Uncovering the Genetic Diversity of *Giardia* Isolates from Outbreaks in New Zealand

Paul Ogbuigwe (p.ogbuigwe@massey.ac.nz)
Massey University https://orcid.org/0000-0002-0553-3207
Patrick J. Biggs
Massey University
Juan Carlos Garcia-Ramirez
Massey University
Matthew A. Knox
Massey University
Anthony Pita
Massey University
Niluka Velathanthiri
Massey University
Nigel P. French
Massey University
David T. S. Hayman
Massey University

Research Article

Keywords: Giardia Intestinalis, genetic diversity, epidemiology, New Zealand, DNA sequencing, metabarcoding

Posted Date: December 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1157774/v1

Abstract

Background

Giardia is one of the most common causes of diarrhoea in the world and is a notifiable disease in New Zealand. Recent advances in molecular techniques, such as PCR and Sanger sequencing, have greatly improved our understanding of the taxonomic classification and epidemiology of this parasite. However, there has been an inability to identify shared subtypes between samples from the same epidemiologically linked cases, due to samples showing multiple dominant subtypes within the same outbreak when characterised using Sanger sequencing.

Methods

Here, NGS was employed to uncover the genetic diversity within samples from sporadic and outbreak cases of giardiasis that occurred in New Zealand between 2010 and 2018.

Results

This strategy exposed the significant diversity of subtypes of *Giardia* present in each sample. The utilisation of NGS and metabarcoding at the glutamate dehydrogenase (*gdh*) locus enabled the identification of shared subtypes between samples from shared outbreaks, providing a better understanding of the epidemiology of outbreaks of giardiasis in New Zealand.

Conclusions

Next-generation sequencing technologies provides a superior tool, when compared to consensus sequencing technologies, for capturing the genetic diversity of Giardia within hosts. This study showed that infections in humans are frequently mixed, with multiple subtypes present in each host.

Key Findings

- Next-generation sequencing uncovers significant within-host diversity of Giardia parasite, previously undetected by Sanger sequencing, in patients from historical outbreaks of giardiasis in New Zealand.
- We use NGS to uncover genetic links between epidemiologically linked cases of giardiasis.
- Similar to historic data, assemblage B has been identified as the most common assemblage found in New Zealand. However, assemblages A and E are commonly found in samples from humans in New Zealand.

Introduction

Giardia is an enteric protozoan parasite with the distinction of being among the most common causes of diarrhoea in humans and farm animals worldwide (Cacciò and Sprong, 2011). *Giardia* infects the epithelial cells of the gastrointestinal tract causing self-limiting diarrhoea in all classes of vertebrates.

This parasite transmits via the faecal-oral route, and in humans particularly, contact with contaminated water sources is the dominant mode of infection and cause of outbreaks.

Approximately 280 million people are infected with this parasite every year, and the prevalence of infections in humans ranges between 0.4–7.5% in high-income countries, and 8–30% in low-/middle-income countries (Feng and Xiao, 2011; Einarsson, Ma'ayeh and Svärd, 2016). Symptoms range in severity, however, the disease can be fatal in immunocompromised individuals and ranks among the leading causes of death in children under the age of 5 (Luján and Svärd, 2011). This is why giardiasis, the disease for which *Giardia* is the causative agent, was recognised by the Word Health Organisation (WHO) in its neglected diseases initiative, highlighting the public health significance of this parasite (Savioli, Smith and Thompson, 2006). Since then, reporting of this parasite has improved in many countries. Exacerbating the burden of this disease is the lack of any effective vaccines against the pathogens.

At present eight species of *Giardia* are recognised, these are: *G. agilis* (associated with amphibians), *G. ardeae* (great blue herons), *G. cricetidarum* (hamsters), *G. intestinalis* (alternatively named *G. duodenalis* or *G. lamblia*, discussed below), *G. microti* (associated with voles and muskrats), *G. muris* (rodents), *G. peramelis* (southern brown bandicoots), and *G. psittaci* (found in budgerigars) (Ryan *et al.*, 2019). According to current understanding, the species responsible for all human infections is *G. intestinalis*, which is further divided into eight assemblages (or subtypes): A-H. These assemblages can be further classified into sub-assemblages. Assemblages A and B are thought to be responsible for most zoonotic infections and cause the majority of infections in humans. However, as molecular techniques have advanced, evidence of infection by other assemblages has been identified in humans (Feng and Xiao, 2011). Assemblages A and B have a wide host range including humans, livestock, domestic and wild animals. The remaining assemblages have narrow known host ranges. Assemblages C and D are associated with dogs and other canids, assemblage E with livestock, assemblages F with cats, assemblage G with rodents, and assemblage H with seals (Feng and Xiao, 2011). Assemblage B is responsible for the majority of human cases of giardiasis in low- and high-income settings, including in New Zealand where this assemblage was identified in 79% of cases between 2009 and 2015 (Garcia–R *et al.*, 2017a).

