Molecular epidemiology of *Giardia duodenalis* in Australia

Amy Jeannette Asher
Bachelor of Science (Honours)
Macquarie University

Department of Biological Sciences
Faculty of Science and Engineering
Macquarie University

This thesis is presented for the degree of Doctor of Philosophy
August 2015

MACQUARIE University

"The identification and characterisation of Giardia is central to investigating and understanding the epidemiology of giardiasis. However, there are significant limitations in detection or diagnosis using conventional microscopic, biochemical, immunological and serological techniques..., such that there has been a need for reliable and practical molecular methods."

Koehler et al., 2014.

Table of Contents

List of Figures

List of Tables

Abstract

This study investigates *Giardia duodenalis* contributing to sporadic clinical cases in NSW between 2010 and 2013, and to cases among children of a high prevalence remote Indigenous community. The study also developed and evaluated two new molecular screening tools, terminal-restriction fragment length polymorphism (T-RFLP) and multilocus next generation sequencing for *Giardia* identification. In NSW, *G. duodenalis* assemblage B was highly prevalent, and patients aged 0 – 39 years were most at risk. Subtyping showed cases were highly similar (BIV). Spatial mapping demonstrated seasonal links in the geographic dispersal of assemblage A across NSW. High *Giardia* carriage rates (64.9%) were detected among remote Indigenous children, and over 50% of cases were not detectable by microscopy. A variety of assemblages, subassemblages/genotypes persisted in the community over a 12 month period, and these assemblages and subassemblages were not linked to child age or gender. Mixed assemblage B population infections were higher among remote community children than children in NSW. Application of T-RFLP and multilocus NGS to community studies enabled powerful detection of *Giardia*, and deep population analyses of individual cases. The study significantly progresses knowledge of *Giardia* transmission in NSW, and among remote community children, and has facilitated new approaches to *Giardia* research.

Declaration

All work described in this thesis was conducted at the Parasitology Laboratory at Macquarie University. All work described in this thesis is original and has not been submitted in any other form for a higher degree at any other university or institution.

The work described in this thesis is an original piece of work, composed by myself. All assistance received in performing this research has been acknowledged, and all sources of information used to prepare this thesis have been appropriately cited.

Amy Asher

Ethics and Funding Declaration

All work carried out for this thesis was performed under the approval of two human ethics committees and one biosafety committee.

Collection of NSW human faecal samples and laboratory faecal sample numbers were performed under approval of the Macquarie University Human Research Ethics Committee, approval number: 5201100403. Screening and analyses of remote community samples was performed under the approval of the Macquarie University Human Research Ethics Committee, under the same approval number.

Acquisition of NSW patient data and analyses of patient data pertaining to faecal samples was performed under approval from the NSW Population Health Services Research Ethics Committee, approval number: HREC/12/CIPHS/87, and with approval from NSW Health.

Collection of faecal samples, irradiation of faecal samples, and processing of faecal samples in the Marsupial Laboratory, Macquarie University, was performed under the approval of the Macquarie University Biosafety Committee, approval number: AMA160513BHA.

Funding for this research was provided by Macquarie University. Travel to Menzie's School of Health Research, Charles Darwin University, was provided by a grant from the Macquarie University Post Graduate Research Fund (PGRF).

Chapter Declarations

Chapter 1: Introduction and Thesis Rationale

I performed the literature review and writing of the thesis chapter. My principal supervisor Michelle Power, and co-supervisor Grant Hose, provided constructive feedback on the chapter.

Chapter 2: Giardiasis in NSW: Identification of *Giardia* subtypes contributing to human and cattle disease, and an epidemiological assessment of human cases

I was responsible for collection of human faecal samples from pathology companies in NSW. I was involved in the preparation of human ethics approval and biosafety approval for this project. I performed all laboratory work for human faecal samples, including DNA extraction, PCR assays, and RFLP. However, Lachlan Byatt of the Marsupal Laboratory, Macquarie University, assisted with DNA extractions. Remo Juliano Bartolini, Sinead Roberston-Cast, and Shanta Nair, also of the Parasitology Laboratory, assisted with a small number of DNA extractions. Patient information, including age, gender, and postcode were provided by Ben Polkingham of NSW Health, who also removed data pertaining to duplicate patient samples. I performed all data analyses for this chapter, including statistical analyses and spatial mapping. I wrote the manuscript, with constructive feedback from Michelle Power, Grant Hose, and Vicky Sheppeard of NSW Health.

Chapter 3: Rapid identification of *Giardia duodenalis* assemblages in NSW using terminal-restriction fragment length polymorphism

I performed all laboratory work for this chapter, as described above, and in addition, I prepared samples for DNA sequence analyses, and prepared fluorescent PCR libraries for T-RFLP analyses. The T-RFLP method was developed under the guidance of Liette Vandine.

This chapter builds upon sample screenings for 40 isolates which were initially screened for my Honours project, and these data have been added to the chapter to support the analyses. The T-RFLP method was comprehensively evaluated for my PhD project, analyses of the data, evaluation of the protocol, and writing of the manuscript was perfomed by myself, for the PhD thesis, with constructive feedback from Michelle Power and Liette Vandine.

Chapter 4: Distribution of *Giardia duodenalis* assemblages A and B among children living in a remote Indigenous community of the Northern Territory

I performed all molecular analyses for this chapter, including PCR assays which were performed at Menzies School of Health Research, Charles Darwin University. I prepared PCR products for DNA sequencing and T-RFLP at the Parasitology Laboratory, Macquarie University. The DNA samples that were screened in this study by PCR were previously extracted by Menzies Researchers who were involved in the primary study from the remote community. Menzies Researchers collected faecal samples and performed microscopy. The microscopy results were kindly provided for this study by the principal investigator of the primary study, Ross Andrews, and Deborah Holt of Menzies School of Health Research. Child age, gender, and collection date were provided for this study by Ross Andrews and Deborah Holt. I performed all data analyses for this chapter, including DNA sequence and T-RFLP analyses, and all distributional analyses relating to screening results, genetic assemblages and participant data. Constructive feed back was provided by Michelle Power, Ross Andrews and Deborah Holt. I wrote the manuscript, with constructive feedback from Michelle Power, Ross Andrews, and Deborah Holt.

Chapter 5: Multilocus next generation sequence typing of *Giardia* from children living in a remote Indigenous community of Australia

I performed all molecular work for this paper. I performed the inital PCR screenings of remote community DNA at Menzies School of Health Research, Charles Darwin University. I prepared the library for next generation sequencing at the Parasitology Laboratory, Macquarie University. I performed all aspects of next generation sequencing data analyses for this chapter. Matthew Lott kindly provided advice on the analyses of raw NGS data. I wrote the manuscript with constructive feedback from Michelle Power and Grant Hose.

Chapter 6: Discussion and conclusions

I evaluated the major findings of the research and the future directions. I wrote the chapter with constructive feed back from Michelle Power and Grant Hose.

Chapter 7: Appendix

These chapters do not form part of the formal PhD thesis and have not been included for examination. These chapters have been included as my additional contributions to *Giardia* research.

Chapter 7.1: I performed the laboratory work and I wrote this chapter for my Honours project. All laboratory and written work was performed under the guidance of Michelle Power and Liette Vandine, additional assistance for the laboratory work was kindly provided by Cristel Cheung-Kwok-Sang.

Chapter 7.2: All work for this chapter was performed and written by Tiffany Delport. The *Giardia* PCR methods used in this chapter builds on the methods I developed for my Honours project (Chapter 7.1), and I provided advice on optimising *Giardia* PCRs and feedback on the paper.

Publications, Conference Presentations and Media

Research chapters:

Asher, A. J., Holt, D. C., Andrews, R. M., Power, M. L. 2014. Distribution of *Giardia duodenalis* Assemblages A and B among Children Living in a Remote Indigenous Community of the Northern Territory, Australia. PloS One, 9: e112058

Asher, A. J., Waldron, L. S., Power, M. L. 2012. Rapid identification of *Giardia duodenalis* in NSW using terminal-restriction fragment length polymorphism. Parasitol 139, 1005-1013

Appendix chapters:

Asher, A. J., Waldron, L. S., Power, M. L. 2012. Evaluation of a PCR protocol for sensitive detection of *Giardia duodenalis* in human faeces. Parasitol Res 110, 853-858

Delport, T. C., Asher, A. J., Beaumont, L. J., Webster, K. N., Harcourt, R. G. 2014. *Giardia duodenalis* and *Cryptosporidium* occurrence in Australian sea lions (*Neophoca cinerea*) exposed to varied levels of human interaction. Int J Parasitol Parasites Wildl 3, 269-279

Conference presentations

IV International *Giardia* and *Cryptosporidium* Conference (2012). Poster presentation 'Fluorescence based detection for *Giardia intestinalis* assemblages'

International Congress for Parasitology (2010). Poster presentation '*Giardia duodenalis*: molecular epidemiology and development of T-RFLP for rapid identification'

Media:

Australian Indigenous Health *InfoNet* , 26 November 2014.

http://www.healthinfonet.ecu.edu.au/about/news/2756

High carriage rates of Giardia among children in a remote Indigenous community

Date posted: 26 November 2014

Recent research in a remote Indigenous community in the Northern Territory has found high carriage rates of the intestinal parasite *Giardia* among children. *Giardia duodenalis* is an enteric parasite that causes gastrointestinal disease (giardiasis) including symptoms of diarrhoea and vomiting, but can lead to malnutrition and wasting among children who are constantly infected.

The study, published by PLOS ONE, was performed by Menzies School of Health Research and Macquarie University, and used DNA screenings to detect *Giardia* in samples collected from children under 15 years of age.

In Australia, the overall prevalence of human giardias is estimated between 2-7%, but is much higher in remote Indigenous communities, particularly among children. Researchers screened 87 samples collected in the community over an 18 month period. The study reports that over 60% of children under 10 years old were positive, and that two different genetic types of *Giardia* (assemblages A and B), known to infect humans, were identified among children.

Macquarie University's Amy Asher says the findings have implications for disease management and improving the health of Indigenous children living in remote communities.

Acknowledgments

I would like to express my deepest gratitude to my principal supervisor, Dr Michelle Power. I would never have dreamt that this was all possible, without your guidance and dedication to teaching me, throughout my entire PhD program. I am deeply grateful for all the time and effort you have given me, for your dedication, and for your constant encouragment and enthusiasm, to teach me throughout this entire time. You have been an endless source of knowledge and inspiration, and I would like to sincerely thank you for all of this.

To my co-supervisor, A/Prof. Grant Hose, I would like to express my deepest gratitude for all your help throughout my PhD program. Thank you for your guidance and assistance, throughout this time, for all your advice and feedback on my manuscripts, and constant support.

I would like to sincerely thank Dr Deborah Holt and Prof. Ross Andrews of Menzies School of Health Research, for giving me the opportunity to visit their labs in Darwin, which I thoroughly enjoyed, and has been a highlight of my PhD. I would also like to thank them for their advice and feedback on my manuscript during my PhD.

I would also like to sincerely thank Dr Liette Vandine, who has kindly provided her time to teaching me fluorescence, and has provided advice and feedback on my manuscript.

I would like to thank everyone in lab; Tiffany Delport, Matthew Lott, Elke Vermeulen, Dr Koa Webster, Lachlan Byatt, Dr Tina Wunderlin and Sabine Schiller. Thank you for your advice, support and making this whole experience wonderful.

Last, but not least, my family, who have been behind me, in every step of this journey. To my parents, Jeannette and Rex Asher, to my partner Broc Townsend, to my sister Ainslie Asher and her partner Henry Hidayat, and to my uncle Phillip Morgan. You have all helped me immensely over the years whilst I have been a perpetual student. I am forever grateful for your endless support and encouragement throughout this time, this would not have been possible without you.

Extended Summary

Giardia duodenalis, a protozoan parasite, is a leading cause of human gastrointestinal disease (giardiasis) worldwide. Several genetic assemblages, subassemblages, and subtypes of *G. duodenalis* infect human and animal hosts, but transmission cycles and risk factors associated with infection of these assemblages and subassemblages are unclear. To understand giardiasis transmission and risk factors requires molecular epidemiological surveys among human populations, in low and high prevalence communities. Highly sensitive and cost effective screening tools are also needed, to improve sample screenings, typing accuracy, and to evaluate the genetic diversity of *Giardia* among human cases.

In Australia, giardiasis is an emerging disease concern, and in particular, in remote Indigenous communities. In the State of New South Wales prevalence of human giardiasis is low, but annual cases of sporadic giardiasis have been steadily increasing since 2002. In remote Indigenous communities of Australia, prevalence rates of *Giardia* among children are high, and are similar to rates observed in developing nations. Chronic giardiasis is linked to malnutrition and failure to thrive, and the burden of chronic gastrointestinal infections among children living in remote Indigenous communities is of increasing concern.

The aim of this study was to investigate human *Giardia* cases in NSW and a remote Indigenous community, and to examine the molecular epidemiology of *G. duodenalis* contributing to human cases in these regions, where giardiasis is an emerging concern. The study also aimed to develop new molecular tools for *Giardia* identification to enable deep assessment of the diversity of assemblages/subassemblages among human cases, in these low and high prevalence communities. In both communities, both human infective *G. duodenalis* assemblages A and B were identified among human cases, by analyses of *Giardia* typing loci (18S rRNA, *gdh*), using DNA sequencing, restriction fragment length polymorphism (RFLP), and terminal–RFLP (T-RFLP). Additional analyses of remote community DNA samples

were performed using four *Giardia* loci (18S rRNA, *gdh*, *tpi*, *β-giardin*) and next generation sequencing.

In NSW, the results of two separate investigations showed that assemblage B was the most common cause of sporadic giardiasis, accounting for a total of 86.5% (205/237) of all patient samples that were successfully genotyped. Spatio-temporal, and demographic distributions, of assemblages A and B were assessed for 165 patient cases occurring in NSW between 2010 and 2013. The overall distribution of assemblage A (<17%) and B (>83%) cases in NSW was similar among total cases examined, including age, gender, season, and NSW local health district groups. The demographic groups most at risk from sporadic giardiasis were infants and children aged 0 – 9 years, and adults aged 30 - 39 years. Among these groups assemblage B prevalence exceeded 90%. The study, however, also demonstrated high prevalence of assemblage B (92.2%) among all patients aged 0 – 39 years, which significantly differed from patients over 40 years, where assemblage B prevalence was 74.5%. For all assemblage A and B NSW cases, subassemblage AII (13.1%), and genotype BIV (81.4%) were most frequently detected. Anthroponotic transmission of subassemblage AII cases is inferred from the limited host distribution of this subassemblage. Assemblage B genotypes are commonly reported from human and animals hosts, and may represent anthroponotic and/or zoonotic cases in NSW.

Spatial mapping of all NSW *Giardia* positive human cases (n = 243), and human assemblage A and B cases (n = 165), identified the Newcastle and lower Hunter region as disease hotspots in NSW. The analysis demonstrated shifts in the geographic dispersal of assemblages A and B, occurring among patients in these hotspot regions, during low and high prevalence seasons of Winter/Spring and Autumn. During these seasons, geographic dispersal patterns were most pronounced for assemblage A. Potential for zoonotic transmission was examined in NSW by PCR-RFLP analyses of 175 NSW cattle DNA samples, at the 18S rRNA, and *gdh* loci. Approximately 37% (64/175) of cattle were PCR

positive, and bovine specific assemblage E (69%), and the human infective assemblage B (31%) were identified. Confirmation of assemblage B cattle cases was only achieved at one *Giardia* locus. These samples, however, were collected from dairy and beef cattle farms in locations proximal to human populations, where assemblage B human cases were identified.

In a remote Indigenous community of the Northern Territory, *Giardia* was detected among 64.9% (48/74) of children <15 years of age, and over 50% of these cases were not detectable by microscopy. Comparison of genotyping results for children (< 15 years) in NSW (n = 57), and the Northern Territory (n = 38), showed highly dissimilar frequencies of assemblages, subassemblages, and genotypes, occurring among children in each region. In NSW, 5.3% and 94.7% of children were infected with assemblages A and B respectively. Of these, only subassemblage AII and genotype BIV were identified. Mixed template infections were not detected among this group. Among remote Indigenous children, 28.9% of cases were assemblage A, and all cases were subassemblage AII. Assemblage B was detected among 71.1% of participating children, including genotypes BIV (37%), BIII (11.1%), and mixed BIII/BIV (22.2%). Further analyses of samples from remote Indigenous children (n = 17) by multilocus next generation sequencing, identified three different subtypes (AII-1, AII-2, AII-3) contributing to subassemblage AII cases. Additionally, five different assemblage B multilocus genotypes were identified, and high levels of genetic diversity were found among all assemblage B cases in this community.

The study developed and applied two new molecular typing tools for *Giardia* identification, including terminal-restriction fragment length polymorphism (T-RFLP), and multilocus next generation sequencing (NGS). Terminal-RFLP and multilocus NGS were applied to molecular screenings in NSW (T-RFLP) and in the Northern Territory (T-RFLP, NGS). Comparison of these tools to established DNA sequencing and RFLP showed consistency in assemblage and subassemblage/genotype assignment by T-RFLP and NGS, in each community application. The high throughput, and high sensitivity of fluorescent

detection by T-RFLP, improved identification of mixed assemblage B templates among NSW samples, and was a rapid diagnostic tool for *Giardia* sample screenings. Similarly, application of NGS to multilocus typing enabled detection and separation of mixed assemblage A and B populations, from a single case in the Northern Territory. Detection of novel assemblage B sequences by T-RFLP, and detection of related sequences by multilocus NGS, demonstrated potential application of these tools to disease source tracking.

