Zealand. Parasitology* 2023;**150**(7):606–11.
31. Zahedi A, Gofton AW, Jian F, Paparini A, Oskam C, Ball A, et al. *Next Generation Sequencing uncovers within-host differences in the genetic diversity of Cryptosporidium gp60 subtypes. Int J Parasitol* 2017;**47**(10-11):601–7.
32. Illumina I. 16S Metagenomic sequencing library preparation. Preparing 16S ribosomal RNA gene amplicons for the illumina MiSeq system. 2013;1:28.
33. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics* 2012;**28**(12):1647–9.
34. Sulaiman IM, Hira PR, Zhou L, Al-Ali FM, Al-Shelahi FA, Shweiki HM, et al. *Unique Endemicity of cryptosporidiosis in children in Kuwait. J Clin Microbiol* 2005;**43**(6):2805–9. https://doi.org/10.1128/jcm.43.6.2805-2809.2005
35. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. *Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol* 2019;**37**(8):852–7.
36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. *DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods* 2016;**13**(7):581–3.

37. McMurdie PJ, Holmes S. *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS One* 2013;**8**(4):e61217.
38. Garcia-R JC, Cox MP, Hayman DT. *Comparative genetic diversity of Cryptosporidium species causing human infections. Parasitology* 2020;**147**(13):1532–7.
39. Thompson R, Ash A. *Molecular epidemiology of Giardia and Cryptosporidium infections. Infect Genet Evol* 2016;**40**:315–23.
40. Kemp A, Lofroth M, Barnes V, Pita A, Hayman DT, Avci E. *Isolation of Parasites by Micropipette Aspiration.* 2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). IEEE; 2022.
41. Ogbuigwe P, Roberts JM, Knox MA, Heiser A, Pita A, Haack NA, et al. *A novel, stain-free, natural auto-fluorescent signal, Sig M, identified from cytometric and transcriptomic analysis of infectivity of Cryptosporidium hominis and Cryptosporidium parvum. Front Cell Infect Microbiol* 2023;**13**:1178576.
42. Bellinzona G, Nardi T, Castelli M, Biffignandi GB, Adjou K, Betson M, et al. *Comparative genomics of Cryptosporidium parvum reveals the emergence of an outbreak-associated population in Europe and its spread to the United States. Genome Res* 2024;**34**(6):877–87.
43. Agyabeng-Dadzie F, Xiao R, Kissinger JC. *Cryptosporidium genomics—current understanding, advances, and applications. Curr Trop Med Rep* 2024;**11**(2):92–103.
44. England R, Harbison S. *A review of the method and validation of the MiSeq FGx™ Forensic Genomics Solution. Wiley Interdiscip Rev Forensic Sci* 2020;**2**(1):e1351.
45. Hornung BV, Zwittink RD, Kuijper EJ. *Issues and current standards of controls in microbiome research. FEMS Microbiol Ecol* 2019;**95**(5):fiz045.
46. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. *Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol* 2017;**35**(11):1077–86.