The clinical effects of giardiasis vary among individuals, ranging from asymptomatic carriage to severe malabsorption syndrome in some acute cases (Luján and Svärd, 2011). However, the mechanisms underlying the differences in phenotypes within these diseases are poorly understood. Previous studies suggest that differences in infectivity exist between assemblages. Experimental observations found that human volunteers inoculated with assemblage B were more likely to succumb to infection and develop symptoms than those inoculated with assemblage A (Cacciò, Lalle and Svärd, 2017). Nevertheless, studies looking at the correlation between symptoms and assemblages have produced contradictory results. The ability to link phenotypic features with assemblages would greatly increase our understanding of transmission patterns. Similarly, investigation of the genetic structure at the population level is essential for the proper inference of the transmission patterns of *Giardia* and its epidemiology.

Outbreaks of giardiasis occur frequently each year across the world. Previous reviews found that between 2011 and 2017 over 140 waterborne outbreaks occurred globally (Efstratiou, Ongerth and Karanis, 2017). Outbreaks might be initiated through waterborne transmission but have the potential to spread further through human-human interaction (Katz *et al.*, 2006). The true burden of this disease is potentially underestimated due to poor reporting in some countries. Giardiasis only became a notifiable disease in the USA, Europe and New Zealand between the late 90s and early 2000s (Adam et al., 2016; Plutzer et al., 2018; Snel, Baker, & Venugopal, 2009). Increased surveillance of *Giardia* and other enteric parasites will give a better idea of the true burden of giardiasis globally. Surveillance data in New Zealand found that *Giardia* was responsible for 7.4% of total outbreaks in the country during 2016, with person-to-person contact being the most common mode of transmission (Institute of Environmental Science and Research Ltd (ESR), 2018). However, there has been an inability to identify the same subtypes of *Giardia* in epidemiologically linked cases in New Zealand. A patient with an infection may carry multiple subtypes of the same infectious agent and the outcome of the competitive interactions between them has an effect on the clinical presentation of the disease, which, in turn, affects the efficacy of treatment (Thompson & Smith, 2011). For this reason, understanding the within-host genetic diversity of a pathogen is essential for effective disease management.

Questions remain as to whether epidemiologically linked cases in New Zealand were all part of the same events or if they represent within- and between-host diversity (Garcia–R *et al.*, 2017a). A possible reason for this could be a lack of resolution due to the standard detection methods used. Over the years the methods for the detection and classification of *Giardia* have progressed from microscopic analysis of physical characteristics to molecular tools such as PCR and Sanger sequencing of notable genes like *bg, gdh, tpi* and *SSU* rRNA genes. However, because Sanger sequencing combines the contribution of all DNA fragments present in the reaction mixture, even this may lack sufficient resolution where mixed assemblages are present. PCR amplification of the *gdh* gene will amplify sequences from any *Giardia* assemblage that is present in the extracted DNA, which can lead to a mixed signal in the resulting Sanger sequence or failure to detect rare assemblage types. These limitations affect disease surveillance and make it difficult to capture within-host diversity. In contrast to Sanger sequencing, next-generation sequencing (NGS) techniques like amplicon-based sequencing allow millions of fragments to be sequenced in a single run allowing the researcher to separate the signal originating from each target molecule, thus allowing the efficient isolation, detection and quantification of rare types. In recent years, researchers have applied NGS techniques to study the epidemiology of cryptosporidiosis and giardiasis, which has led to great advances in the understanding of these infectious diseases (Ortega-Pierres *et al.*, 2018).

In this study, NGS techniques are used to gain a better understanding of the genetic diversity of giardiasis outbreaks in New Zealand. Taking human faecal samples from three outbreaks of giardiasis that occurred between 2010 and 2017 in various regions across the country and some samples from routine surveillance, and utilising amplicon-based metabarcoding at the glutamate dehydrogenase (*gdh*) locus, the hypothesis that epidemiologically linked cases share subtypes undetectable with consensus sequencing technologies was tested. In addition, NGS was used to detect the degree of genetic diversity

(strictly, richness) present in samples from patients diagnosed with giardiasis. Comparing these results to the results of Sanger sequencing at the same locus it was possible to detect the presence of mixed infections and gained a better understanding of the assemblages of *Giardia* present in New Zealand. This study shows that amplicon-based sequencing provides better tools for painting a clearer picture of the role of protozoan genetic diversity in giardiasis outbreaks in New Zealand, which could lead to a better perception of protozoan outbreak epidemiology.

Materials And Methods

Sampling

The Protozoa Research Unit (PRU) at the Hopkirk Institute, Palmerston North, New Zealand, receives human faecal samples diagnosed as positive by accredited diagnostic laboratories from routine surveillance and outbreaks of giardiasis in New Zealand. All the samples included in this study, both from outbreaks and routine surveillance, were from patients diagnosed with giardiasis. A list of the samples from routine surveillance and outbreaks of giardiasis that occurred in New Zealand between 2010 and 2018 can be found in Table 1.