The results of this thesis demonstrate complex disease links between the diversity of *Giardia*, prevalence and frequency of human giardiasis transmission, and environmental conditions in Australian communities that contribute to the dispersal of different types of human infective *Giardia*. Development and application of new molecular diagnostic methods provides tools to better understand human infections, and the diversity of *Giardia* contributing to these infections. Increased knowledge of disease factors that contribute to human giardiasis, and the availability of highly sensitive, and informative molecular diagnostic tools, are essential for future disease management and control, to accurately identify *Giardia* cases, and provide baseline information that informs public health strategies.

1. Introduction and Thesis Rationale

1.1. *Giardia* and giardiasis: a neglected and re-emerging disease concern

Giardia is a parasitic protozoan that infects the gastrointestinal tract of vertebrate hosts (Ankarklev et al., 2010; Caccio and Ryan, 2008). *Giardia* was first identified by Antonie van Leeuwenhoek in 1676, who discovered the organisms in his own stool (Dobell, 1920; Lipoldova, 2014), and then in 1681 he provided the first description of *Giardia* occurring intermittently in stool, and during diarrhoeal episodes (Dobell, 1920). Although *Giardia* was well documented as an intestinal organism, its association to human disease was not fully recognised for a further 300 years (Escobedo and Cimerman, 2007; Lipoldova, 2014). The species *Giardia duodenalis* (syn. *Giardia lamblia, Giardia intestinalis*), which infects a wide range of mammalian (human, domestic animals, wildlife) hosts (Thompson and Monis, 2012), was first listed as human pathogen in 1981 by the World Health Organisation (Escobedo and Cimerman, 2007).

Giardia duodenalis has since been recognised as a leading cause of human diarrheal disease (giardiasis), and one of the most common intestinal parasites among humans and domestic animals worldwide (Feng and Xiao, 2011; World Health Organisation, 1987). Children, immuno-compromised individuals, and populations in high prevalence regions are most at risk from the disease (Caccio and Ryan, 2008; Monis and Thompson, 2012). In 1995, giardiasis was listed as a re-emerging infection in the United States by the U.S. National Science and Technology Council, due to increasing incidence among infants attending day care centres (Lederberg, 1996). Most recently, giardiasis was recognised as a neglected disease, due to high prevalence and persistence in regions of low socio-economic status, and due to associations between giardiasis and ongoing poverty (Feasey et al., 2009; Savioli et al., 2006).

Since the 1980's advances in DNA detection, amplification and sequence based technologies have enhanced our understanding of human and animal giardiasis, by enabling genetic characterisation of the parasite (Monis et al., 1999). Genetic sub-structuring within *G. duodenalis* is well documented, but the ecology and diversity of genetic types, and associations to disease transmission and infection are currently unresolved. Molecular screening tools to accurately characterise *G. duodenalis* diversity are still emerging. PCR amplification of *Giardia* DNA across numerous loci remains problematic, limiting the application of molecular tools and the feasibility of research to perform large scale screenings (Breathnach et al., 2010; Caccio and Ryan, 2008).

Globally, *G. duodenalis* poses a continuous public health threat to human populations, through the waterborne transmission of the parasite, potential for cross-species transmission between humans and non-human hosts, and the potential for chronic disease progression among vulnerable populations (Hanevik et al., 2009; Thompson and Monis, 2012; Wang et al., 2014). Giardiasis is increasingly linked to chronic morbidity, including developmental impairments among children with persistent infections (McDonald and Bailie, 2010; Thompson, 2008), and the development of gastric disorders in adults (Robertson et al., 2010). The high prevalence of giardiasis among developing nations, increasing incidence in developed nations, and increased knowledge of chronic disease outcomes, has highlighted giardiasis as a re-emerging, global health concern (Ignatius et al., 2012; Thompson, 2000).

1.2 *Giardia* species: taxonomy, biology and lifecycle

1.2.1 Taxonomy and biology of *Giardia*

Six recognised species of *Giardia* infect mammals (*G. duodenalis*), birds (*G. psittaci* and *G. ardeae*), rodents (*G. microti* and *G. muris*), and amphibians (*G. agilis*) (*c.f.* Thompson et al., 2000). Identification of *Giardia* species is based on morphological differences, such as the shape and length/width of the trophozoite and median body, or electron microscopic

features of the cyst (Adam, 2001; Monis et al., 2009; Thompson et al., 2000). *Giardia* trophozoites are pear shaped, ovular, or round in appearance, range from 10 – 31µm in length, possess a unique median body organelle, four pairs of flagella, and a ventral sucking disk (Adam, 2001; Thompson et al., 2000).

The mammalian infective species, *G. duodenalis*, has a broad host range, including humans, domestic animals (cattle, sheep, pigs, dogs, cats) and wildlife (non-human primates, marsupials, pinnipeds) (Caccio and Ryan, 2008; Lasek-Nesselquist et al., 2010). *Giardia duodenalis* cysts are small and ovular, measuring 8-10µm (Karanis et al., 2007) and containing a 0.5µm protective cell wall (Adam, 2001). The species is characterised by the trophozoite's pear shaped appearance (12 – 15µm length by 6 – 8µm width) (Figure 1), and two claw shaped median bodies (Thompson et al., 2000).

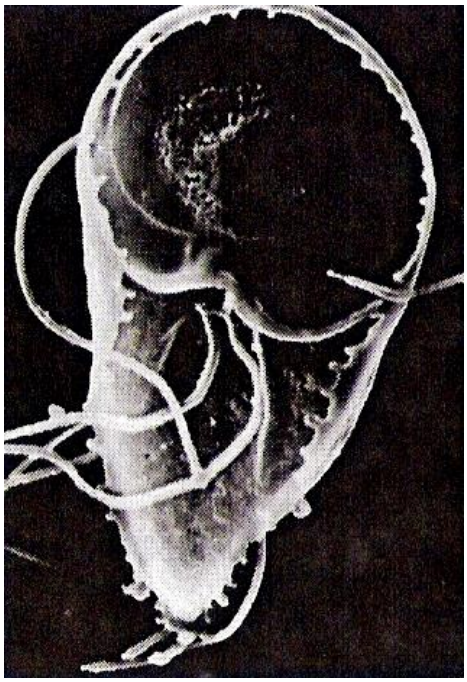

Figure 1.1 Pear shaped *Giardia duodenalis* trophozoite (12 – 15µm length), viewed from the ventral surface. The ventral sucking disk appears at the top of the image, and four pairs of flagella are visible (Cox, 2004).

1.2.2 *Giardia* lifecycle

Transmission of *Giardia* between susceptible hosts occurs via the faecal oral-route, and is facilitated by the parasite's two-stage life cycle (Figure 1.2) (Ankarklev et al., 2010; Xiao and Fayer, 2008). *Giardia* cysts are excreted within the faeces of an infected host (Ankarklev et al., 2010). An individual host can shed up to 10^9 cysts per day, and cysts are viable in the environment for several weeks (Xiao and Fayer, 2008). In water, at approximately 4°C, *Giardia* cysts may be infectious for up to 11 weeks (Olsen et al., 1999). *Giardia* has a low infectious dose of 10 to 100 cysts, and susceptible hosts acquire *Giardia* through ingestion of excreted cysts (Rendtorff, 1954; Svard et al., 2003). Transmission occurs directly through contact with infective faeces, or indirectly through contaminated drinking and recreational water, and to a lesser extent, contaminated surfaces such as fresh food and animal coats (Thompson and Monis, 2012).

Once ingested, *Giardia* cysts undergo excystation, a process of cell breakdown and the release of *Giardia* trophozoites into the host's small intestine (Ankarklev et al., 2010). Excystation is triggered by passage of the cyst through the host's stomach and exposure to an acidic environment (Ankarklev et al., 2010; Svard et al., 2003). For each cyst ingested, two *Giardia* trophozoites are formed (Adam, 2001). Trophozoites are the active stage of *Giardia*'s lifecycle and reproduce by asexual binary fission (Monis and Thompson, 2003). Trophozoites attach to the intestinal microvilli but do not invade host cells (Ankarklev et al., 2010). Detachment from the microvilli and migration towards the lower small intestine exposes trophozoites to biliary fluids, which triggers encystation (Ankarklev et al., 2010; Svard et al., 2003). During this process, trophozoites form the environmentally robust and infectious cysts, prior to environmental excretion by the infected host (Monis and Thompson, 2003).

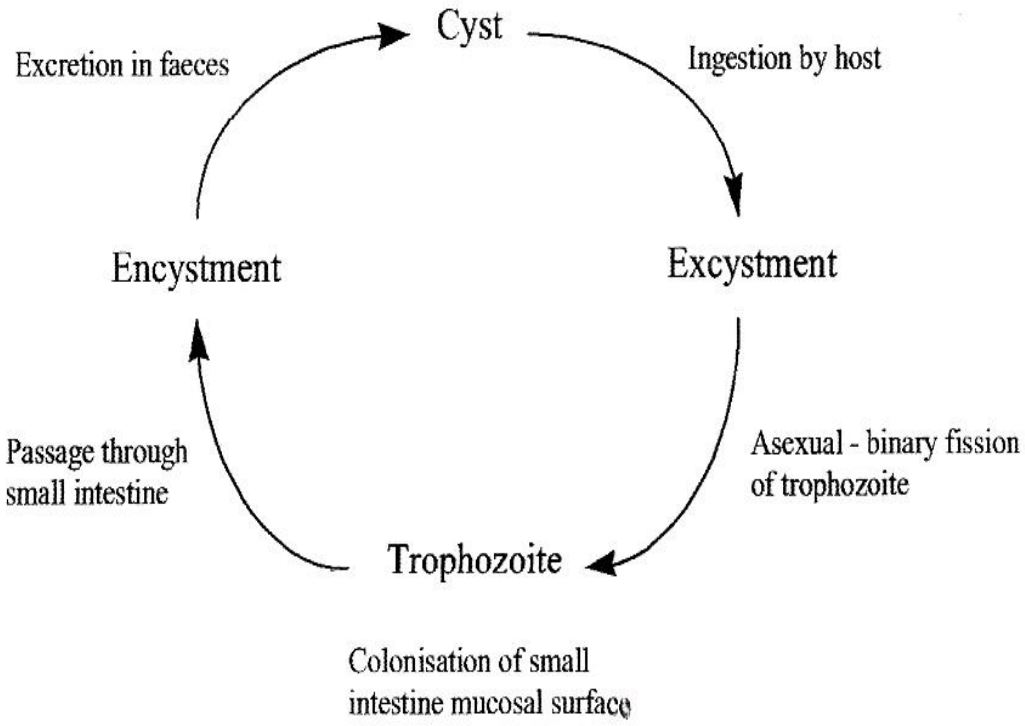

Figure 1.2 *Giardia* has a two-stage life cycle consisting of an active trophozoite which colonises the host's small intestine, and an environmentally robust and infective cyst, that forms prior to environmental excretion and facilitates parasite transmission between susceptible hosts (Monis and Thompson, 2003).

Recently, the presence of a sexual or parasexual lifecycle stage has been suggested by several authors (Andersson, 2012; Caccio and Sprong, 2010; Poxleitner et al., 2008; Tederovic et al., 2007). Evidence for genetic recombination has been provided by fluorescent *in-situ* hybridization experiments, demonstrating plasmid transfer between the nuclei of a single cyst during encystation (Poxleitner et al., 2008). Meiotic gene homologues have also been identified within the *Giardia duodenalis* genome, but the function of these genes are not known and direct evidence of sexual recombination is lacking (Caccio and Sprong, 2010; Ramesh et al., 2005)

1.3 Epidemiology of giardiasis: clinical manifestations, prevalence and disease sources

1.3.1 Clinical manifestions of giardiasis

Giardiasis produces a broad spectrum of acute and chronic gastrointestinal symptoms; including; diarrhea, vomiting, nausea, abdominal pain, and weightloss (Cotton et al., 2011; Roxstrom-Lindquist et al., 2006). The disease is also commonly asymptomatic, and can occur intermittently (Cotton et al., 2011). Chronic infections are associated with wasting, malnutrition, and failure to thrive, particularly among children with persistent giardiasis (Thompson, 2008). Patients with poor immune status, due to conditions such as lymphomas, and hypogammaglubulinaemia are at an increased risk of giardiasis and chronic disease progression (Escobedo and Cimerman, 2007; Nash et al., 2001; Robertson et al., 2010). Giardiasis is not considered an opportunistic infection associated with HIV/AIDs, however, the risk of chronic and symptomatic disease is higher (Escobedo and Cimerman, 2007; Espelage et al., 2010; Nash et al., 2001). Development of life threatening giardiasis has also been reported in an HIV positive patient (Aronson et al., 2001).

Giardiasis symptoms occur between 1 and 2 weeks following infection and the duration of giardiasis varies between several days to several weeks and months (United States Environmental Protection Agency, 1999; Upcroft and Upcroft, 2001). The longevity and severity of giardiasis is highly variable between patients, even when infection is acquired from a common disease source. Following an outbreak in Norway, 38% to 41% of patients reported ongoing abdominal pain and fatigue two years after the initial *Giardia* infection (Hanevik et al., 2007). Giardiasis has been linked to post infection irritable bowl syndrome and functional dyspepsia, six months after successful treatment for giardiasis (Hanevik et al., 2009). Treatment refractory disease has also been reported (Nash et al., 2001). Conversely, the disease often presents as mild and self resolving, may be limited in duration, or successfully treated with 5-nitroimidazoles (Robertson et al., 2010).

Clinical manifestations of giardiasis are poorly understood. Attachment of *Giardia*

trophozoites to the mucosal surface of the host's intestinal microvilli elicits an immune response, but *Giardia* has no known mechanism of inducing disease symptoms among infected hosts (Ankarklev et al., 2010; Escobedo and Cimerman, 2007). Close association between trophozoites and host epithelium most likely produces diarrhoea, through villus atrophy, brush boarder shortening, and induction of the host immune response (Cotton et al., 2011). Switching of variable surface proteins (VSPs) enables evasion of the host immune response by trophozoites (Prucca et al., 2009) but the frequency of VSP switching is unclear (Ankarklev et al., 2010). The motility of trophozoites, and ability to attach to the epithelial surface via the ventral sucking disk provide successful means of host colonisation (Ankarklev et al., 2010). The broad range of symptoms exhibited by giardiasis most likely represents complex host-parasite interactions, influenced by the host's age, immune status, response to treatment and the genetic variability of the parasite (Ankarklev et al., 2010; Roxstrom-Lindquist et al., 2006).

1.3.2 Giardiasis prevalence among human populations

An estimated 280 million cases of symptomatic giardiasis occur globally each year (*c.f.* Caccio and Ryan, 2008). The contribution of *Giardia* to gastrointestinal infections among humans worldwide is likely underreported due to a high rate of asymptomatic disease, which may occur in up to 75% of *Giardia* cases (Hoque et al., 2002; United States Environmental Protection Agency, 1999). The overall prevalence of human giardiasis has been estimated at one to eight percent in developed countries, and between 30% and 40% in developing countries and regions of low socio-economic status (Feng and Xiao, 2011; Thompson and Monis, 2012; Upcroft and Upcroft, 2001). Poor sanitary and hygiene infrastructure and overcrowded living conditions enhance faecal-oral transmission of the parasite and contact between susceptible hosts in these regions (McDonald et al., 2008).

Globally giardiasis is highest among children and infection rates range between one and 36% (*c.f.* Feng and Xiao, 2011; United States Environmental Protection Agency, 1999). High rates of *Giardia* positivity, exceeding 50% and 60%, among children under 15 years have been documented in Southern India and Western Uganda (Johnston et al., 2010; Kang et al., 1998). Bimodal age distributions are common among developed nations, including the United States (Naumova et al., 2000; Yoder et al., 2012), New Zealand (Hoque et al., 2002), Canada (Laupland et al., 2005), Australia (NSW Department of Health, 2011), and among countries of the European Union (European Centre for Disease Prevention and Control, 2014). Notifications peak among children under 5 years, and a secondary peak occurs among adults aged approximately 25 to 49 years. Conversely, in high prevalence regions, steady declines in giardiasis infection with increasing age are commonly reported (Cifuentes et al., 2000; Kang et al., 1998; Mukherjee et al., 2014). Males and females are equally affected by giardiasis (Breathnach et al., 2010; Choy et al., 2014; Hoque et al., 2002), but slightly higher prevalence rates among infant males have been reported by some studies (Laupland et al., 2005; Mahdy et al., 2008; Yoder et al., 2012), and most likely represent behavioural differences.

1.3.3 Disease sources and risk factors for infection

Giardia duodenalis is globally distributed, and human giardiasis occurs both sporadically and in outbreaks of the disease (Caccio and Ryan, 2008; Sprong et al., 2009). Waterborne transmission is a primary vehicle for giardiasis associated with outbreaks, which have been well documented in developed nations (Karanis et al., 2007; Sprong et al., 2009). Giardiasis is the most frequent cause of waterborne outbreaks in New Zealand (Hoque et al., 2002) and contributes to almost 30% of waterborne outbreaks in the United States (1971 – 2006) (Craun et al., 2010). Consumption of contaminated water, from rivers and lakes, recreational swimming water (including pools), unfiltered and/or unchlorinated water, and

from water treatment deficiencies have contributed to 6.1% to 21.2% of waterborne giardiasis outbreaks in Australia, New Zealand, United Kingdom, Europe, Japan, and North America (Karanis et al., 2007).

In developed nations attendance at day centres presents a major risk factor for giardiasis among children and their carers, including parents and staff (Centres for Disease Control and Prevention, 2012). An estimated 7% – 54% of children attending day care centres in the United States have *Giardia*, and transmission of the infection occurs in up to 35% of their carers (United States Environmental Protection Agency, 1999). These figures indicate that over 1.6 million people in the United States are at risk of acquiring giardiasis through day care centres alone (United States Environmental Protection Agency, 1999). In New Zealand, exposure to human waste by nursing mothers and through occupational exposure has been identified as a significant risk factor for acquiring giardiasis (Hoque et al., 2002). Travel to highly endemic regions of Africa, Asia, and the South Pacific, frequent swimming, storage of untreated water for domestic use, consumption of green salads, and contact with farm animals have also been associated with giardiasis transmission in developed nations (Espelage et al., 2010; Hoque et al., 2002).

In high prevalence regions, contamination of drinking water through poor sanitary infrastructure, storage of water in uncovered vessels, and not boiling water have been associated with increased risk of giardiasis (Choy et al., 2014; Cifuentes et al., 2000; Mahdy et al., 2008). Additional risk factors such as not washing hands frequently, lacking a functional household toilet, bathing in lakes, consumption of raw market vegetables, and occupational contact with farm animals and manure have been documented (Cifuentes et al., 2000; Mahdy et al., 2008; Wegayhu et al., 2014).

1.4 Genetic diversity of *Giardia duodenalis*: prevalence and epidemiology of genetic assemblages

1.4.1 Molecular detection of *Giardia duodenalis* and description of a species complex

Giardia duodenalis cysts and trophozoites are morphologically identical; however, genetic diversity within the species is extensive (Thompson et al., 2000). Early studies detected genetic variation within axenic isolates of *G. duodenalis* using isoenzyme electrophoresis of numerous *Giardia* loci, southern blotting and enzymatic digestion of DNA (Andrews et al., 1989; Mayrhofer et al., 1995; Monis et al., 1996; Monis et al., 2003; Nash et al., 1985). Genetic variation within *G. duodenalis* was first detected by Andrews et al., in 1989. The genetic differences between *G. duodenalis* isolates that were comparable to species level diversity (Andrews et al., 1989), this observation was later supported by the findings of Mayrhofer et al. (1995). Among human infective isolates, two distinct human infective genetic groups were defined, which were initially referred to as 'Polish' and 'Belgium' groups, and later recognised as two genetic assemblages, A and B (Table 1) (Homan et al., 1992; Mayhofer et al., 1995; Monis et al., 1999; Nash et al., 1985). Non-human infective assemblages were also identified, and in 2003, Monis et al. linked isolates from seven *G. duodenalis* assemblages (A to G) to broad mammalian host groups.

Application of PCR amplification to *G. duodenalis* DNA (Weiss et al., 1992) has facilitated genetic characterisation of *G. duodenalis* directly from clinical and environmental samples, removing the need to culture isolates (Monis et al., 1999; Traub et al., 2005b). Numerous studies have since documented the genetic substructuring within *G. duodenalis*, and international surveys of publically available *G. duodenalis* sequence data have confirmed links between genetic assemblages and host groups described by earlier studies (Sprong et al., 2009; Weilinga and Thompson et al., 2007). Identification of *G. duodenalis* assemblages is achieved by PCR amplification and sequence analyses of several *G. duodenalis* genes (Lebbad et al., 2011; Traub et al., 2005b), most commonly, the small subunit ribosomal RNA (*18S rRNA*), glutamate dehydrogenase (*gdh*), triosphosphate isomerase (*tpi*), and beta giardin (*β-giardin*) (Ankarklev et al., 2012; *c.f.* Feng and Xiao, 2011).

Table 1.1 *Giardia duodenalis* assemblages A to H, and relationship to mammalian host range, *Giardia duodenalis* comprises eight genetic assemblages (A to H) including the human infective assemblages A and B, and host specific assemblages C to H. Host range has been documented by numerous studies and international surveys, and recently, species names have been proposed for assemblages A to G.

Assemblage	Host	Proposed species name[a]	References
A	Human, non-human primates, ruminants, pigs, horses, canines, marmoset, rodents,	*Giardia duodenalis*	Andrews et al., 1989; *c.f.* Caccio and Ryan, 2008; Ey et al., 1997; Hopkins et al., 1997; Hopkins et al., 1999; Lasek-Nesselquist et al., 2010; Mayrhofer et al., 1995; Monis et al., 1996; Monis et al., 1998; Monis et al., 1999; Monis et al., 2003; *c.f.* Monis et al., 2009; Sprong et al., 2009; *c.f.* Thompson and Monis 2012; Weilinga and Thompson, 2007; *c.f.* Feng and Xiao., 2011
B	Human, non human primates, cattle, canines, marsupials guinea pig, cat, alpaca, beavers	*Giardia enterica*	
C	Canines	*Giardia canis*	
D	Canines	*Giardia canis*	
E	Ruminents, pigs	*G.bovis*	
F	Cat	*G. cati*	
G	Rodents	*G. simondi*	
H	Seals		

[a]Proposed species names for *Giardia* assemblages A to G, according to Monis et al., 2009

1.4.2. *Giardia duodenalis* assemblages, subassemblages and genotypes, and relationship to mammalian host groups

Giardia duodenalis is considered to be a species complex that includes eight genetic assemblages (A to H) (Table 1.1). Phylogenetic studies that show that *Giardia* assemblages are distinct evolutionary lineages, associated with several mammalian host groups, and consist of broad clusters of genetically related isolates (Monis et al., 2003; Sprong et al., 2009). Assemblages A and B infect humans, domestic animals, and wildlife (Table 1.1). Worldwide, studies show that humans are only infected by assemblages A and B, whilst domestic animals and wildlife are most commonly infected by host specific assemblages C to H (Table 1.1.) (*c.f.* Caccio & Ryan, 2008; *c.f.* Feng and Xiao, 2011Sprong et al., 2009). Within assemblages A and B, several subassemblages (AI to AIV; BI to BIV) have been identified (Table 1.2) (Read et al., 2004; Sprong et al., 2009). Analyses of human and animal isolates using sequence data from multiple *Giardia* loci show subassemblages AI to AIII are associated with different host groups and comprise of several distinct genotypes (Table 1.2) (Caccio et al., 2008; Sprong et al., 2009).

Within assemblage B, subassemblages BI to BIV have been detected (Table 1.2), and until recently, BIII and BIV were described as human infective subassemblages (Monis et al., 2003; Read et al., 2004). Several studies, however, have demonstrated inconsistent typing results for subassemblages BIII and BIV, showing some isolates are assigned to different subassemblages, referred to as 'assemblage swapping', based on the screening locus (Lebbad et al., 2011; Sprong et al., 2009; Yang et al., 2010). Sequence heterogeneity, detected as double peaks in DNA sequence chromatograms, has also been frequently reported for assemblage B isolates (*c.f.* Caccio and Sprong, 2010). Phylogenetic studies have not supported groupings within assemblage B, which is now considered to consist of numerous genotypes, with no clear subassemblage or genotype sub-structuring (Lebbad et al., 2011; Sprong et al., 2009).

Table 1.2 *Giardia duodenalis* assemblages (Assem) A and B, subassemblages (Subas.) and genotypes and detection among human and non human hosts

A) *Giardia duodenalis* assemblages and host range			**B) *Giardia duodenalis* subassemblages**			**C) *Giardia duodenalis* genotypes**		
Assem.	Host	Reference	Subas.	Host	Reference	Genotype	Host	Reference
A	Human, dog, cat, alpaca,	Monis et al., 2003 Read et al., 2004	AI	Human	Monis et al., 2003	AI-1	Human, cattle, water buffalo, cat, pig	Caccio et al., 2008
	non-human primates					AI-2	Cat	Caccio et al., 2008
						AI-3	Human, animal	Sprong et al., 2009
	pig		AII	Human	Monis et al., 2003	AII-1	Human	Caccio et al., 2008
	ruminents					AII-2	Human, cat	Caccio et al., 2008; Sprong et al., 2009
	rodents							
	marsupials					AII-3	Human	Caccio et al., 2008
						AII-4	Human	Caccio et al., 2008
						AII-5	Human	Caccio et al., 2008
						AII-6	Human	Caccio et al., 2008
						AII-7	Human	Caccio et al., 2008
			AIII	Animals	Monis et al., 2003	AIII-1	Fallow deer, wild boar, cat	Caccio et al., 2008 Sprong et al., 2008
B	Human, dog, marmoset	Monis et al., 2003 Read et al., 2004	BI	Monkey, dog	Monis et al., 2003	-		
	siamang		BII	Monkey	Monis et al., 2003	-		
	non-human primates		BIII	Human	Monis et al., 2003	BIII-1		Caccio et al., 2008 Sprong et al., 2009
	cattle, rabbit beaver		BIV	Human	Monis et al., 2003	BIV-1		Caccio et al., 2008 Sprong et al., 2009

1.4.3 Prevalence of *Giardia* assemblages among human populations

Giardia duodenalis assemblages A and B are globally distributed. Recent international reviews of *Giardia* research have revealed that overall, assemblage B appears to be more common among humans globally; 57% to 60% human cases documented by Caccio and Ryan (2008), and Feng and Xiao (2011) were assemblage B. More assemblage B cases have been identified in both developed countries (58.4% of cases), and developing countries (55% cases)(Feng and Xiao, 2011). Mixed infections, however, with *Giardia* assemblages A and B are higher in developing regions (7.84% cases), compared with developed regions (1.35% cases) (*c.f.* Feng and Xiao, 2011). Less than one percent of human cases are attributed to other *Giardia* assemblages (C-F) (Sprong et al., 2009), and are more commonly reported from developing regions (*c.f.* Feng and Xiao, 2011).

The prevalence of assemblages A and B among humans fluctuates between countries and within global geographic regions (Table 1.3a-c). Several studies in Australia have reported a dominance of assemblage B among human cases which directly contrasts findings in New Zealand, where assemblage A has been more commonly detected (Table 1.3a). Distributions of assemblages A and B also vary between countries of Europe, and between North and South America (Table 1.3b and 1.3c), whilst studies in Africa and the Middle East demonstrate similar frequencies of assemblages A and B among populations (Table 1.3c). Giardiasis transmission is linked to socio-economic, demographic and climatic conditions (Savioli et al., 2006) which vary among neighbouring countries and regions, and most likely alter transmission dynamics of assemblages A and B. Global research findings, however, have been difficult to compare due to diverse screening methodologies, sample sizes, and target populations (Table 1.3a-c), and disease factors that enhance transmission of assemblages A or B are unclear.

1.4.4 Prevalence of subassemblages and genotypes among human populations

Recent analyses of publically available sequence data indicates that regional differences in the frequencies of subassemblages and genotypes occurs among human populations (Sprong et al., 2009). Among human assemblage A cases, subassemblage AII is most commonly identified in Africa, Europe, and the Middle East; subassemblage AI is most commonly identified in Asia and Australia; and similar distributions of AI and AII are detected in Central and North America. Among assemblage B human cases, genotype BIII is most common in Africa, Asia, the Middle East and Central America; whilst North America has more BIV cases, and BIII and BIV occur in equal frequency among populations in Australia and Europe (Sprong et al., 2009). Although subassemblage/genotype substructuing within assemblage B is unclear, these findings would indicate that the underlying mechanisms (heterogeneity, mixed templates) producing different assemblage B genotypes among human populations vary geographically. Furthermore, sequence data analysed by Sprong et al. (2009) contrasts Australian research findings, where subassemblage AII has been most frequently reported from assemblage A cases (Table 1.3a). Further research is required to resolve the distribution of subassemblages and sub-structuring within assemblage B.

1.4.5 Prevalence of *Giardia duodenalis* assemblages A and B among non-human hosts

In non-human hosts (domestic animals and wildlife), up to 20% of infections may be caused by human infective assemblages (*c.f.* Feng and Xiao, 2011). Assemblage A is most commonly detected, and has been reported among domestic animals in Sri Lanka, United Kingdom, China, Australia, and Thailand (Abeywardena et al., 2014; Minette et al., 2014; Traub et al., 2005b; Traub et al., 2009; Wang et al., 2014), and among native animals (kangaroo, wombat) and wild animals (rabbit, deer, waterbird, dog) in Australia (Nolan et al., 2013). Less frequently, assemblage B has been identified in domestic animals, including in Europe, Canada, New Zealand, and Australia (Coklin et al., 2007; Learmonth et al., 2003; Nolan et al., 2010; Sprong et al., 2009). Transmission cycles between humans and animals

are unresolved, but host contact, competitive interactions, and host immunity to different *Giardia* assemblages may determine the frequency human infective assemblages among animal populations (Thompson and Monis, 2012). In high prevalence remote Indigenous communities of Australia, assemblages A and B were detected among humans, whilst among numerous dogs living in the same community, all but one sample were typed as host specific assemblages C and D (Hopkins et al., 1997). Conversely in India, humans and dogs living in the same community were infected with assemblages A and B (Traub et al., 2004). Assemblage B has also been recently reported among marine mammals with close proximity to human settlements in Australia (Delport et al., 2014).

1.4.6 Assemblage linked differences and difficulties in molecular typing

Differences between *Giardia* assemblages A and B, including biotypic and genetic differences have been widely documented, and support the designation of these assemblages as separate species (Adam et al., 2013; Jerlstrom-Huitvist et al., 2010; Xu et al., 2012). In-vitro growth rates, metabolism and establishment of experimental infections differ for assemblages A and B (*c.f.* Caccio et al., 2005; *c.f.* Caccio and Sprong, 2010). Studies of disease manifestations have documented assemblage-linked symptoms among humans living in the same communities (Ignatius et al., 2012; Lebbad et al., 2011; Read et al., 2002). Assemblage B has been associated with increased likelihood of children in Rwanda being underweight (Ignatius et al., 2012), and asymptomatic children attending day care centres in Western Australia (Read et al., 2002). In the same Western Australian study, assemblage A infections were associated with diarrhoea in children (Read et al., 2002). Conversely, assemblage B has been detected among small number of Dutch patients with persistent diarrhoea (Homan and Mank, 2001). Globally, results are conflicting and no clear links have been established between *Giardia* symptoms and infecting assemblage, or differences in the pathogenicity of assemblages A and B (Caccio and Ryan, 2008).

Comparative genomic studies show assemblage A and B genomes are genetically diverse; that approximately 9% of the genomes contain assemblages specific genes, including VSPs, (Jerlstrom-Huitvist et al., 2010; Takumi et al., 2012; Xu et al., 2012). Allelic sequence heterozygosity (ASH), however, within assemblage A is extremely low (<0.01%), and comparatively lower than assemblage B (0.5%) (Morrison et al., 2007). Genetic recombination has been proposed as a mechanism maintaining low ASH (Caccio and Sprong, 2010) and is supported by indirect evidence of recombination associated with *Giardia*'s lifecycle (section 1.2.1). It is unclear if recombination occurs between isolates of a single assemblage (intra- assemblage recombination), or between assemblages (inter-assemblage) and how frequently recombination events would occur (Caccio and Sprong, 2010).

Further evidence is provided by inconsistent typing results/assemblage swapping, which have been reported between assemblages A and B, and between subassemblages AI and AII, AII and BIV, and more frequently among isolates of assemblage B (BIII and BIV) (Amar et al., 2002; Berrili et al., 2012; Lalle et al., 2009; Traub et al., 2004). Inconsistencies are commonly reported from clinical, animal or environmental samples, where isolates have been typed using several *Giardia* loci. These results may represent the presence of mixed templates within the host/environment, and preferential amplification of different assemblages across numerous loci (Caccio and Sprong, 2010). Allelic sequence heterozygosity, however, has been detected from single *Giardia* parasites, from both axenic cultures, and clinical samples (Ankarklev et al., 2012).

Subassemblage/genotype diversity, host range, and distribution are also poorly understood due to complexities associated with their identification (Table 1.3a-c). *Giardia* DNA is difficult to amplify; subtyping genes (section 1.4.1) display various levels of conservation and amplify inconsistently; and application of nomenclature for genetic subgroups has been inconsistently applied, limiting comparisons of results between studies (Caccio and Ryan, 2008; Nantavisai et al., 2007). A new nomenclature for assemblages,

subassemblages and genotypes has been recently proposed (Caccio et al., 2008) (Table 1.2), and molecular methods to accurately characterise and determine intra-assemblage diversity are still emerging. Complexities in subtype identification have limited subtype screenings, which have been less frequently performed for molecular epidemiological studies (Table 1.3a-c).

1.5 Detection and molecular characterisation of *Giardia* assemblages

1.5.1 Routine diagnosis of clinical giardiasis

Routine diagnosis of human giardiasis is performed by immuno-detection assays, and confirmed by visualisation of *Giardia* cysts using microscopic examination of faecal smears (Traub et al., 2005a). *Giardia duodenalis* cysts and trophozoites are morphologically identical, and genetic assemblages cannot be differentiated by microscopy (Thompson et al., 2000; Traub et al., 2005a). Assemblage-specific immunoassays have not been developed, and as such it is not possible to identify *G. duodenalis* assemblages contributing to human infections during routine diagnosis of the disease. Assemblage identification is performed by amplification of *Giardia* DNA and DNA sequence analyses, which is limited to molecular and epidemiological research (Traub et al., 2005a).