Table 1
List of samples from outbreaks and routine surveillance along with the regions in which they occurred. Outbreaks where 'Organism' is annotated with (*) highlight situations in which *Cryptosporidium* and *Giardia* were identified in the same sample. A full list of the samples used in this study can be found in Supplementary Table 1.

Year	Region	Organism	Sample origin	Number of Cases
2010	Hawke's Bay	*Giardia*	Giardiasis Outbreak	3
2014	Gisborne	*Giardia**	Giardiasis Outbreak	5
2015	Hawke's Bay	*Giardia*	Giardiasis Outbreak	5
2016	Christchurch	*Giardia*	Routine Surveillance	1
2017	Auckland	*Giardia**	Cryptosporidiosis Outbreak	1
2017	Palmerston North	*Giardia*	Routine Surveillance	1
2017	Otago	*Giardia*	Routine Surveillance	1
Total				17

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 at the glutamate dehydrogenase (*gdh*) locus,

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 Massey Genome Services (Palmerston North, New Zealand).

PCR

A partial fragment of the glutamate dehydrogenase (*gdh*) gene was amplified by nested PCR using a previously established PCR programme and set of primers (Read, Monis and Thompson, 2004). 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 (432 bp) from all the PCR reactions. A blank containing deionised H_2O was used as a negative control, and DNA from a sample that had already been verified by PCR and Sanger sequencing as containing *Giardia* DNA was used as a positive control.

Next-generation sequencing (NGS)

The PCR products for all 17 samples were cleaned according to Illumina recommended protocols (Illumina Inc., 2013). The DNA concentration in each sample was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, United States), the samples diluted to a 5 ng/µl concentration according to the Illumina protocol referenced above 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.

Construction of a gdh database

Through our collaboration with the New Zealand Ministry of Health the PRU receives anonymised faecal samples from patients diagnosed with giardiasis. The samples are analysed through PCR and Sanger sequencing at the *gdh* locus. The assembly of sequences and compilation of databases was done using Geneious v.10.2.6 (Kearse *et al.*, 2012). Using the *Giardia intestinalis* sequences from our in-house database a separate database was compiled consisting of 858 unique *gdh* sequences, most had previously been submitted to GenBank by Garcia–R et al. (2017) and can be found in GenBank with accession numbers MT265681 – MT265802. To capture the greatest possible extent of known diversity of *Giardia gdh* sequences, a dataset of all available *gdh* sequences for *G. intestinalis* from GenBank (Benson *et al.*, 2013) was extracted and imported into Geneious. The search strategy employed one search string (*Giardia*) and included the keywords glutamate dehydrogenase, and *gdh*. The sequences were trimmed to the length of the primers employed in this study and all sequences less than 393 bp were discarded. This left 337 unique sequences from GenBank. The 337 GenBank sequences were combined with the 858 sequences extracted from our in-house database, then duplicate sequences were extracted

to create a collection of 1109 unique sequences covering most of the assemblages of *G. intestinalis* that have been characterised at the *gdh* locus.

Sequence processing

The Illumina sequence reads for the 17 samples involved in this study were analysed inside the Quantitative Insights Into Microbial Ecology 2 (QIIME 2) environment (Bolyen *et al.*, 2019). The dada2 methodology (Callahan *et al.*, 2016) 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. Then our database of 1109 known unique sequences was used as a reference to assign taxonomy to the merged sequences. To remove the impact of index hopping or PCR error, from the processed and merged sequences only the top 1971 sequences, based on the expected sequence length, were imported from dada2 into the phyloseq R package (McMurdie and Holmes, 2013) for plotting, ranking of the most expressed sequences and creation of a heatmap. The resulting table of sequences was run against the reference database to exclude any sequences that did not match known sequences of *G. intestinalis* then put through phyloseq again for further analysis. Only the top 50 sequences present across all the samples were used for the creation of bar plots and heatmaps to reduce the possibility of sequencing errors being included in the analysis.

Results

Overview of sample data

Of the 17 historical faecal samples from cases of giardiasis that had occurred in New Zealand between 2010 and 2018 fragments of the *gdh* gene were successfully amplified for all of them using nested PCR. The samples for which the assemblage according to Sanger sequencing were known and the most dominant assemblage according to NGS are shown in Table 2. There were no disagreements in assigned dominant assemblage between the two sequencing methods. According to the NGS data, and focusing on the dominant assemblage in each sample, 11/17 samples were found to belong to sub-assemblage BIV, 1/17 to BIII, 2/17 to AII, 1/17 to AIII, 2/17 to E. Fig. 1 provides a comparison of the assemblage assigned by Sanger sequencing and the diversity captured by NGS. It shows that even in genetically diverse samples, like the one from the outbreak in Hawke's Bay in 2015, there are agreements between the Sanger sequence data and the NGS data. Analysis of the NGS data was conducted to probe the intra-sample diversity of these samples.

Table 2
Sample assemblages according to results of Sanger sequencing compared with most abundant assemblages according to NGS. “Unspecified” denotes samples for which the assemblage could not be determined. The number of reads generated by NGS from each sample after filtering, trimming and dereplication are shown for reference.

Sample No.	ID	Sanger	NGS	NGS Reads
1	1997	BIV	BIV	55235 reads in 15404 unique sequences
2	1998	BIV	BIV	113042 reads in 21061 unique sequences
3	1999	BIV	BIV	136387 reads in 25493 unique sequences
4	10015	AII	AII	118718 reads in 23734 unique sequences
5	10046	BIV	BIV	141257 reads in 24507 unique sequences
6	10047	BIV	BIV	95812 reads in 19814 unique sequences
7	10048	BIV	BIV	95836 reads in 28269 unique sequences
8	10049	BIV	BIV	106343 reads in 24744 unique sequences
9	10936	BIV	BIV	8184 reads in 3121 unique sequences
10	10937	BIV	BIV	144483 reads in 39869 unique sequences
11	10938	AII	AII	116354 reads in 33518 unique sequences
12	10939	BIV	BIV	121820 reads in 21446 unique sequences
13	10940	BIII	BIII	20678 reads in 6331 unique sequences
14	13273	BIV	BIV	103784 reads in 22832 unique sequences
15	14201	Unspecified	E	75624 reads in 15918 unique sequences
16	11359	Unspecified	AIII	112267 reads in 19785 unique sequences
23	13805	Unspecified	E	87460 reads in 11150 unique sequences

Metabarcoding analysis

The diversity of assemblages found in each sample after processing and analysis of the NGS reads are shown in Fig. 2. Similar to the results from the Sanger sequencing, the most abundant assemblage in most samples was assemblage B, specifically sub-assemblage BIV. This assemblage was present at some level in 16/17 samples, only one sample (13805_S23) did not have any variants of sub-assemblage BIV present in it. There was evidence of mixed infections in 13/17 samples. The majority of the genetic diversity within those 13 samples was due to the presence of multiple variants within assemblage B, for example, samples 10937_S10 and 10940_S13 showed evidence of multiple variants corresponding to assemblage B. The second most common assemblage present in this study was assemblage A, with 7/17 samples showing the presence of at least one variant of that assemblage.

Three samples from routine surveillance were included in this study (see Table 1) to compare the genetic diversity between samples from outbreaks and samples from sporadic cases. No significant differences were observed. Two samples (11359_S16 & 14201_S15) from routine surveillance represented the first report of sub-assemblage AIII and assemblage E in human samples from the South Island in New Zealand. These samples were analysed further in another study (Garcia-R *et al.*, 2021).

Identifying links between outbreak cases

The primary aim of this study was to utilise NGS to detect a genetic link between epidemiologically linked cases. To this end, an analysis of the outbreaks that occurred in Gisborne in 2014 and Hawke's Bay in 2015 was conducted. These outbreaks were selected based on the fact that although the samples within each outbreak were epidemiologically linked, according to Sanger sequence data the samples did not share the same dominant genotype.

Of the 5 samples from the outbreak that occurred in Gisborne in 2014, 4/5 were characterised as sub-assemblage BIV and 1/5 as AII according to Sanger and NGS data. Fig. 3A is a heatmap showing the genetic diversity, captured by NGS, within the samples involved in this outbreak. From this, it is evident that a single variant of sub-assemblage BIV is shared by all the samples in this outbreak. Also, a copy of sub-assemblage AIII is present in one of the samples from this outbreak (10049_S8).

Of the 5 samples received from the outbreak in Hawke's Bay in 2015, 3/5 were identified as sub-assemblage BIV, 1/5 as BIII, and 1/5 as AII according to Sanger and NGS data. From the heatmap shown in Fig. 3B it is evident that, despite the differences in dominant assemblages, sub-assemblage BIV is shared between all the samples from this outbreak. Sample 13273_S14 represented the only sample from an outbreak of cryptosporidiosis in Auckland. According to the NGS data this sample was also positive for *G. intestinalis* sub-assemblage BIV (Fig. 2). This represents an example of a mixed-species infection. The NGS abundance data for the rest of the outbreaks is available in Supplementary Figure 1.