1.5.2 Molecular detection of *Giardia* assemblages

Several *Giardia* genes including the *18S rRNA*, and the housekeeping genes; *gdh*, *tpi*, *β-giardin* (section 1.4.1), are commonly used to identify *Giardia* assemblages and subassemblages/genotypes in molecular typing and epidemiological research. Amplification of several other genes and/or gene regions has also been applied (Caccio et al., 2008; Caccio et al., 2010; Nash and Mowatt, 1992; Ng et al., 2005). The high nucleotide sequence conservation of the *18S rRNA* limits detection to *Giardia* assemblages only, whilst the *gdh*, *tpi*, and *β-giardin* are more variable and enable delineation of subassemblages (Lalle et al., 2005; Read et al., 2004; Sulaiman et al., 2003). Identification of genotypes within

subassemblages AI, AII and assemblage B (Table 1.2) requires multilocus genotyping of all three genes (*gdh*, *tpi*, and *β-g*) and construction of multilocus profiles (Caccio et al., 2008; Sprong et al., 2009). PCR amplification across numerous loci, however, remains problematic due to DNA polymerase inhibitors, degraded DNA in clinical and environmental samples, non-specific binding of PCR primers, and intermittent parasite shedding (Caccio & Ryan, 2008; Nantavisai et al., 2007; Traub et al., 2005a). Many studies have applied cyst concentration methods prior to DNA extraction (Pelayo et al., 2008; van Keulen et al., 2002; Winkworth et al., 2008), and performed PCR amplification using highly sensitive DNA polymerases (Asher et al., 2012a; Sulaiman et al., 2003), which is lengthy and/or expensive. To provide rapid and inexpensive sequence analyses of PCR products, tools such as restriction fragment length polymorphism (RFLP) and real-time assemblage specific PCRs have been developed as alternatives to DNA sequencing (Ng et al., 2005; Read et al., 2004). Amplification of numerous *Giardia* loci (two or more) for accurate molecular typing remains difficult and expensive for large scale studies (Asher et al., 2012a; Breathnach et al., 2010).

1.5.3. Emerging tools for *Giardia* identification: terminal-restriction fragment length polymorphism and next generation sequencing

Emerging tools for *Giardia* identification include highly sensitive screening methods which have previously been applied to detect microbial diversity (Boers et al., 2012; Marsh, 1999; Pilloni et al., 2012). Fluorescent detection by terminal-restriction fragment length polymorphism (T-RFLP) and clonal amplification of PCR products by next generation sequencing are highly sensitive, and are increasingly used in diverse research applications. These methods enable detection of microbial populations within individual hosts, and across populations when applied in an epidemiological context (Marsh, 1999; Waldron et al., 2009).

Terminal-RFLP is a fluorescence based method which has been developed from RFLP and relies upon the generation of a fluorescent terminal-restriction fragment (T-RF) (Marsh,

1999). A terminal-RF is produced by fluorescent PCR amplification and enzymatic digestion of the PCR product, which is detected by an automated DNA sequencer (Marsh, 1999). Fluorescent detection is highly sensitive; sequence polymorphisms among PCR products generate T-RFs of different nucleotide lengths, and enable detection of population diversity (Marsh, 1999; Waldron et al., 2009). Terminal-RFLP has been recently applied to typing of *Cryptosporidium* isolates (Waldron et al., 2009), and has been shown to be a rapid, cost-effective, and sensitive alternative to DNA sequencing and RFLP (Waldron et al., 2009).

Next generation sequencing is a high-through put alternative to DNA (Sanger) sequencing that enables simultaneous screening of individual amplicons across multiple samples and loci (Pilloni et al., 2012), without the need for cloning. Recent improvements in sequence read lengths and the ability to barcode sequences from individual samples has provided a platform to develop high-throughput multilocus genotyping (Boers et al. 2012). Multilocus genotyping has recently been applied to bacterial (Boers et al., 2012), fungal (Chen, 2015) and protozoan (Grinberg et al., 2013) species, but has not been developed for *Giardia*.

1.6 Research objectives:

1.6.1 Giardiasis in Australia

In Australia giardiasis is a notifiable disease in several jurisdictions including, New South Wales (NSW Health Notifiable Conditions Information Management System, 2015), Australian Capital Territory (ACT Government Health, 2015) and Tasmania (Department of Health and Human Services, Tasmania, 2010). The overall prevalence of human giardiasis observed in Australia is estimated between 2% and 7% (Thompson and Monis, 2012), however, prevalence fluctuates across communities. In New South Wales a large number of sporadic cases are reported annually, over 2900 cases were reported to NSW Health in 2014 (NSW Health Notifiable Conditions Information Management System, 2015). Disease

surveillance between 2002 and 2012 shows increasing numbers of annual giardiasis cases in NSW (13.0 to 27.6 per 100 000 population) and annual incidence rates vary between 5.0 to 54.1 (per 100 000 population), across both rural and urban NSW communities (NSW Department of Health, 2013, 2011).

In remote Indigenous communities of Australia giardiasis prevalence is high, ranging from 15% to 36% (Thompson et al., 2001; Jones, 1980). Prevalence rates among children living in remote Indigenous communities range between 32% to 65%, and are among the highest rates in the world (Meloni et al., 1993; Reynoldson et al., 1998; Thompson et al., 2001). The prevalence of *Giardia* among remote Indigenous populations is poorly understood. These communities experience high rates of gastrointestinal disease and chronic morbidity, and the contribution of *Giardia* infections to these manifestations are unclear (McDonald et al., 2001; Holt et al., 2010).

1.6.2 Research aims

Giardiasis is an emerging disease concern in New South Wales and in remote Indigenous communities of Australia. Genetic assemblages/subassemblages contributing to human giardiasis in Australians are unknown, and associations between genetic types and risk factors for infection, including demographic groups and spatio-temporal links are not documented. Public health service planning and disease management requires a comprehensive understanding of parasite ecology to evaluate public health risks associated with different types of *Giardia* infections. Emerging molecular screening tools such as T-RFLP and next generation sequencing are potential avenues for developing streamlined and cost effective multilocus genotyping tools, to examine the diversity of *Giardia duodenalis* in low and high prevalence communities, and provide tools for large scale sample screenings. Analyses of *G. duodenalis* in high prevalence communities is particularly required, to examine the extent of genetic diversity within assemblages A and B, in environments with

increased competition and selective pressures (Thompson and Monis, 2012). Comparison of data from low and high prevalence communities in Australia will enable greater understanding of *Giardia* transmission and factors contributing to the continued high prevalence of giardiasis in remote Indigenous communities, and the increasing prevalence among New South Wales populations.

The overall aim of this study is to improve our understanding of *G. duodenalis* epidemiology in the Australian context, and to develop molecular tools that provide a platform to improve molecular epidemiological research. The overall aim of the project will be addressed through the following specific aims:

1. Determine the prevalence of *Giardia* assemblages/subassemblages in low prevalence communities of New South Wales, Australia
2. Examine risk factors for disease in New South Wales by assessing links between genetic assemblages identified among human and animal hosts, and among patient demographic groups and spatio-temporal distributions in New South Wales
3. Determine the prevalence of *Giardia* among children in a remote Indigenous community of Australia, and examine the contribution of *Giardia* assemblages among participant age and gender groups over time
4. Develop and evaluate new molecular screening tools for *Giardia* identification, and examine the efficiency of these tools to existing protocols
5. Examine the extent of genetic diversity within *Giardia duodenalis* by applying highly sensitive molecular screening tools to *Giardia* identification, in communities with low and high prevalence rates

1.7 Positioning of the research chapters

Chapter 2: Giardiasis in NSW: Identification of *Giardia* subtypes contributing to human and cattle disease, and an epidemiological assessment of human cases

Increasing incidence of sporadic giardiasis cases in NSW has highlighted this disease as an emerging public health concern. This chapter presents the spatial and temporal distributions of *Giardia* assemblages A and B among human cases in NSW local health districts, and among patient age and gender groups. The study demonstrates localised dispersal patterns of assemblages A and B, identified by geographical mapping of human cases. Our results show similar frequencies of assemblages A and B throughout NSW, but indicate that risk factors for assemblage A or B infection may vary within communities and across time. Additionally, human infective assemblages were detected among cattle in NSW, providing evidence of transmission between human and animal hosts.

Addresses objectives 1 and 2

Chapter 3: Rapid identification of *Giardia duodenalis* assemblages in NSW using terminal-restriction fragment length polymorphism

Identification of *Giardia* assemblages/subassemblages requires molecular analyses and screenings are limited by the expense of DNA sequencing or insensitivity of restriction fragment length polymorphism (RFLP). This study describes the development and application of a fluorescent method, terminal-RFLP, for detection of *Giardia* at the diagnostic glutamate dehydrogenase gene. Using this method, a dominance of assemblage B cases were detected among humans in NSW, consistent with findings of chapter 2. This study is the first application of T-RFLP to *Giardia* and provides a rapid, sensitive, and simple method for high throughput detection of human infective assemblages and subassemblages.

Addresses objectives 1, 4 and 5

Chapter 4: Distribution of *Giardia duodenalis* assemblages A and B among children living in a remote Indigenous community of the Northern Territory

High rates of giardiasis persist in remote Indigenous communities in Australia, and children are most at risk from chronic disease outcomes. In this study we demonstrate high carriage rates (66%) of *Giardia* among children under 15 years in a remote community of the Northern Territory. Both assemblages A and B were detected in the community over an 18 month period. The majority of cases (75%) were assemblage B and a high proportion (28%) were mixed templates. Over 50% of cases were not detectable by microscopy, and were identified by 18S rRNA PCR. This study applies a combined DNA sequencing and T-RFLP approach, and demonstrates the need for highly sensitive screening methods in high prevalence communities. The high proportion of sub microscopic and mixed template cases suggests unique disease dynamics in this high prevalence community.

Addresses objectives 3, 4 and 5

Chapter 5: Multilocus next generation sequence typing of *Giardia* from children living in a remote Indigenous community of Australia

Despite advances in molecular technologies, the genetic diversity *Giardia duodenalis* remains unclear due to complexities in subtype identification. Next generation sequencing (NGS) has emerged as an affordable tool for multilocus sequence typing and for examining genetic variation. This study describes the development and application of NGS to *Giardia* typing across four diagnostic loci, and examines the genetic diversity of *Giardia* in a high prevalence remote Indigenous community. The study demonstrates detection of mixed assemblage A and B populations from a single sample, and identification of eight multilocus genotypes from ten children in the community. Application of multilocus NGS to *Giardia* typing in this study should facilitate further development of this tool, for *Giardia* epidemiological research.

Addresses objectives 4 and 5

1.8 Tables 1.3a to 1.3c

Table 1.3a Distribution of *Giardia duodenalis* assemblages A and B among human populations in countries of the Australia/Pacific and Asian region, including target population, sample size, and detection methods for *Giardia* assemblages and subtypes

Region **& country** Study description	***Giardia*** **detection**			***G. duodenalis*** **assemblages**			
	Detection methods	Loci	Positivity	A	B	Subtyping	Reference
Australia & Pacific							
Australia: children <5yrs, Day care centres; *n* = 353	Microscopy PCR, DNA-sequencing	*18S rRNA*	27/353	7/23	16/23		Read et al., 2002
Australia: remote Indigenous community; <1 to >80yrs; *n* = 19	Microscopy PCR, Southern Blotting, DNA- sequencing	*18S rRNA*	12/19	1/12	11/12		Hopkins et al., 1997
Australia: human samples *n* = 8	Microscopy PCR-RFLP, DNA sequencing	*18S rRNA* *gdh*	all cases	2/8	6/8	AI: 1/8 AII: 1/8 BIV: 6/8	Read et al., 2004
Australia: sporadic cases <5 to >70yrs; *n* = 124	Microscopy confirmed cases PCR sequencing	*18S rRNA* *gdh*	all cases	31/124	93/124	AII: 27/109 BIII: 44/109 BIV: 38/109	Yang et al., 2010
New Zealand: patients with active infections; *n* = 30	Cyst isolation & DNA extraction, PCR sequencing	*β-giardin*	all cases	23/30	7/30		Winkworth et al., 2008
Asia (Central, South, South/East)							
Bangladesh: patients with diarrhoea; *n* = 35	Microscopy, antigen detection assemblage specific PCR	*Scorpion probe assay*	all cases	3/35	32/35		Ng et al., 2005
Cambodia: PCR positive human cases*; n* = 40	PCR-DNA sequencing	*18S rRNA, tpi, gdh*	all cases	11/40	29/40	AII: 11/40 BIII: 29/40	Inpankaew et al., 2014
India: sporadic, *n* = 10	Microscopy, PCR-DNA sequencing	*18S rRNA* *tpi*	all cases	0/10	10/10		Sulaiman et al., 2003
India (Assam): human faecal samples; *n* = 328	Microscopy, PCR-DNA sequencing	*18S rRNA, ef1-α, tpi*	29/328	5/24	2/24	Mixed A, B, C, D typing results	Traub et al., 2004

Table 1.3a Distribution of *Giardia duodenalis* assemblages A and B among human populations in countries of the Australia/Pacific and Asian region, including target population, sample size, and detection methods for *Giardia* assemblages and subtypes

Region **& country** Study description	***Giardia*** **detection**			***G. duodenalis*** **assemblages**			
	Detection methods	Loci	Positivity	A	B	Subtyping	Reference
Asia (Central, South, South East) continued							
Malaysia: *n* = 154	Microscopy, PCR, DNA sequencing	*tpi, β-giardin, gdh*	154/1330	69/138	69/138		Choy et al., 2014
Nepal: *n* = 1096	Microscopy, qPCR, RFLP	*18S rRNA*	45/1096	7/35	26/35	Mixed A/B: 2/35	Singh et al., 2009
*Thailand (Temple communities);*human faecal samples *n* = 204. Eighty-five selected for analysis	Microscopy, IF antibody, ELISA, PCR-DNA sequencing	*18S rRNA*	78/85	26/35	1/35	Mixed A/B: 2/35 Mixed, assemblages not reported: 6/35	Traub et al., 2009

Table 1.3b Distribution of *Giardia duodenalis* assemblages A and B among human populations in countries of the American and European region, including target population, sample size, and detection methods for *Giardia* assemblages and subtypes

Region **& country** Study description	***Giardia*** **detection**			***G. duodenalis*** **assemblages**			
	Detection methods	Loci	Positivity	A	B	Subtyping	Reference
Americas (North, Central, South)							
United States: foodborne outbreak *n* = 2	Microscopy PCR DNA sequencing	*tpi* *β-giardin*	all cases	0/2	2/2		Sulaiman et al., 2003
United States: sporadic cases *n* = 14	PCR cloning Sequencing/RFLP	*18S rRNA*	all cases	14/14	0/14		van Keulen et al., 2002
Canada: outbreak *n* = 12	PCR cloning DNA sequencing/RFLP	*18S rRNA*	6/10	6/6	0/6		van Keulen et al., 2002
Cuba: children n = 95	Cyst isolation, PCR	*β-giardin, gdh*	20/95	9/20	11/20		Pelayo et al., 2008
Argentina: *Giardia* positive cases; 1 – 33 yrs; *n* = 60	Microscopy PCR RFLP	*tpi*	43/60	3/43	40/43	AII: 3/43	Minvielle et al., 2008
Brazil: patients with diarrhoea *n* = 37	Microscopy cyst concentration PCR	*gdh*	all cases	29/37	8/37	AII: 29/37	Souza et al., 2007
Peru: sporadic cases; n = 25	Micrscopy, PCR, DNA-sequencing	*18S rRNA* *tpi*	all cases	6/25	19/25		Sulaiman et al., 2003
Europe							
Belgium: symptomatic patients n = 373, and positive *Giardia* cases *n* = 59	Immunofluorescene, PCR for positive cases, DNA-sequencing	*tpi, β-giardin* *gdh*	15/373 (4); + cases (n = 59)	18/72	54/72	Mixed A/B: 2/72 AII: 8/72 BIV: 22/72 BIII: 6/72	Geurden et al., 2009
France: sporadic, *n* = 26	Microscopy, PCR of positive cases, RFLP	*tpi, gdh*	all cases	9/25	16/25	Assemblage swapping detected	Bertrand et al., 2005
Italy: *n* = 28	PCR, DNA sequencing	*18S rRNA* *tpi, β-g, gdh*		28/61	33/61		Caccio et al., 2008

Table 1.3c Distribution of *Giardia duodenalis* assemblages A and B among human populations in countries of Europe, Africa and the Middle East, including target population, sample size, and detection methods for *Giardia* assemblages and subtypes

***Region*/country** Study description	***Giardia* detection**			***G. duodenalis* assemblages**			
	Detection methods	Loci	Positivity	A	B	Subtyping	Reference
Europe continued							
Sweden; *Giardia*+ cases (175 sympt. cases), *n* = 214	Microscopy, assemblage specific PCR, DNA sequencing, RFLP	*tpi, β-g, gdh*	207/214	73/207	128/207	Mixed A/B: 6/207	Lebbad et al., 2011
United Kingdom; sporadic cases, *n* = 35. Outbreak cases, *n* = 24	Microscopy, assemblage specific PCR, DNA sequencing, RFLP	*tpi*	33/35	sporadic 9/35 outbreak 0/21	sporadic 21/35 outbreak 21/21	Mixed AII/B: 3/35	Amar et al., 2002
United Kingdom: positive cases, *n* = 819	PCR,DNA sequencing	*18S rRNA* *tpi*	217 screened	48/199	145/199	Mixed A/B: 6/199	Breathnach et al., 2010
Africa & the Middle East							
Algeria: Sahrawi children, 8 – 13 yrs, *n* = 120	Microscopy, PCR, DNA sequencing, RFLP	*tpi, gdh*	41/120	12/30 (*tpi*)	18/30 (*tpi*)	AI: 6/30 AII: 6/30 Mixed A/B: 2/30	Lalle et al., 2009
Cote d'Ivoire: 2 – 63yrs, n = 307	Microscopy, PCR, DNA sequencing, Real time PCR	*18S rRNA, β-g*	61/307	21/61	36/61	Mixed A/B: 4/61	Berrili et al., 2012
Egypt: *n* = 52	Cyst concentration PCR-DNA sequencing	*tpi*	18/52	1/18	14/18	Mixed: 1/18 Assemblage E: 2/18	Foronda et al., 2008
Rwanda: children <5 yrs, *n* = 583	Microscopy, PCR, DNA sequencing	*tpi*	368/583	28/208	179/208	Mixed: 1/208	Ignatius et al., 2012
United Arab Emirates: healthy participants, *n* = 111	Microscopy, PCR-DNA sequencing	*18S rRNA* *tpi*	67/111	21/67	19/67	Mixed: 6/67 Unknown: 21/67	ElBakri et al., 2014

2: Giardiasis in NSW: Identification of *Giardia* subtypes contributing to human and cattle disease, and an epidemiological assessment of human cases

A. J. Asher [a, *], G. Hose [a], M. L. Power [a]

[a] *Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia*

Keywords: *Giardia duodenalis*, assemblages, humans, cattle, New South Wales

2.1 Abstract

Giardiasis is a communicable gastrointestinal disease caused by the protozoan parasite *Giardia duodenalis*. Two genetic assemblages, A and B, infect humans, domestic animals and wildlife. In New South Wales, over two-thousand cases of sporadic human giardiasis cases are notified annually, but disease sources and links between sporadic cases are unknown. This study investigated *Giardia* assemblages contributing to human and cattle cases in NSW, and examined demographic, spatial, and temporal distributions of NSW human infections and *Giardia* assemblages. Human samples were collected from NSW pathology companies between 2010 and 2013. Successful genotyping of 165 *Giardia* positive human faecal samples, by PCR-restriction fragment length polymorphism of the glutamate dehydrogenase (*gdh*) gene identified *Giardia* assemblage B as the most common (86%) cause of human infection. Among assemblage B cases, 97.9% were typed as subassemblage BIV. All assemblage A cases (12.7%) were subassemblage AII, suggesting anthroponotic transmission of these cases. Similar frequencies of assemblages A (<17%) and B (>83%) were detected among patient groups, however, assemblage B prevalence was higher among 0 – 39 year olds. Within NSW Local Health Districts, geographical mapping showed seasonal shifts in the dispersal of human assemblage A across NSW communities. From a DNA library of NSW cattle faecal DNA samples, approximately 37% (n = 64/175) were positive for *Giardia* by PCR amplification of the small subunit ribosomal rRNA and/or the *gdh*. These samples were identified as *Giardia* assemblage E (69%) or B (31%). The results of this study indicate *Giardia* assemblage B is highly prevalent among humans in NSW and is a persistent of sporadic human giardiasis. Detection of assemblage B among cattle samples suggests potential for cross-species transmission. Further research is required to elucidate transmission cycles of assemblage B, to evaluate risk factors associated with seasonal and geographic dispersal patterns of

assemblages A and B, and to determine the potential for disease transmission between humans and farm animals in NSW.