Discussion

This investigation into the intra-sample diversity of *G. intestinalis* in patients from historical outbreaks of giardiasis in New Zealand compares the capabilities of NGS and Sanger sequencing technologies. The strength of Sanger sequencing lies in its ability to detect the dominant sequence within a sample. The results outlined here show that NGS is also capable of the same level of discernment with regards to the identification of dominant sequences, shown by the agreements between the data from Sanger sequence and NGS of samples from cases of giardiasis that occurred in New Zealand between 2010 and 2018. The aim of this study was to use NGS to capture the diversity within samples, and this is where the benefit of NGS over Sanger shows itself. NGS is capable of sequencing multiple reads in each sample, compared to the one consensus read per sample achieved with consensus sequencing technologies. Here, NGS was employed to uncover the genetic diversity present within cases of giardiasis in this country. As stated in the introduction, the clinical manifestation of giardiasis can differ between individuals. Further work needs to be done to ascertain if there is a link between (sub)assemblage and clinical presentation,

however recent advances in the *in vitro* culture of *Giardia* (Liu *et al.*, 2020) have the potential to address this. So, the ability to capture the genetic diversity within samples from cases of giardiasis and link them to the symptoms displayed by the patient could greatly advance our understanding of the disease mechanisms of this parasite.

The data presented here suggest that assemblage B is still the most common assemblage of *Giardia* in New Zealand, present in 16/17 samples. However, the ability to capture the diversity of assemblages within samples showed that, although they might not be dominant, assemblages A and E reported with increased frequency in New Zealand, as evidenced by their presence in 7/17 and 4/17 samples respectively. This is particularly significant since assemblage E was thought to be exclusively infectious to livestock. However, recent studies have shown that it is increasingly present in humans as well (Abdel-Moein and Saeed, 2016). The significance of this finding is discussed in a study published by our research group (Garcia-R *et al.*, 2021).

The subtyping in this study was carried out at only the *gdh* locus. This presents a potential limitation since other studies have shown that sequencing typing at different loci can result in assignation of multiple subtypes (Feng and Xiao, 2011; Brynildsrud *et al.*, 2018) and is why more recent studies utilise multi locus sequence typing (MLST) (Seabolt, Konstantinidis and Roellig, 2021). However, this study sought to compared data from NGS to samples that had previously been characterised by Sanger sequencing at the *gdh* locus. For this reason, metabarcoding at the same locus was considered appropriate for this study. No no-template controls or DNA extraction reagent blanks were included in the library prep for NGS. These are usually used as an indication of the level of lane-hopping or environmental contamination present in the sequenced samples. Here, the use of nested PCR resulted in the amplification of specifically the *Giardia* DNA at the specific locus analysed in this study. In addition, while index hopping might be present it is usually between 0.1 to 1% on the Illumina MiSeq platform (Sinha *et al.*, 2017; Hornung, Zwittink and Kuijper, 2019; England and Harbison, 2020), NGS sequencing in this study produced millions of reads and the low quality and abundance reads were removed from the study. Furthermore, only the top 50 sequences were used when analysing the diversity across all samples and within each outbreak. Also, each outbreak had a different pattern of amplicons, generally with different dominant subtypes, which suggests there was little cross-contamination present. Another limitation was the low number of samples from outbreaks of giardiasis. This was due to the fact that only a subset of samples from outbreaks that occurred in New Zealand between 2010 and 2018 are sent to our laboratory for molecular characterisation.

Another aim of this study was to use NGS to uncover genetic links between epidemiologically linked samples. It was hypothesised that epidemiologically linked cases share assemblages undetectable with consensus sequencing technologies. The outbreaks that occurred in Gisborne in 2014 and Hawke's Bay in 2015 provided a perfect case study for this. In those outbreaks there were multiple dominant assemblages present in the samples within each outbreak. By applying NGS and metabarcoding at the *gdh* locus it was shown that sub-assemblage BIV was shared between all samples from the Gisborne and

Hawke's Bay outbreaks, thereby verifying the hypothesis. This improves our understanding of the epidemiology of these outbreaks.

Conclusions

In conclusion, this study highlights the importance of utilising NGS technologies to uncover the genetic diversity of *Giardia* in humans to gain a better understanding of the risk factors associated with the disease. Out of 17 samples, 13 showed the presence of multiple variants of *Giardia*. This suggests that labelling a human sample using consensus sequencing technologies as belonging to one assemblage is insufficient and does not capture the true genetic diversity that can exist in one individual. In addition, these results suggest that *Giardia* frequently invades humans as part of a mixed infection. Further work needs to be carried out to ascertain the relative contribution of each assemblage to the disease phenotype. This will give us a better understanding of the disease mechanisms of the parasite and create a clearer epidemiological picture that will inform public health services in the development of better strategies to combat this persistent and prevalent parasite by allowing them to properly pinpoint all potential sources of infections and disrupt transmission pathways.

Declarations

Supplementary material

The supplementary material for this article can be found at [doi]

Ethical Approval and Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Sequence data is available in the Sequence Read Archive with the accession numbers PRJNA716067 and PRJNA785019.

Competing interests

Not applicable

Funding

This work was supported by funds from MicroAquaTech, Massey University, Royal Society Te Aparangi grant RDF-MAU170, and New Zealand Ministry of Health contract number 355766-02.

Authors' contributions

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

Acknowledgements

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

Authors' information

David Hayman is the senior author

Affiliations

Massey University, School of Veterinary Sciences, Palmerston North, New Zealand

All authors

References

Abdel-Moein, K. A. and Saeed, H. (2016) 'The zoonotic potential of *Giardia intestinalis* assemblage E in rural settings', *Parasitology Research*. Springer Verlag, 115(8), pp. 3197–3202. doi: 10.1007/s00436-016-5081-7.

Adam, E. A. *et al.* (2016) 'Giardiasis outbreaks in the United States, 1971-2011', *Epidemiology and Infection*. Cambridge University Press, pp. 2790–2801. doi: 10.1017/S0950268815003040.

Benson, D. A. *et al.* (2013) 'GenBank', *Nucleic Acids Research*. Nucleic Acids Res, 41(D1). doi: 10.1093/nar/gks1195.

Bolyen, E. *et al.* (2019) 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*. Nature Publishing Group, pp. 852–857. doi: 10.1038/s41587-019-0209-9.

Brynildsrud, O. *et al.* (2018) '*Giardia duodenalis* in primates: Classification and host specificity based on phylogenetic analysis of sequence data', *Zoonoses and Public Health*. John Wiley & Sons, Ltd, 65(6), pp. 637–647. doi: 10.1111/ZPH.12470.

Cacciò, S. M., Lalle, M. and Svärd, S. G. (2017) 'Host specificity in the *Giardia duodenalis* species complex', *Infection, Genetics and Evolution*. doi: 10.1016/j.meegid.2017.12.001.

Cacciò, S. M. and Sprong, H. (2011) 'Epidemiology of Giardiasis in Humans', in *Giardia*. Springer Vienna, pp. 17–28. doi: 10.1007/978-3-7091-0198-8_2.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Nature Publishing Group, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Efstratiou, A., Ongerth, J. E. and Karanis, P. (2017) 'Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011–2016', *Water Research*. Pergamon, 114, pp. 14–22. doi: 10.1016/J.WATRES.2017.01.036.

Einarsson, E., Ma'ayeh, S. and Svärd, S. G. (2016) 'An up-date on *Giardia* and giardiasis', *Current Opinion in Microbiology*. Elsevier Current Trends, 34, pp. 47–52. doi: 10.1016/J.MIB.2016.07.019.

England, R. and Harbison, S. (2020) 'A review of the method and validation of the MiSeq FGx™ Forensic Genomics Solution', *Wiley Interdisciplinary Reviews: Forensic Science*. John Wiley & Sons, Ltd, 2(1), p. e1351. doi: 10.1002/WFS2.1351.

Feng, Y. and Xiao, L. (2011) 'Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis', *Clinical Microbiology Reviews*. American Society for Microbiology Journals, 24(1), pp. 110–140. doi: 10.1128/CMR.00033-10.

Garcia-R, J. C. *et al.* (2021) 'First report of novel assemblages and mixed infections of *Giardia duodenalis* in human isolates from New Zealand', *Acta Tropica*. Elsevier, 220, p. 105969. doi: 10.1016/j.actatropica.2021.105969.

Garcia–R, J. C. *et al.* (2017a) 'Local and global genetic diversity of protozoan parasites: Spatial distribution of *Cryptosporidium* and *Giardia* genotypes', *PLoS Neglected Tropical Diseases*, 11(7), pp. 1–20. doi: 10.1371/journal.pntd.0005736.

Garcia–R, J. C. *et al.* (2017b) 'Local and global genetic diversity of protozoan parasites: Spatial distribution of *Cryptosporidium* and *Giardia* genotypes', *PLOS Neglected Tropical Diseases*. Edited by A. Caccone. Public Library of Science, 11(7), p. e0005736. doi: 10.1371/journal.pntd.0005736.

Hornung, B. V. H., Zwittink, R. D. and Kuijper, E. J. (2019) 'Issues and current standards of controls in microbiome research', *FEMS Microbiology Ecology*. Oxford Academic, 95(5), p. 45. doi: 10.1093/FEMSEC/FIZ045.

Illumina Inc. (2013) '16S Metagenomic Sequencing Library Preparation - Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System', *16S Metagenomic Sequencing Library Preparation Manual*.

Institute of Environmental Science and Research Ltd (ESR) (2018) *Annual Summary of Outbreaks in New Zealand 2016*.

Katz, D. E. *et al.* (2006) 'Prolonged outbreak of giardiasis with two modes of transmission', *Epidemiology and Infection*. Cambridge University Press, 134(5), pp. 935–941. doi: 10.1017/S0950268805005832.

Kearse, M. *et al.* (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*. Oxford Academic, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.

Liu, L. *et al.* (2020) '*Giardia duodenalis* induces apoptosis in intestinal epithelial cells via reactive oxygen species-mediated mitochondrial pathway in vitro', *Pathogens*, 9(9), pp. 1–14. doi: 10.3390/pathogens9090693.