2.2 Introduction

Giardiasis is one of the most common communicable gastrointestinal diseases of humans and domestic animals worldwide (Thompson and Monis, 2012). The aetiological agent, *Giardia duodenalis*, is a protozoan parasite of mammals and is transmitted via the faecal-oral route (Caccio and Ryan, 2008). Acute disease typically produces diarrhoea, vomiting, nausea and fatigue but symptoms can be intermittent, progress to chronic disease and wasting, or infection can remain asymptomatic (Caccio and Ryan, 2008; Roxstrom-Lindquist et al., 2006). Globally, an estimated 2.8×10^8 humans are affected by symptomatic giardiasis each year, and the disease burden and economic losses associated with infected livestock remains unclear (Caccio and Ryan, 2008; Feng and Xiao, 2011; Lane and Lloyd, 2002). Disease sources include consumption of contaminated water and contact with faeces of an infected host (Ng et al., 2011).

In Australia, human giardiasis is largely a sporadic disease and overall prevalence is estimated between 2 – 7% (Thompson and Monis, 2012). In New South Wales, a large number of sporadic human cases are notified annually, over two-thousand human cases have been reported each year since 2009 (NSW Health Notifiable Conditions Information Management System, 2015). Giardiasis was first listed as a notifiable disease in NSW in August 1998 following the Sydney Water Crisis, in which *Giardia* and *Cryptosporidium* were identified as contaminating Sydney's water supply (NSW Department of Health, 1998). Disease surveillance since 1998 shows a steady increase in the number of human giardiasis cases reported each year in NSW (NSW Department of Health, 2010; NSW Department of Health, 2011). Between 2002 and 2012, the incidence of giardiasis notifications per 100 000 population has increased from 13.0 to 27.6 (NSW Department of

Health, 2011; NSW Department of Health, 2013) and has highlighted *Giardia* as an emerging public health concern.

Across NSW giardiasis notifications fluctuate across geographic, temporal and demographic scales. Annual notifications peak between January and April each year and are highest among children aged 0 to 4 years, and adults aged 30 to 39 years (NSW Health Notifiable Conditions Information Management System, 2015). Incidence rates (per 100 000 population) also vary across communities, ranging 5.0 to 54.1; and high rates (~43.0) are reported from both rural and urban health districts (NSW Department of Health, 2011). In New South Wales, distributions of human giardiasis among NSW residents are similar to patterns observed in developing nations (Centres for Disease Control and Prevention, 2012) but sources of sporadic infection in NSW communities are largely unknown. It is unclear if associations exist between high risk age groups, and whether fluctuations across NSW communities reflect different disease dynamics. It is also unclear whether infected animals pose a significant risk to human health.

Recent advances in molecular technologies have improved *G. duodenalis* identification and genetic typing is now used to differentiate isolates, identify disease sources, and assess host infectivity (Feng and Xiao, 2011). Eight different *G. duodenalis* assemblages (A to H) infect mammalian host groups, and are defined as clusters of genetically related isolates (Caccio and Ryan, 2008; Lasek-Nesselquist et al., 2010; Thompson and Monis, 2012). Humans are only infected by assemblages A and B which are zoonotic, and most likely represent two separate species of *Giardia* (Thompson and Monis, 2012). Livestock, companion animals and wildlife are infected by assemblages A and B, and also host specific assemblages C to H (canine assemblages C and D; bovine assemblage E; feline assemblage F; rodent assemblage G; piniped assemblage H) (Caccio and Ryan, 2008; Lasek-Nesselquist et al., 2010). Four subtypes/subassemblages (AI – AIV) have been defined within assemblage A (*c.f.* Ryan and Caccio, 2013).

Subassemblage AII is largely restricted to humans, whilst subassemblage AI has been identified in both human and non-human hosts (*c.f.* Ryan and Caccio, 2013; Yang et al., 2010). Assemblage B consists of multiple genotypes and although two subassemblages have been identified (BIII and BIV), subassemblage structuring remains unclear (*c.f.* Feng and Xiao, 2011).

Identification of *Giardia* assemblages/subassemblages is currently limited to epidemiological research (Caccio and Ryan, 2008; Traub et al., 2005a). Routine pathology screenings determine only the presence/absence of *Giardia* by various methods such as microscopic examination of faecal smears, or immuno-assays which detect *Giardia* antigens from stool samples (Traub et al., 2005a). In the present study we aimed to identify *Giardia* assemblages/subassemblages causing disease among humans and animals in NSW; and to evaluate links between *Giardia* subtypes and patient age groups, geographic location and seasonal distributions, which has not previously been assessed in NSW. In performing these analyses we aimed to determine whether fluctuations in giardiasis notifications across NSW are associated with different types of *Giardia*. This information is essential for informing disease management and control strategies and understanding the continuing rise in giardiasis cases in NSW.

2.3 Materials and Methods

2.3.1 Ethics statement

All work performed for this study including collection of faecal samples, laboratory sample numbers, and patient information received ethical approval from the Macquarie University Research Ethics Committee (Reference: 5201100403) and the NSW Population and Health Services Human Research Ethics Committee (Reference: HREC/12/CIPHS/87).

2.3.2 Sources of human faecal samples and patient information, and DNA from cattle

Human faecal samples were obtained from pathology companies in NSW between 2010 and 2013. The samples were identified as being positive for *Giardia* by pathology companies using a Remel ProSpect *Giardia*/*Cryptosporidium* microplate immunoassay (Thermo Fisher Scientific, 2010) and by visualising *Giardia* cysts in faecal smears using microscopy. Samples were accompanied with a laboratory faecal sample number to enable retrospective collection of patient data from the NSW Notifiable Diseases Database (NDD). At the time of collection, no other information, other than the laboratory number was provided. Samples were transported to Macquarie University, and to enable processing outside of a PC2 laboratory, all samples were irradiated using $Cobalt^{60}$ exposure (1500 Ci, 30 min). Samples were assigned a Macquarie University *Giardia* (MQG) number and stored at 4°C prior to DNA extraction.

Patient data (age, gender and postcode) were obtained for 243 faecal samples collected between 2010 and 2013. Patient data were provided by a Data Custodian at NSW Health, and only individual cases were included. Multiple laboratory samples/sample numbers collected for a single patient at the same time were excluded from the analyses.

A DNA library representing 205 NSW cattle faecal DNA samples (Waldron et al., 2011a) were provided for the study, and were screened for *Giardia* to investigate *Giardia* prevalence and to determine *Giardia* assemblages among any positive samples. Cattle samples were collected as part of the previous study, from beef and dairy farms in NSW locations of Camden, Richmond and Wagga Wagga, and included samples from both adult and juvenile cattle (Waldron et al., 2011a).

2.3.3 DNA extraction PCR amplification of the *gdh* gene for human and animal samples

DNA extraction for human faecal samples (n = 243), was performed using approximately 150 mg of faecal material and the ISOLATE Fecal DNA Kit (Bioline,

Sydney, Australia). DNA extracted from NSW cattle faecal samples (n = 175) as part of a previous study (Waldron et al., 2011a) was made available for the present study.

To identify *Giardia* assemblages/subassemblages, all human and cattle DNA samples were screened by PCR amplification of a partial fragment (430 bp) of the *Giardia* glutamate dehydrogenase gene (*gdh*). All *gdh* PCRs were performed following the previously described semi-nested PCR protocol by Read et al. (2004), using PCR primers GDHeF/GDHiR for the primary reaction, and GDHiF/GDHiR for the secondary reaction. Primary and secondary reactions were prepared in 25 μL volumes with 1 μL of DNA template and using the GC-RICH Kit, dNTPack (Roche Diagnostics, IN, USA) as previously described (Asher et al., 2012), or using *Tth plus* DNA polymerase (Fisher Biotec, Australia). *Tth plus* DNA polymerase reactions were prepared using 1 unit of DNA polymerase, 200 nM of dNTPs, and 20 pmol/uL of PCR primers for primary and secondary reactions. All *gdh* reactions were performed in an Eppendorf Mastercycler (Eppendorf, North Ryde, Australia), following the thermocycling conditions described by Read et al. (2004).

DNA samples from cattle faeces (n = 175) were screened at two *Giardia* loci. All samples were screened at the *Giardia* small subunit ribosomal RNA (*18S rRNA*) and the *gdh* loci to confirm the presence of *Giardia*, and to perform subassemblage analysis using *gdh* PCR-RFLP. PCR conditions for the *18S rRNA* were performed following the nested PCR protocol described by Hopkins et al. (1997) and Read et al. (2002), using primers RH11/RH4LM (primary reaction) and GiAR18SeR/GiAR18SiR (secondary reaction). Primary and secondary reactions were prepared in 25 µL volumes with *Tth plus* DNA polymerase (Fisher Biotec, Australia) using the PCR chemistry described above.

All PCR products were analysed by agarose gel electrophoresis (2% w/v, 100 V for 30 min) in TBE (Tris, EDTA, boric acid, pH 8) with gel staining using SYBR safe (Invitrogen, Mulgrave, Australia). The bp length of *gdh* and *18S rRNA* PCR products were

determined using HyperLadder^{TM} II DNA marker (Bioline, Sydney, Australia), all products were visualised under UV-light. Positive *gdh* amplicons were purified using a QIAquick PCR purification kit (Qiagen, Melbourne, Australia) following the manufacturer's instructions.

2.3.4 Identification of *Giardia* assemblages/subassemblages by *gdh*-restriction fragment length polymorphism

Giardia assemblages/subassemblages were identified from positive human and cattle DNA samples by restriction fragment length polymorphism of purified *gdh* PCR products, following the RFLP protocol described by Read et al. (2004). Purified products were digested separately with two restriction endonucleases, *Nla IV* (New England Biolabs) and *Rsa I* (Roche Diagnostics, IN, USA). Digestion reactions were performed at 37°C for 3 hours, using 1 unit of endonuclease and 10 µL of purified PCR product. Following digestion, restriction fragments were resolved by agarose gel electrophoresis (100 V for 50 min) as previously described, but using 3.5% w/v agarose and 10 µL of digested product.

2.3.5 Statistical analyses and mapping of spatial data

The frequencies of *Giardia* assemblages among patient age and gender groups were evaluated by chi-squared statistical analyses. Geographical mapping of NSW human cases were performed in ArcMap 10.2.2 (ESRI Inc, 2014) using patient postcode data, and using the online basemap; Local Government Area NSW 2011 (Credits; Australian Bureau of Statistics (ABS), Commonwealth of Australia). The basemap was created by Esri using ArcGIS® software and are used herein under license. New South Wales is divided into 8 metropolitan Local Health Districts (LHDs) and 7 rural/regional LHDs (NSW Department of Health, 2013). Patient postcode data was used to match NSW cases included in this

study to NSW LHDs, to determine the frequencies of *Giardia* cases and *Giardia* assemblages for each district.

2.4 Results

2.4.1 PCR screenings for human samples and identification of *Giardia* assemblages/ subassemblages by RFLP

DNA extraction was successfully performed for 241/243 human faecal samples. Two pathology samples contained no faecal material and could not be analysed further. PCR amplification of the *gdh* gene was successful for 169/241 (70.1%) faecal DNA samples. Fifty-four samples (22.4%) were PCR negative, whilst for an additional 18 samples, amplification produced large PCR smears, and positivity could not be determined. Analyses of all *gdh* positive PCR products by RFLP was successful for 165/169 (97.6%) human faecal samples. Two samples could not be visualised after digestion and a two further samples produced multiple banding patterns, indicating multiple restriction fragments. Multiple banding patterns were inconsistent with typing profiles, and these samples were removed from further analyses.

Both *Giardia duodenalis* assemblages A and B were identified from restriction profiles of *gdh* products. For all samples, 21/165 (12.7%) were assemblage A, whilst 142/165 (86.1%) were assemblage B. Two samples (1.2%) were typed as mixed infections of assemblages A and B. The difference in assemblage A and B infections (excluding mixed template infections) was statistically significant (Chi-square = 88.3, df = 1, p = <0.00). Differences in the frequencies of human assemblages A and B in NSW, when compared to frequencies reported from Western Australia (assemblage A, 25%; assemblage B, 75%) (Yang et al., 2010), were statistically significant (chi-sq = 13.54, df = 1, p = <0.00).

Analyses of restriction profiles indicated that all assemblage A infections were subassemblages AII. Subassemblage AI was not identified. Among assemblage B infections, 139/142 (97.9%) were subassemblage BIV, whilst profiles for three samples were consistent with mixed templates of subassemblages BIII and BIV.

2.4.2 Spatial distribution of sporadic human cases and *Giardia* assemblages A and B in NSW

Geographical mapping showed sporadic giardiasis cases (n = 243) were distributed across a large area of NSW between 2010 and 2013. Locations included metropolitan areas of Sydney and the Blue Mountains, and regional inland and coastal centres of NSW (Figure 2.1a). Many cases, however, were from regional areas, in particular the Hunter region, including Newcastle, Lake Macquarie and surrounding areas of Port Stephens, Maitland, and the Upper Hunter (Figure 2.1a). Allocation of postcode data to Local Health Districts supported this; a total of 198/243 (81.5%) samples were from rural/regional LHDs. The Hunter New England LHD accounted for 49.4% (120/243) of all samples, and of these 91/120 (75.8%) were from the Hunter region. A total of 45/243 (18.5%) were from metropolitan LHDs of Western Sydney (33/45, 73.3%) and Nepean Blue Mountains (9/45, 20.0%). A small number of samples were from the Central Coast (n = 1), Illawarra (n = 2), and Sydney (n = 1), whilst there were no samples from the Far West, Northern Sydney, South-Eastern Sydney, or South-Western Sydney LHDs.