Luján, H. D. and Svärd, S. G. (2011) *Giardia: A Model Organism*, *Springer Link*. doi: 10.1017/CBO9781107415324.004.

McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLoS ONE*. Edited by M. Watson. Public Library of Science, 8(4), p. e61217. doi: 10.1371/journal.pone.0061217.

Ortega-Pierres, M. G. *et al.* (2018) 'Recent advances in the genomic and molecular biology of *Giardia*', *Acta Tropica*, 184, pp. 67–72. doi: 10.1016/j.actatropica.2017.09.004.

Plutzer, J. *et al.* (2018) 'Review of cryptosporidium and giardia in the eastern part of Europe, 2016', *Eurosurveillance*. doi: 10.2807/1560-7917.ES.2018.23.4.16-00825.

Read, C. M., Monis, P. T. and Thompson, R. (2004) 'Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP', *Infection, Genetics and Evolution*. Elsevier, 4(2), pp. 125–130. doi: 10.1016/J.MEEGID.2004.02.001.

Ryan, U. *et al.* (2019) '*Giardia*: an under-reported foodborne parasite', *International Journal for Parasitology*. Pergamon, 49(1), pp. 1–11. doi: 10.1016/J.IJPARA.2018.07.003.

Savioli, L., Smith, H. and Thompson, A. (2006) '*Giardia* and *Cryptosporidium* join the "Neglected Diseases Initiative"', *Trends in Parasitology*. Elsevier Current Trends, 22(5), pp. 203–208. doi: 10.1016/J.PT.2006.02.015.

Seabolt, M. H., Konstantinidis, K. T. and Roellig, D. M. (2021) 'Hidden Diversity within Common Protozoan Parasites Revealed by a Novel Genomotyping Scheme', *Applied and Environmental Microbiology*, 87(6),

pp. 1–17. doi: 10.1128/aem.02275-20.

Sinha, R. *et al.* (2017) 'Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium', *Nature Biotechnology 2017 35:11*. Nature Publishing Group, 35(11), pp. 1077–1086. doi: 10.1038/nbt.3981.

Snel, S. J., Baker, M. G. and Venugopal, K. (2009) 'The epidemiology of giardiasis in New Zealand, 1997–2006', *THE NEW ZEALAND MEDICAL JOURNAL*, 122(5), p. 660. doi: 10.14219/jada.archive.1980.0221.

Thompson, R. and Smith, A. (2011) 'Zoonotic enteric protozoa', *Veterinary Parasitology*. Elsevier, 182(1), pp. 70–78. doi: 10.1016/J.VETPAR.2011.07.016.

Figures

Figure 1

The taxonomic distribution of (sub) assemblages in samples from the routine surveillance and the multiple outbreaks included in this study. The X axis shows the (sub) assemblage of each sample according to Sanger sequence data and the Y axis displays the number of samples corresponding to each assemblage; the colour codes in each bar represent the genetic diversity within each sample according to NGS.