Spatial distributions of assemblage A and B cases (n = 165) showed that both assemblages were widely dispersed across NSW, and were not geographically isolated (Table 2.1). Approximately 86% of all assemblage A infections occurred at the same location as an assemblage B infection. Both assemblages A and B were identified in all rural/regional health districts included in the study, and from the metropolitan health districts of Western Sydney and Nepean Blue Mountains. Samples from the Central Coast

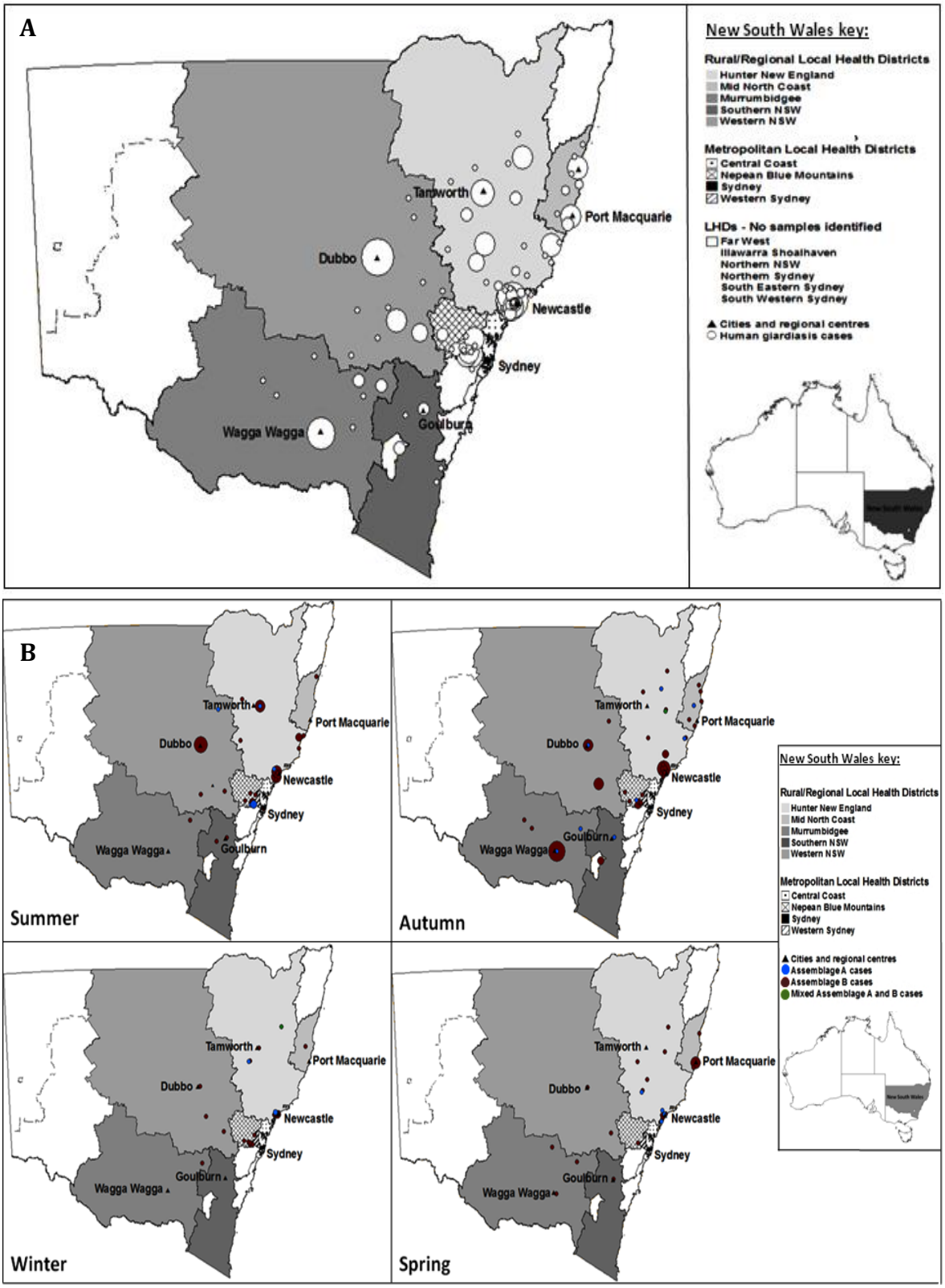

Figure 2.1 Geographic distributions of giardiasis cases and seasonal dispersal of *Giardia* assemblages A and B across NSW, and among NSW local health districts. The figure shows **a)** geographic distributions of positive cases (*n* = 243) occurring in NSW between 2010 and 2013; and **b)** spatio-temoral distributions of assemblage A and B cases (*n* = 165) occurring in NSW between 2010 and 2013.

and Sydney LHDs (n = 2) were both identified as assemblage B infections, whilst genotyping for Illawarra samples (n = 2) was unsuccessful.

The majority of cases detected in NSW LHDs were caused by *Giardia* assemblage B, which was identified in 83.7% to 92.3% of cases per district (Table 2.1). Assemblage A infections were identified in 7.7% - 16.6% of cases per district (Table 2.1). Differences in the frequencies of assemblage A and B infections were significant for Hunter New England, Mid North Coast, Murrumbidgee, Western NSW, and Western Sydney LHDs. (Table 2.1). Sample sizes for other LHDs were too small to determine if differences were significant. Comparison of rural/regional LHDs (assemblage A, 17/134, 12.7%; assemblage B, 115/134, 85.8%) to metropolitan LHDs (assemblage A, 4/31, 12.9%; assemblage B, 27/31, 87.1%) produced similar results, for both assemblages and for each region.

2.4.3 Distribution of sporadic human cases and *Giardia* assemblages A and B among NSW residents by patient age and gender

Patient ages were obtained for 242/243 cases, and patients were aged between 0 years to over 70 years. The largest number of cases were collected from patients aged 0 - 4 years (48/242, 19.8%) and 5 - 9 years (26/242, 10.7%), followed by 30 - 34 years (24/242, 9.9%) and 35 - 39 years (22/242, 9.1%). The total number of males (122/243, 50.2%) and females (121/243, 49.8%) included in the study was approximately equal (Table 2.2).

For the 165 cases genotyped, assemblage B infections were the most common cause of infection among all age groups (Table 2.2). Assemblage A was also widely distributed among the age groups; however, it was not identified among patients aged between 5 years and 19 years (Table 2.2). Within male and female groups, significant differences in the assemblage A and B frequencies were also observed. For males, assemblages A and B were identified in 13/88 (14.8%), and 74/88 (84.1%) cases respectively (Table 2.2).

Among females, 8/77 (10.4%) cases were assemblage A, whilst 68/77 (88.3%) cases were assemblage B (Table 2.2). Mixed assemblage cases (A and B, *n* = 2) were identified from one male and one female. Among high risk age groups assemblage B contributed to 93.9% and 91.6% of cases among 0 – 9 year olds, and 30 – 39 year olds respectively, and between these ages (10 – 29 years) 90.3% of cases were assemblage B. Assemblage A cases were more frequently detected among patients aged 40 to over 70 years (Table 2.2). The difference in assemblage A and B frequencies between patients aged 0 – 39 years, and 40 – 70+ years was statistically significant (Chi-square = 18.9, df = 1, p = <0.00).

2.4.4 Seasonal distribution of assemblages A and B among NSW human cases

Both *Giardia* assemblages A and B were detected among NSW residents during the seasons of Summer, Autumn, Winter, and Spring (Table 2.3). Frequencies of assemblage A and B infections were similar to total cases described above, and the majority of cases for each season were assemblage B (Table 2.3). Assemblage B cases were widely dispersed throughout NSW during each season, and were identified in all LHDs (Table 2.3). Assemblage A was identified from the majority of LHDs during Autumn and was distributed across a large area of NSW during this season (Figure 2.1b). During Summer, assemblage A was only identified among cases from the Hunter New England, Western NSW and Western Sydney LHDs, whilst during Winter and Spring, assemblage A was only detected in the Hunter New England LHD (Figure 2.1b). Seasonal distributions showed that cases typed as mixed assemblage A and B (*n* = 2), and cases typed as mixed subassemblages BIII/BIV (*n* = 3) occurred during different seasons (Figure 2.1b). Clusters of giardiasis cases were visualised in the Newcastle and lower Hunter region throughout all seasons. During Autumn, all cases in these communities were caused by assemblage B (Figure 2.1b).

2.4.5 *Giardia* positivity and assemblages among NSW cattle samples

PCR screening of cattle faecal DNA samples using the *18S rRNA* and *gdh* identified a positive response, consistent with *Giardia*, for 64/175 (36.6%) samples. Of the positives, 22/64 (34.4%) samples were positive at both loci, whilst 33 samples were only positive at the *18S rRNA*, and nine samples were positive only at the *gdh* locus. Analyses of *gdh* products by RFLP was successful for 29/31 cattle samples, and both *Giardia* assemblages B and E were detected among the cattle samples. For samples positive at both loci (n = 20/29, 69%), all *gdh* RFLP profiles were consistent with assemblage E. For 9/29 (31%) *gdh* positive only samples, only assemblage B was identified.

Samples positive for *Giardia* were from four farms in Wagga Wagga (dairy calves, n = 7), Richmond (beef cattle, n = 8), and Camden NSW (dairy calves, n = 16) (Waldron et al., 2011a). All samples typed from the Camden farm (n = 15) were identified as assemblage E. Among samples from the Richmond farm, 2/8 (25%) were assemblage E, and 6/8 (75%) were assemblage B. Samples collected from Wagga Wagga farms were typed as assemblage E (3/6, 50%) and B (3/6, 50%).

Table 2.1 Distribution of *Giardia* assemblages/subassemblages identified from human cases in NSW by Local Area Health District

		Assemblages			Subassemblage/Genotype			
Local Health District	Genotyped	A	B	A + B	AII	BIV	AII+BIV	BIII+BIV
Rural/Regional								
Hunter New England	80	11	67	2	11	67	2	0
Mid North Coast	13	1	12	0	1	12	0	0
Murrumbidgee	13	2	11	0	2	11	0	0
Southern NSW	6	1	5	0	1	5	0	0
Western NSW	22	2	20	0	2	18	0	2
Total Rural/Reg (%)	134	17 (12.7)	115 (85.8)	2(1.5)	17 (12.7)	113 (84.3)	2 (1.5)	2 (1.5)
Metropolitan								
Central Coast	1	0	1	0	0	1	0	0
Nepean Blue Mountains	6	1	5	0	1	5	0	0
Sydney	1	0	1	0	0	1	0	0
Western Sydney	23	3	20	0	3	19	0	1
Total Metropolitan (%)	31	4 (12.9)	27 (87.1)	0	4 (12.9)	26 (83.9)	0	1 (3.2)
Total NSW (%)	165	21 (12.7)	142 (86.1)	2 (1.2)	21 (12.7)	139 (84.2)	2 (1.2)	3 (1.8)

Table 2.2. Distribution of *Giardia* assemblages/subassemblages identified from human cases in NSW by patient age (years) and gender.

		Gender		***Giardia* assemblages**		
Age (years)	*n*	Male	Female	A	B	A+B
0-4	32	23	9	3	29	0
5-9	17	13	4	0	17	0
10-14	8	4	4	0	8	0
15-19	4	3	1	0	4	0
20-24	7	4	3	1	6	0
25-29	12	4	8	2	10	0
30-34	19	8	11	2	17	0
35-39	17	7	10	1	16	0
Subtotal 0-39yrs (%)	116	66 (56.9)	50 (43.1)	9 (7.8)	107 (92.2)	0
40-44	14	5	9	1	12	1
45-49	9	7	2	3	5	1
50-54	6	2	4	2	4	0
55-59	4	0	4	2	2	0
60-64	7	5	2	3	4	0
65-69	6	1	5	0	6	0
70+	3	2	1	1	2	0
Subtotal 40+ yrs (%)	49	22 (44.9)	27 (55.1)	12 (24.5)	35 (71.4)	2 (4.1)
Total NSW (%)	165	88 (53.3)	77 (46.7)	21 (12.7)	142 (86.1)	2 (1.2)

Table 2.3 Seasonal distribution of *Giardia* assemblages identified from NSW patients by Local Health District (LHD), during Summer, Autum, Winter and Spring.

	Summer		**Autumn**			**Winter**			**Spring**		
Giardia assemblage	A	B	A	B	A+B	A	B	A+B	A	B	**Total**
Rural/Regional LHDs											
Hunter New England	2	24	2	23	1	3	9	1	4	11	80
Mid North Coast	0	1	1	5	0	0	1	0	0	5	13
Murrumbidgee	0	1	2	7	0	0	1	0	0	2	13
Southern NSW	0	2	1	2	0	0	0	0	0	1	6
Western NSW	1	6	1	8	0	0	3	0	0	3	22
Total Rural/Reg (%)	3 (8.1)	34 (91.9)	7 (13.2)	45 (84.9)	1 (1.9)	3 (16.7)	14 (77.8)	1 (5.6)	4 (15.4)	22 (84.6)	134 (81.2)
Metropolitan LHDs											
Blue Mountains	0	3	1	2	0	0	4	0	0	0	10
Sydney	0	1									1
Western Sydney	3	4	0	7	0	0	5	0	0	1	19
Total Metropolitan (%)	3 (37.5)	8 (72.7)	1 (10)	9 (90)	0	0	9 (100)	0	0	1 (100)	31 (18.8)
Total genotyped (%)	6 (12.5)	42 (87.5)	8 (12.7)	54 (85.7)	1 (1.6)	3 (11.1)	23 (85.2)	1 (3.7)	4 (14.8)	23 (85.2)	165

2.5 Discussion

Analyses of 243 cases of sporadic human giardiasis showed that infections were widely dispersed across eastern regions of NSW (Figure 2.1a), where the majority of the NSW population resides (Australian Bureau of Statistics, 2011). Among the NSW LHDs most (81.5%) sporadic cases occurred in rural/regional LHDs, with clusters observed in coastal centres of Newcastle and Port Macquarie, and in inland centres of Wagga Wagga, Dubbo, and Tamworth. Fewer sporadic cases were identified among the metropolitan LHDs, and were largely confined to Western Sydney and the Blue Mountains. Similar findings have been reported from a Western Australian study, where most sporadic giardiasis cases occurred in rural populations (Yang et al., 2010).

It is unclear why only a small number of metropolitan cases appeared in the present study. Giardiasis notification data for NSW (NSW Health) shows high incidence rates of giardiasis are reported annually from several Sydney LHDs (NSW Department of Health, 2011), but many of these locations were not identified in this study. Many more samples were collected for this study, but we were unable to obtain patient data for these samples, and the locations of these cases are unknown. Samples were collected from private pathology companies, which service all regions of NSW, including metropolitan and regional/rural areas. Several public and private pathology companies service NSW, and it is also possible that the distribution of cases reflects differences in patient use of these services between urban and rural/regional communities.

Genotyping of 165 cases showed that *Giardia* assemblage B was the most common cause of human infection in NSW during the study period. Approximately 86% of all human infections were caused by assemblage B, whilst 12.7% were assemblage A, and mixed A and B infections were identified among two (1.2%) NSW residents. These results are consistent with previous Australian studies, which show assemblage B contributes to 70% - >90% of human cases in low and high prevalence communities (Hopkins et al.,

1997; Read et al., 2002; Yang et al., 2010). The proportion of sporadic assemblage B cases in NSW (86%) is significantly higher than Western Australia (assemblage B, 75%), where a comparable study has been performed (Yang et al., 2010).

Consistent with NSW disease surveillance (NSW Department of Health, 2013), positive human cases were highest during Autumn, and Summer, and among patients aged 0 – 9 years, and 30 – 39 years. Similar frequencies of assemblage A (<17%) and B (>83%) were consistently detected among the total infections in NSW, for each season, in each local health district, and among males and females. *Giardia* assemblage B was the most common cause of infection in each of these groups. The frequencies of assemblages A and B, however, among patients aged 0 – 39 years, and 40 -70+ years were significantly different (Table 2.2). The high prevalence of assemblage B (92.2%) among patients aged 0 – 39 years may be indicative of transmission occurring between, high risk groups, such as infants, children, parents, and carers (United States Environmental Protection Agency, 1999; Hoque et al., 2002; Laupland et al., 2005).

Seasonal shifts in the dispersal patterns of assemblages A and B were observed among the local health districts and were detected by geographical mapping of sporadic cases. Disease clusters were consistently identified in the Newcastle and lower Hunter region (Hunter New England LHD) during all seasons, but assemblage A cases were not detected in these communities during Autumn. Conversely, during Winter and Spring, assemblage A was restricted to the Hunter New England LHD and was not detected in any other NSW LHD (Figure 2.1b). Both assemblages A and B were widely dispersed throughout NSW during Summer and Autumn, where cases were detected in rural/regional and metropolitan centres (Figure 2.1b). Among the inland communities during summer, cases were centralised to the North-Western and Western towns of Tamworth and Dubbo, whilst during Autumn, prevalence in Western and South-Western communities of Dubbo, Orange and Wagga Wagga was evident. Geographical mapping of *Giardia* assemblages

has not been previously performed in NSW, and our results provide a significant insight into NSW cases. Our results indicate the overall prevalence of assemblages A and B among humans is similar across NSW local health districts, but that local differences, in the dispersal of these assemblages, may exist within communities over time.

Further analyses of RFLP profiles showed that within assemblages A and B, three subassemblages/subtypes, AII, BIII, BIV, were present among human cases. All assemblage A cases were subassemblage AII; subassemblage AI was not identified in this study. To date, only one human case of subassemblage AI has been reported from Australia (Read et al., 2004). Identification of subassemblage AII suggests cases are linked to anthroponotic transmission of sporadic giardiasis, due to the restricted host range of this subassemblage (Yang et al., 2010).

Almost all (97.9%) NSW human cases that were caused by assemblage B, were subtype BIV, which is significantly higher than sporadic BIV cases (34.9%) reported in Western Australia (Yang et al., 2010). Mixed template infections of subtype BIII and BIV were identified in three NSW cases. We have identified subtypes here as BIII and BIV for consistency with previous studies, but our results cannot be confirmed as genetically distinct assemblage B infections. Genetic heterogeneity of assemblage B isolates is currently unresolved (Ryan and Caccio, 2013) and inconsistent typing results have been reported by numerous studies (Lebbad et al., 2011; Sprong et al., 2009; Yang et al., 2010). *Giardia* assemblage B is most common to humans, but is also infective to a range of non-human hosts (*c.f.* Caccio & Ryan, 2008; Sprong et al., 2009). Host specificity and transmission cycles associated with different types of assemblage B have not been established.

Detection of the same subtype (BIV) among 84% of total cases in this study suggests genetic similarity between the majority of human cases in NSW. Further screening at additional loci would be required to resolve subtype diversity. Additionally, analysis by

RFLP may underestimate the proportion of mixed template cases due to the lower resolving power of RFLP compared with DNA sequencing and fluorescent methods (Waldron et al., 2009). Restriction fragment length polymorphism was used in this study to facilitate mapping of large sample numbers. The low frequency of mixed A and B (1.2%), and mixed BIII/BIV (1.8%) detected in NSW concords with previous Australian studies, indicating these cases are infrequent.