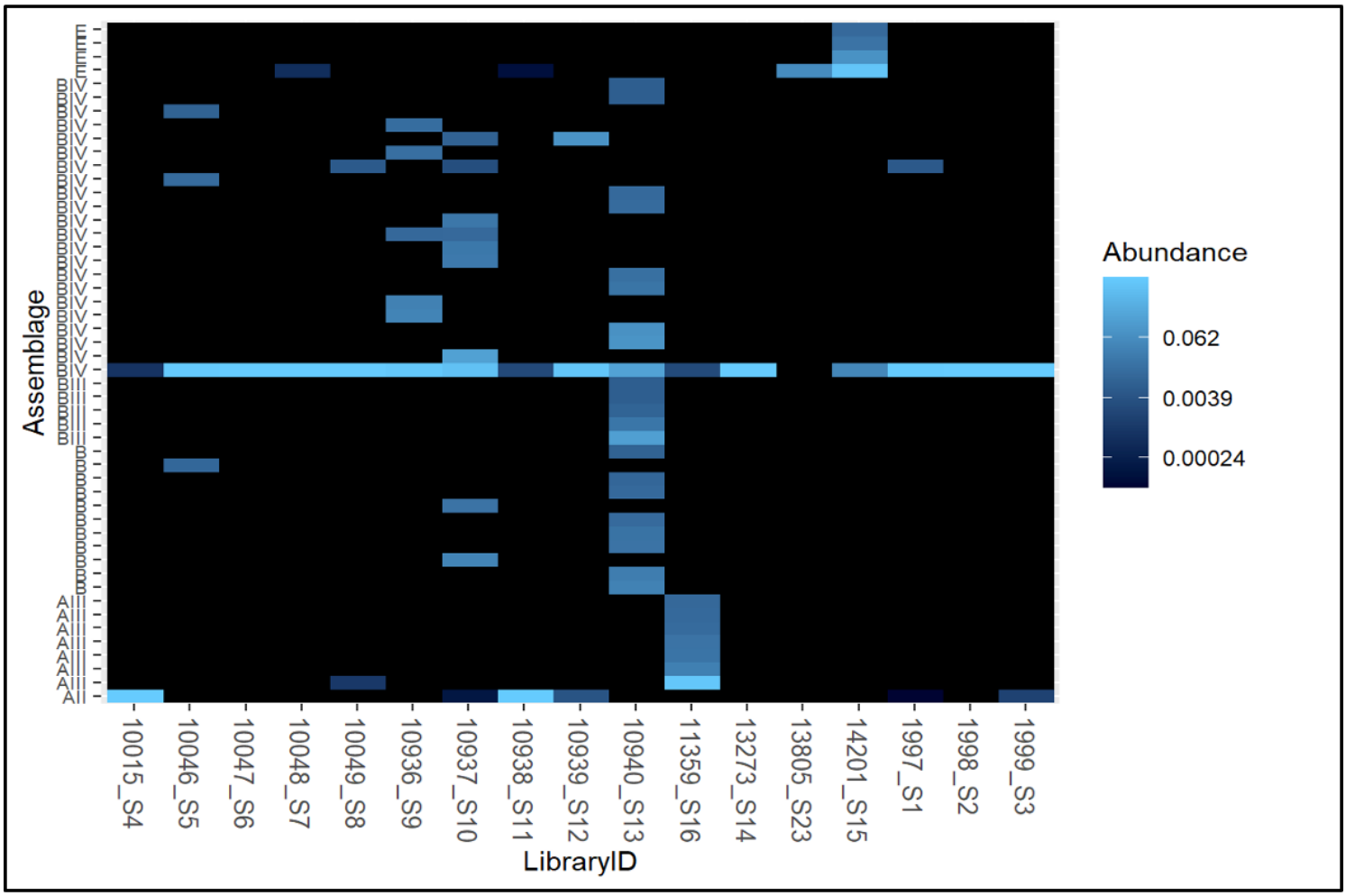
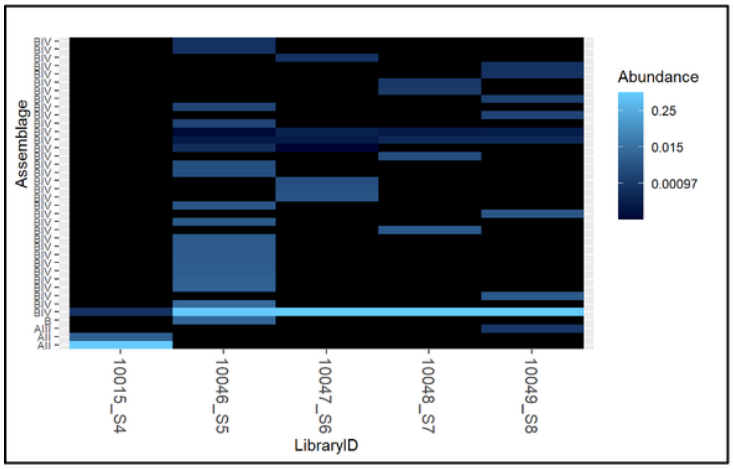
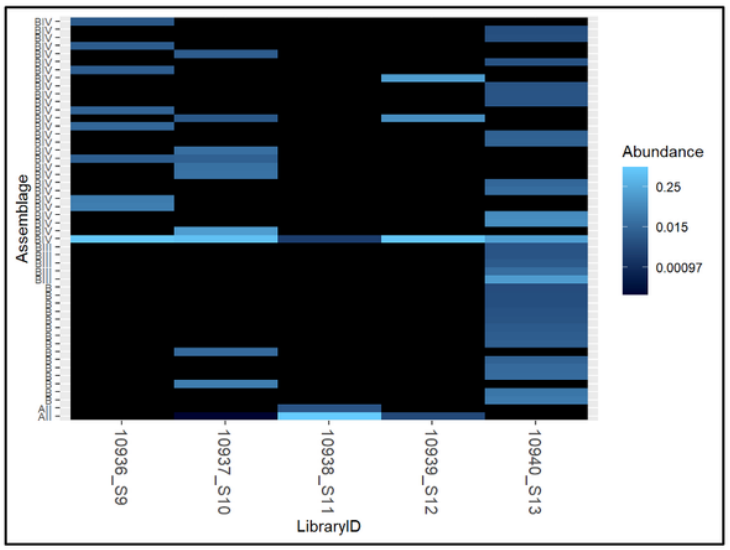


Figure 2

Heatmap showing the relative abundance of the top 50 sequences in each sample. The multiple variants of each assemblage present in each sample are displayed on the Y axis.



A



B

Figure 3

Heatmap showing the relative abundance of the top sequences in each sample from the outbreaks of giardiasis that occurred in Gisborne in 2014 (A) and in Hawke's Bay in 2015 (B). The multiple variants of each assemblage present in each sample are displayed on the Y axis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx
- SupplementaryFig.1..png