Among the 64/175 (36.6%) cattle DNA samples positive for *Giardia*, bovine specific assemblage E (69%) was most commonly detected. Human infective assemblage B was detected among nine samples by *gdh* RFLP analyses, but PCR amplification of the same samples at the 18S rRNA locus was unsuccessful. For these samples, further assignment of subassemblages by gdh RFLP was not performed due to inconsistent PCR screening results and PCR positivity which could not be confirmed at two *Giardia* loci. Assemblage B identified among cattle in this study occurred in locations proximal to human populations, in Wagga Wagga and Richmond, which are within the Murrumbidgee and Nepean Blue Mountains LHDs. Transmission of *Giardia* between humans and domestic animals may occur in either direction (zoonotic and reverse zoonotic transmission), and both host species are potential reservoirs for infection, by excrement and environmental contamination (Thompson and Monis, 2012). Previous NSW studies have identified assemblage B and subassemblage AI among non-human hosts (Ng et al., 2011; Vermeulen et al., 2015). The high prevalence of assemblage B among humans in NSW, and absence of AI cases suggests separate transmission cycles of human infective assemblages occurring in NSW, among humans and non-human hosts (assemblage B), and among non-human hosts only (subassemblage AI). (Thompson et al., 2008; Vermeulen et al., 2015).

The results of this study have demonstrated high prevalence of assemblage B among sporadic human cases occurring during the study period. Australia has among the highest reported prevalence of human assemblage B cases in developed nations (*c.f.* Feng and Xiao,

2011), and our results suggest prevalence in NSW may be higher than in other Australian jurisdictions. The similar frequencies of assemblages A and B that were detected among the total number cases (LHD, season) suggests the overall frequencies among these NSW population groups are stable, however, assemblage A may be more common among older patients, and vary in its seasonal geographic dispersal. Among rural/regional LHDs, disease clusters were consistently identified in the Newcastle and lower Hunter region, which persisted throughout all seasons, and has highlighted this area as a disease hotspot in NSW. Detection of assemblage B cases among NSW cattle and human populations in the same geographic locations suggests potential for cross-species transmission.

Essential to disease management of giardiasis is knowledge of *Giardia* assemblages among susceptible host populations, and the distributions of these assemblages across spatial, temporal and demographic scales. The results of the present study have highlighted differences among assemblage A and B human cases, and suggest humans are not a primary source of subassemblage AI infections among non-human hosts. Further research is required to identify cases across metropolitan regions, to identify environmental sources of assemblage B infections in NSW, and determine the potential risk posed by cattle infected with *Giardia* assemblage B.

2.6 Acknowledgments

Financial assistance for this research was provided by Macquarie University. We would like to thank Jenny Musto and Ben Polkinghorne from the NSW Ministry of Health for their invaluable help in providing patient data used in this study. We would also like to thank Douglass Hanly Moir Pathology and Symbion Health Laverty Pathology for kindly providing faecal samples used in this research.

3: Rapid Identification of *Giardia duodenalis* assemblages in NSW using terminal-restriction fragment length polymorphism

A. J. Asher [a], L.S. Waldron [a], and M.L. Power [a],

[a] *Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia*

Keywords: *Giardia duodenalis,* assemblages, subassemblages, T-RFLP, fluorescence

Asher, A.J., Waldron, L.S., Power, M.L. 2012. Rapid identification of *Giardia duodenalis* assemblages in NSW using terminal-restriction fragment length polymorphism. Parasitology 139, 1005-1013.

3.1 Abstract

Humans are infected by two genetic assemblages (A and B) of *Giardia duodenalis*, a protozoan parasite that causes gastrointestinal disease. Subassemblages AI, AII, BIII and BIV are commonly identified in human cases. Detection requires amplification of *G. duodenalis* loci. Subsequent DNA sequencing, or restriction fragment length polymorphism (RFLP), identifies subassemblages but is expensive (DNA sequencing) or insensitive (RFLP). This study investigated a fluorescence based detection method, using terminal-restriction fragment length polymorphism (T-RFLP) of the glutamate dehydrogenase gene, to characterise human infections. Clinical samples (n = 73), positive for *Giardia* were collected in New South Wales, Australia, and were used to evaluate T-RFLP detection. The accuracy and sensitivity of T-RFLP detection was established by comparison to DNA sequencing and RFLP. Subassemblage assignment by T-RFLP identified BIV as the common subtype in NSW cases, whilst AI, AII and BIII were also detected. When compared to DNA sequencing and RFLP, analysis by T-RFLP was a reliable and reproducible method. Automated fluorescent detection enabled accurate sizing of restriction fragments, and provided a sensitive alternative to RFLP. Discrimination of subassemblages by T-RFLP was comparable to DNA sequencing, but was efficient and inexpensive. The protocol described here provides a rapid and sensitive diagnostic tool for routine sample screenings in epidemiological research.

3.2 Introduction

Giardia duodenalis is a protozoan parasite of mammals that causes giardiasis, a disease characterised by diarrhoea, nausea, fatigue, vomiting and weight loss (Caccio and Ryan, 2008; Roxstrom-Lindquist et al., 2006). Previously, human infection has been estimated at 280 million cases per year, worldwide (Caccio and Ryan, 2008; Lane and Lloyd, 2002). Parasite transmission via the faecal-oral route contributes to a high prevalence in areas of poor sanitary and hygiene conditions (Caccio and Ryan, 2008; Savioli et al., 2006). Mammals are infected by eight morphologically indistinguishable genetic assemblages of *G. duodenalis* (syn. *G. intestinalis*, *G. lamblia*) (Read et al., 2004; Thompson et al., 2000; Lasek-Nesselquist et al., 2010). These are referred to as assemblages A through to H, and are only identifiable through molecular analyses (Caccio and Ryan, 2008; Read et al., 2004; Lasek-Nesselquist et al., 2010). Assemblages A and B can be further differentiated into several discrete subassemblages (Read et al., 2004). Humans are infected by assemblages A and B, with subassemblages AI, AII, BIII and BIV most commonly identified (Read et al., 2004; Monis et al., 1999).

The epidemiology of human giardiasis is poorly understood due to complexity in species identification (Smith et al., 2006). Molecular data are required to document the occurrence of human infective subassemblages, evaluate transmission cycles and infection sources, and understand disease aetiology (Caccio and Ryan, 2008; Monis et al., 2003). Identification of assemblages and subassemblages is typically achieved by DNA-sequencing and PCR-restriction fragment length polymorphism (RFLP) (Lebbad et al., 2008; Read et al., 2004). These methods are commonly used in *Giardia* epidemiological research (Yang et al., 2010; Traub et al., 2005a; Read et al., 2004). DNA-sequencing is a robust, highly sensitive, and reliable method that can be applied to all diagnostic loci. However, the expense of sequence analysis often presents as a limitation to large scale studies (Waldron et al., 2009). Rapid assemblage identification is achieved by RFLP of

the diagnostic locus glutamate dehydrogenase (*gdh*) (Read et al., 2004). Variation within the gene differentiates *G. duodenalis* assemblages and human infective subassemblages AI, AII, BIII and BIV (Read et al., 2004). Although RFLP is a cost-effective alternative, the method is limited by gel electrophoresis. Low PCR yields as a result of low parasite numbers and low molecular weights of digestion products impair analysis (Waldron et al., 2009).

This study investigated fluorescence-based detection using terminal-restriction fragment length polymorphism (T-RFLP) as an alternative tool for *G. duodenalis* identification. Fluorescence-based detection offers several advantages over conventional methods. Sample analysis is automated by the use of capillary electrophoresis, enabling high precision, high throughput analysis that is rapid and sensitive (Waldron et al., 2009; Marsh, 1999). Terminal-RFLP is similar to RFLP; DNA is digested with restriction enzymes to produce assemblage and subassemblage specific profiles from sequence polymorphisms (Marsh, 1999). Terminal-RFLP, however, incorporates a fluorescently labelled primer to amplify the target gene, and genotyping is based on the size of only one restriction fragment, the terminal-restriction fragment (Waldron et al., 2009; Marsh, 1999). Upon digestion with restriction endonucleases, only the terminal-restriction fragment (T-RF) adjacent to the fluorescent primer maintains fluorescence, and only the T-RF is detectable by capillary electrophoresis (Marsh, 1999). Initially developed to analyse complex genetic diversity within bacterial communities (Marsh, 1999), T-RFLP now has wide applications including the differentiation of *Cryptosporidium* species and subtype families (Waldron and Power, 2011; Waldron et al., 2009). The purpose of this study was to develop and evaluate T-RFLP for identification of *G. duodenalis* subassemblages that are of importance to human disease. In developing T-RFLP for *G. duodenalis*, our aim was to provide a diagnostic tool that is a rapid, cost-effective, and a sensitive alternative to current molecular techniques.

3.3 Materials and Methods

3.3.1 Clinical samples

In total, 73 clinical samples, positive for *Giardia*, were sourced from pathology laboratories in Sydney, Australia. Sample screenings at pathology companies included a Remel ProSpecT *Giardia/Cryptosporidium* microplate immunoassay (Thermo Fisher Scientific© 2010), and microscopy to visualise *Giardia* cysts in positive faecal smears. Upon collection, samples were irradiated at Macquarie University using a Colbalt^{60} source (1500 Ci, 30 min) to allow for processing external to PC2 laboratory. All samples were assigned a Macquarie University *Giardia* (MQG) number, and stored at 4°C for DNA extraction.

3.3.2 DNA extraction and PCR amplification

DNA was extracted from 100 mg of clinical sample using an ISOLATE Fecal DNA Kit (Bioline, Sydney, Australia), following the manufacturer's instructions. Samples were confirmed as *gdh* competent by PCR amplification following the semi-nested protocol outlined by Read et al. (2004). Primary and secondary reactions (25 μl volumes) were prepared using the GC-RICH PCR System, dNTPack (Roche Diagnostics, Indianapolis, IN), using the reaction conditions previously described (Asher et al., 2012a; Sulaiman et al., 2003). Two positive controls containing DNA extracted from laboratory cultivated *G. duodenalis* trophozoites (QIMR stocks BRIS/91/HEPU/1279 and BRIS/87/HEPU/713) (Upcroft et al., 1995) were included. PCR reactions were performed in an Eppendorf Mastercycler (Eppendorf, North Ryde, Australia). *Gdh* amplicons were resolved by agarose gel electrophoresis (2% w/v, 110V for 30 min) in TBE (Tris, EDTA, boric acid, pH 8) with SYBR safe (Invitrogen, Mulgrave, Australia) using a HyperLadder™ II DNA marker (Bioline, Sydney, Australia) to estimate amplicon size (432 bp fragment).

3.3.3 DNA Sequencing, analyses and phylogenetic inference

To establish a set of reference samples (representatives of AI, AII, BIII and BIV) for T-RFLP development and evaluation, a subset of clinical samples (n = 40) were genotyped by DNA sequencing. Two trophozoite DNA controls were also included. Samples were amplified as described above, and secondary *gdh* amplicons were purified using a QIAquick PCR Purification Kit (Qiagen, Melbourne, Australia), following the manufacturer's instructions. Products were sequenced in the forward direction on a 3130x1 genetic analyser (Applied Biosystems, Foster City, California), using the standard run protocol for a 50cm, 16 capillary array using a BigDye ™ terminator kit (Applied Biosystems). *Giardia duodenalis* subtypes were identified by aligning nucleotide sequences to previously described *gdh* DNA sequences retrieved from the NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html) for subassemblages AI, AII, BIII and BIV (accession numbers; L40509, L40510, AF069059, L40508) (Monis et al., 1996; Monis et al., 1999); and assemblages C, D, E, F and G (accession numbers; U60982, U60986, AY178740, AF069057, AF069058) (Ey et al., 2002 unpublished observations; Monis et al., 1999; Monis et al., 1998). For assemblage H, only partial sequences of the 432 bp amplicon were available in GenBank, and assemblage H could not be included in sequence analyses. DNA sequences were trimmed in GeneiousPRO version 5.0.3 (Biomatters Ltd, Auckland, New Zealand). Trimmed sequences were aligned using Clustal W (Thompson et al., 1994) in *MEGA* version 5 (Tamura et al., 2011). For phylogenetic analyses, nucleotide substitution models were tested for maximum likelihood in *MEGA*5. Akaike Information Criterion corrected (AICc) values were used to determine the optimal parameters. A phylogenetic tree was constructed using maximum likelihood (Tamura 3-parameter distance model with the gamma distribution parameter) and bootstrap analysis (1000 replicates). A *gdh* DNA sequence for *Giardia ardae* (Accession no. AF069060) (Monis et al., 1999) was used as the outgroup.

3.3.4 Design and Development of T-RFLP protocol

T-RFLP design was based on the previously described *gdh* RFLP protocol (Read et al., 2004). Nucleotide sequences (*gdh*) for sub-assemblages AI, AII, BIII and BIV were used to identify T-RFs, and calculate expected sizes. *Nla IV* and *Rsa I* restriction sites were mapped within the secondary primer region, and terminal-restriction profiles were determined for each subassemblage.

The reference group of samples (see above) comprised four clinical samples and two trophozoite DNA controls, identified as sub-assemblages AI, AII, BIII and BIV from DNA sequencing. The reference group was cloned to overcome potential for mutliple T-RFs, from mixed infections in clinical samples. Cloning was performed using the pGEM plasmid vector (Promega, Madison, USA) system and α-Select Gold Efficiency *E. coli* cells (Bioline, Sydney, Australia). Recombinant plasmids were purified using a QIAprep Miniprep Kit (Qiagen, Australia), sequenced, and subassemblages confirmed as previously described.

Gdh PCR was optimised to incorporate a fluorescent primer and generate a fluorescent PCR product. The reverse primer GDHiR was labelled with a 5' 6-carboxyfluorescein (6-FAM) (Sigma-Aldrich, Sydney, Australia) and used in the secondary reaction. A gradient PCR (temperature range 52-60°C, 2°C increments) was performed using positive DNA controls, to determine the optimal annealing temperature for the fluorescent primer.

Purified fluorescent PCR products were digested separately using two restriction enzymes *Nla IV* (New England Biolabs) and *Rsa I* (Roche Diagnostics, Indianapolis, IN) according to the protocol developed by Read et al., (2004).

3.3.5 Capillary electrophoresis for detection of fluorescent terminal-restriction fragments

Digested fluorescent *gdh* amplicons were analysed by capillary electrophoresis at the

Macquarie University Sequencing Facility, as previously described (Waldron et al., 2009), using the DNA size standard LIZ 500 (Applied Biosystems). Outputs were analysed using GeneScan software, version 4.0 (Applied Biosystems). A complete *gdh* PCR sequence (432 bp) was included in each run to control for peak shifts that may arise from run to run variability (Waldron et al., 2009).

3.3.6 Comparison of electrophoretic detection methods for identification of restriction fragments

Detection of diagnostic restriction fragments for identification of *Giardia duodenalis* assemblages is achieved by T-RFLP and RFLP analyses. These methods differ by the electrophoretic technique used to detect base pair lengths of restriction fragments. To examine the sensitivity of T-RFLP for *Giardia* identification and to compare T-RFLP to RFLP, fluorescent PCR amplification and enzymatic digestion was conducted on an additional 33 clinical samples. Digested products were separated by agarose gel electrophoresis (3.5% w/v, 100V for 50 min) using 10 μl of product. These samples were also analysed by capillary electrophoresis (T-RFLP) and results were compared.

3.4 Results

3.4.1 Identification of *G. duodenalis* assemblages by DNA sequencing and phylogenetic analyses

Prior to evaluating T-RFLP, a subset of samples with known *gdh* sequences were needed to develop the method. DNA sequencing was conducted on 40 clinical samples, and phylogenetic analysis was performed to confirm genotyping. Sequencing was successful for the 40 clinical samples selected. For 36 of these, *gdh* nucleotide sequences were 100% homologous to GenBank references for subassemblages AII and BIV. Trophozoite DNA controls were identified as AI. Four clinical samples (MQG95,

MQG103, MQG127, MQG136) contained several single nucleotide polymorphisms (Table 3.1). Nucleotide sequence searches in the GenBank database identified these samples as subassemblages BIII or BIV (Table 3.1). MQG136 could not be resolved below the assemblage level due to sequence diversity.

Table 3.1 Nucleotide sequence diversity observed in DNA sequences for four clinical samples

Position (bp)[a]	GenBank sequences		Clinical samples			
	AF069059 (BIII)	L40508 (BIV)	MQG95	MQG103	MQG127	MQG136
183	T	T	T	T	T	C
345	C	C	T	C	C	T
366	C	T	C	T	T	C
387	C	T	C	C	C	C
396	C	C	C	T	T	C
423	C	C	C	T	C	C
438	G	A	G	A	A	G
Similarity (%) to GenBank sequences			99.7 (BIII)	99.2 (BIV)	99.4 (BIV)	98.6 (BIII) 98.6 (BIV)

[a] Position (bp) refers to the published sequence for L40508

Phylogenetic analysis grouped clinical samples with GenBank references for subassemblages AII, BIII and BIV (Figure 3.1). The phylogenetic tree consisted of two major clusters and seven distinct branches representing assemblages A to G. Human infective assemblages A and B formed two highly divergent groups, as depicted by their phylogenetic distance in the tree. Samples containing higher nucleotide sequence diversity, MQG103 and MQG127, were grouped together in a clade that was sister to subassemblage BIV. Sample MQG136 was positioned internally to subassemblages BIII and BIV. Bootstrap analyses (1000 replicates) showed strong support for clustering of clinical samples within Assemblages A and B. All *gdh* sequences generated in this study were submitted to NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html) under Accession numbers JQ700387 – JQ700436.

3.4.2 Development and evaluation of T-RFLP protocol

T-RF lengths (bp) were determined using *gdh* nucleotide sequences that were retrieved from GenBank. Both restriction endonucleases were required to differentiate T-RFs at the assemblage and sub-assemblage level. At the assemblage level, the expected T-RFs were 131 bp (assemblage A) and 288 bp (assemblage B), whilst the predicted T-RFs for subassemblages were; 146 (AI), 69 (AII), 131 (BIII), and 428 (BIV).

Eight clones containing the recombinant plasmid for subassemblages AI, AII, BIII and BIV were generated for T-RFLP development, and were used as representative controls of known genotypes. T-RFLP on cloned sequences showed that T-RFs were consistent with those expected for assemblages A and B, and sub-assemblages AI, AII, BIII and BIV. T-RFs were within three to four base pairs of the predicted size (Figure 3.2). This result was in agreement with variation observed for the internal control (undigested *gdh* PCR amplicon), detected at 428 bp.

3.4.3 Evalutation of T-RFLP: comparison to DNA sequencing and RFLP

Fourty clinical samples were initially genotyped by DNA sequencing, and of these 39 were screened by T-RFLP. One sample could not be re-amplified by fluorescent PCR, and did not undergo T-RFLP analysis. Genotyping by T-RFLP matched DNA sequencing data for 38 (97%) of the 39 samples screened by both methods (Table 3.2). T-RFLP detected a mixed template in one clinical sample, which was not detected by DNA sequencing.

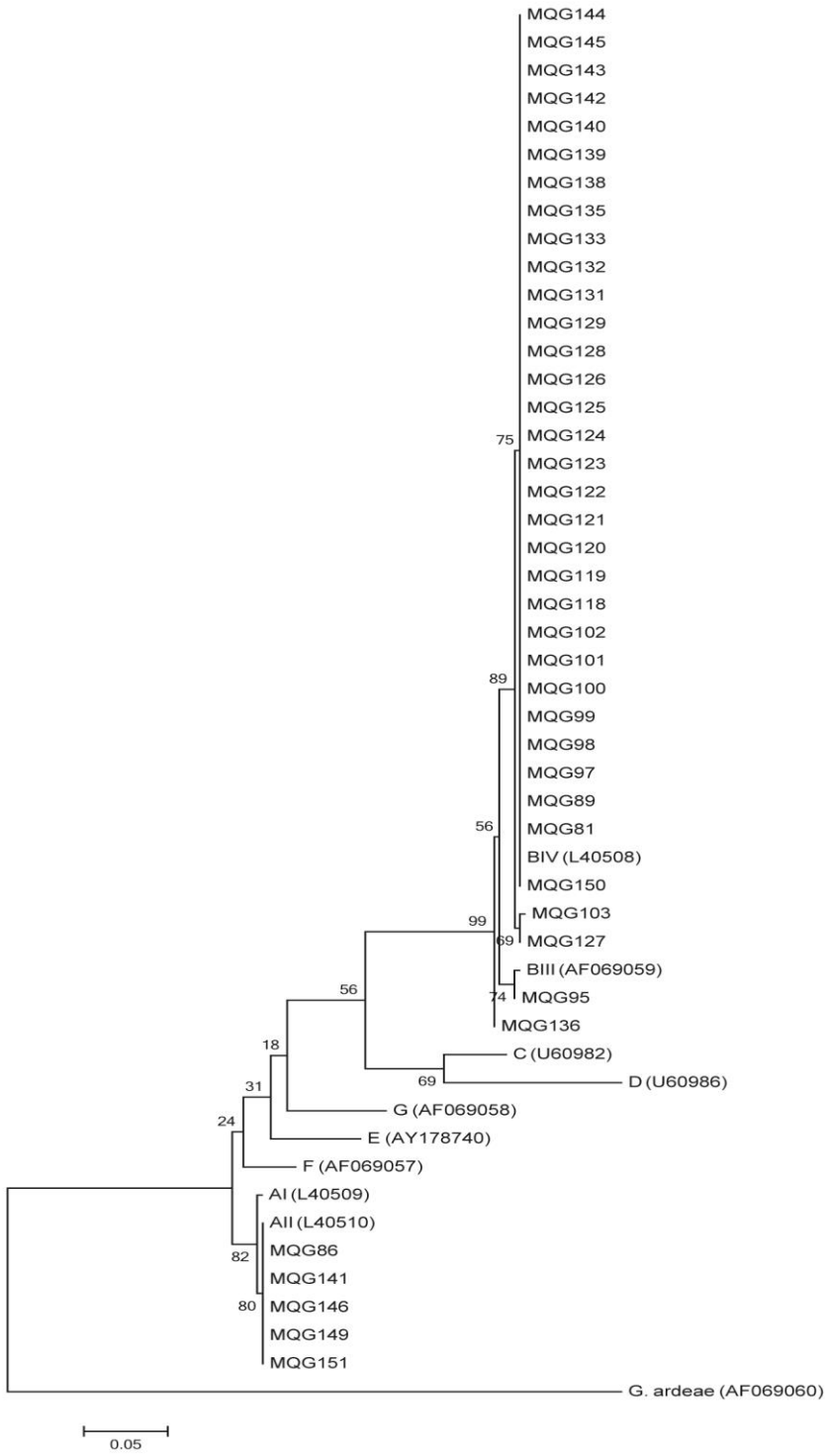

Figure 3.1 Phylogenetic tree of *Giardia duodenalis gdh* sequences inferred by maximum likelihood. The phylogenetic tree shows gdh sequences (359bp) for 40 clinical samples and 10 reference sequences obtatined from GenBank. Clinical samples are divided into two main clusters of assemblages A and B. Bootsrap values (%) supporting the topology of the tree are placed to the left of each node.

An additional 33 clinical samples were genotyped using T-RFLP and RFLP. For RFLP, many restriction fragments could not be resolved by eye following gel electrophoresis. High PCR yields did not produce clear restriction profiles. The sizes of restriction fragments below 100 bp could not be determined, and fragments that were of similar sizes could not be distinguished. We were able to genotype 28 (85%) of the 33 clinical samples by RFLP. Following digestion with *Nla IV* and gel-electrophoresis restriction profiles for five samples contained a large fragment at approximately 120 bp. Diagnostic fragments below 100 bp could not be differentiated, and RFLP profiles could not be accurately determined. For these samples, capillary electrophoresis (T-RFLP) detected T-RFs at 65 bp, and was consistent with subassemblage AII. For subassemblage BIV, a 4 bp difference between the full *gdh* nucleotide sequence (432 bp) and a large restriction fragment for BIV (428 bp) could not be differentiated by gel electrophoresis. A novel restriction profile was generated for one clinical sample. For this sample, T-RFLP and RFLP data were consistent. Profiles did not correspond to any *G. duodenalis* assemblage or subassemblage, and the sample was not identified.

Table 3.2 Genotyping results for seventy-two giardiasis clinical samples

Genotyping method	Sub-assemblages detected						
	Total	AI	AII	BIII	BIV	BIII/BIV	Unknown
DNA sequencing & T-RFLP[a]	40		5	1	31	3	
RFLP and T-RFLP	33	1	5	2	23	1	1
Total	73	1	10	3	54	4	1

[a] Thirty-nine of forty clinical samples with DNA sequencing data were screened by T-RFLP

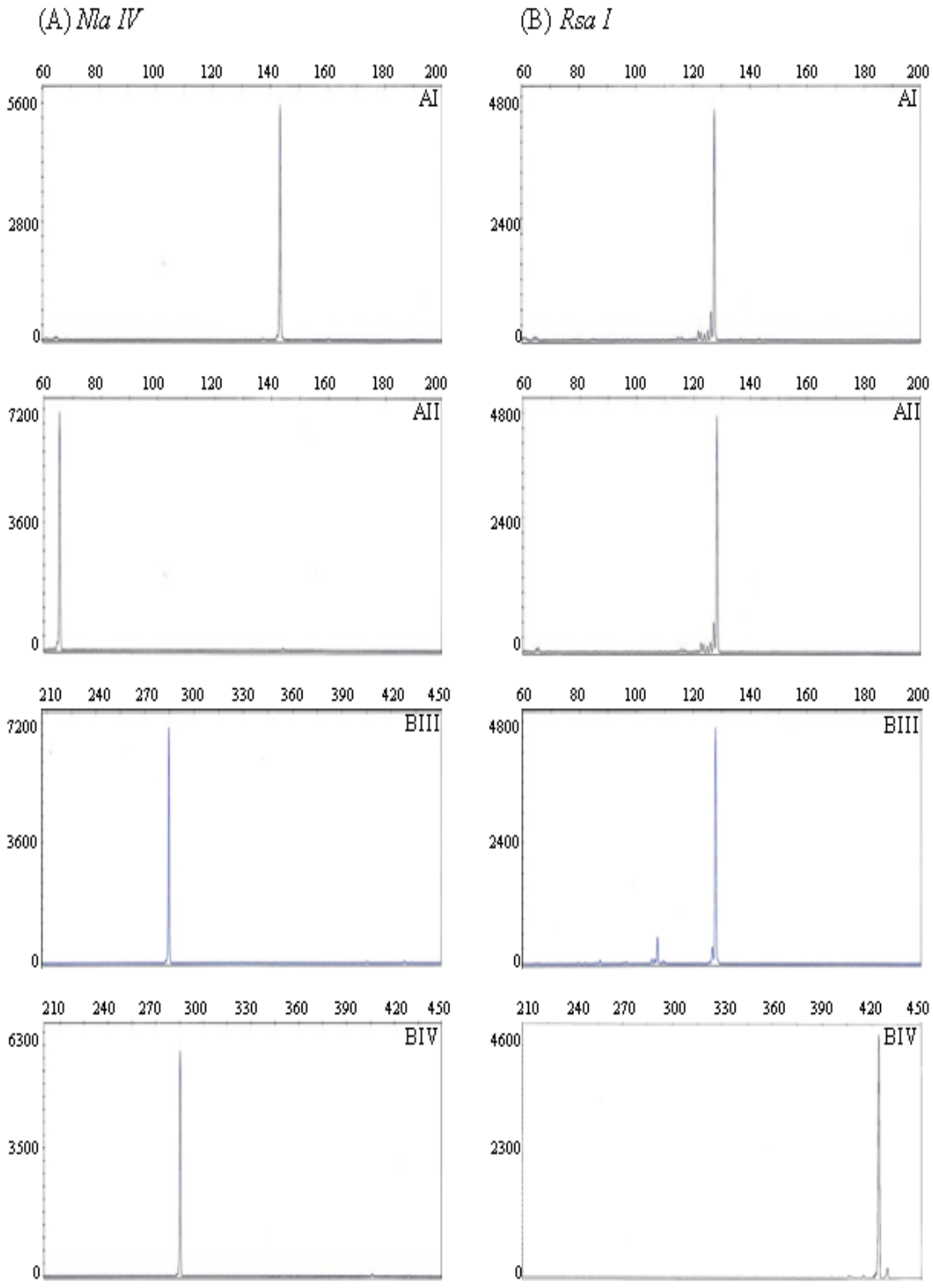

Figure 3.2 Electropherograms showing assemblage and subassemblage specific T-RFs following digestion with *Nla IV* and *Rsa I*. Digestion with *Nla IV* generated T-RFs observed at 143 bp (AI), 65 bp (AII), and 285 bp (assemblage B). Subassemblages BIII and BIV are not differentiated by *Nla IV*. Digestion with *Rsa I* produced T-RFs observed at 128 bp (BIII), 424 bp (BIV), and 128 bp (assemblage A). A minor secondary peak at 108 bp is visible in the electropherograms for BIII (Rsa I) and is not typical of the BIII T-RFLP profile. All T-RFs are within 3 to 4 bp of the predicted sizes.

3.4.4 Diversity of *G. duodenalis* subtypes in human samples from NSW

Seventy-two clinical samples were successfully genotyped (Table 3.2). Assemblage B was detected in 61 samples (85%) and of these, 54 (88.5%) samples were subassemblage BIV, three were sub-assemblage BIII, and four clinical samples generated multiple T-RFs, consistent with mixed BIII/BIV templates (Table 3.2). Assemblage A was detected in 11 clinical samples (15%), with sub-assemblages AI (n = 1) and AII (n = 10).
Unique T-RFLP profiles were generated for 17 assemblage B clinical samples. Profiles contained additional T-RFs that were observed as low intensity secondary fluorescent peaks in the electropherograms (fluorescence intensity approximately ≤ 20% of primary genotyping peak). These minor peaks did not correspond to any other *G. duodenalis* assemblage. Secondary peaks that were consistent with a mixed template, or that were of high fluorescent intensity, were investigated further by comparing T-RFLP results to sequencing data.

3.4.5 Investigation of mixed T-RFs and sequence diversity

Sequencing data was available for three clinical samples (MQG123, MQG127, MQG136) and one cloned sequence (clones 95-1) that contained multiple peaks in the T-RFLP profiles. Additional T-RFs for samples MQG123, MQG127 and MQG136 were consistent with mixed BIII/BIV templates. Investigation of DNA chromatograms revealed double peaks at position 366 of the published gene (L40508), which corresponded to an additional *Rsa I* restriction site, producing a T-RF for subassemblage BIII. The additional restriction site for subassemblage BIII could not be located in the chromatogram for sample MQG123, and DNA sequencing/T-RFLP data were not consistent for this sample.

Two reverse primer binding sites for GDHiR were identified in the DNA chromatogram for cloned sequence 95-1. These sites generated two secondary PCR products, at 432 bp and 499 bp. Multiple T-RFs observed in the T-RFLP profile for clone 95-1 (*Nla IV* and

Rsa I) were consistent with digestion of both PCR products and BIII restriction sites (Figure 3.3).

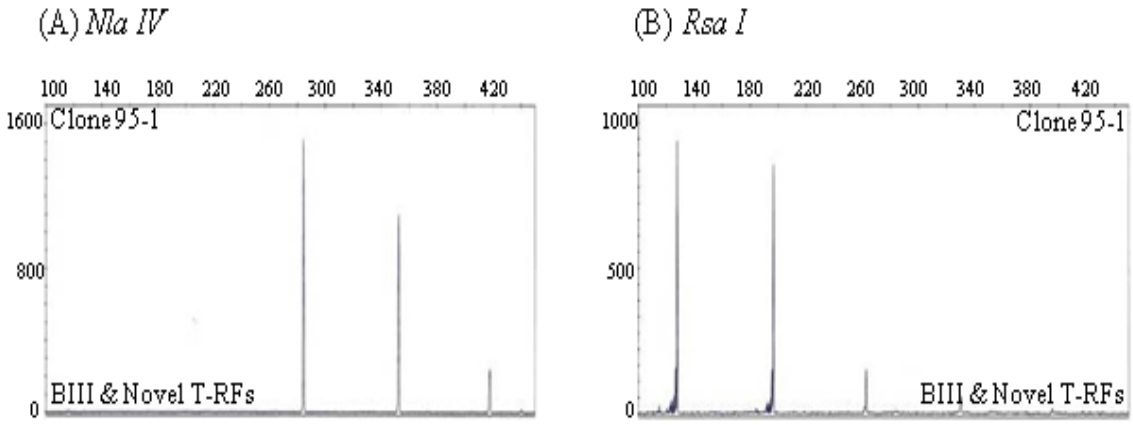

Figure 3.3 Electropherograms showing multiple T-RFs consistent with nucleotide sequence diversity. Digestion of cloned sequence 95-1 with Nla IV (A) generated a primary T-RF observed at 285 bp and secondary T-RFs at 352 bp and 418bp. Digestion with RsaI (B) produced a primary T-RF observed at 128 bp and secondary T-RFs at 196bp, 263 bp and 329 bp. Primary T-RFs are consistent with subassemblage BIII. Secondary T-RFs were generated by an additional reverse primer binding site in the nucleotide sequence.

3.5 Discussion

This study investigated T-RFLP as a cost effective and alternative molecular diagnostic tool for *G. duodenalis* identification. *G. duodenalis* is globally distributed and poses a continuous threat to human populations (Caccio and Ryan, 2008). Understanding complex host-parasite relationships and disease transmission cycles requires identification of genetic subtypes, and the development of reliable but inexpensive molecular tools for routine sample screening.

Analyses of 73 clinical samples showed that T-RFLP is a sensitive, robust and effective diagnostic tool for *G. duodenalis* identification. Fluorescent *gdh* amplification,

followed by enzyme digestion, produced fluorescent T-RFs that were assemblage and subassemblage specific. Terminal-RFs were within three to four base pairs of the predicted size due to shifts in the detection system (Waldron et al., 2009, Schutte et al., 2008). Terminal-RFs are detected by capillary electrophoresis and base pair length is measured against a size standard (Marsh, 1999, Schutte et al., 2008). Similar to RFLP, factors such temperature and the amount of DNA product can cause slight variations in the size that the T-RF is detected at. By including a *gdh* nucleotide sequence of known length to control for these shifts, the size of each T-RF could be accurately determined.

We evaluated the sensitivity and accuracy of T-RFLP by comparing the method with DNA sequencing and RFLP for the *gdh* gene. Consistent results between DNA sequencing and T-RFLP data demonstrated that fluorescent detection of the T-RF is a reliable and reproducible method. Of the 73 clinical samples used in this study, we were able to screen 72 by T-RFLP. Whilst DNA sequencing is still required to resolve ambiguous samples, we found T-RFLP is a fast and effective genotyping tool to routinely screen samples, and a low cost alternative. Automated detection enabled high efficiency, high throughput processing (Marsh, 1999), and rapid identification of subassemblages. Although DNA sequencing is automated, additional thermal cycling and purification prior to capillary electrophoresis is required, and increases sample turn-around times by comparison. To evaluate the protocol, we chose to analyse *Nla IV* and *Rsa I* digestions for each sample separately. T-RFLP for both digestions per sample was performed at 33% of the DNA sequencing price, but could be processed at approximately 16%, if digested products for each sample were combined for capillary electrophoresis.

In comparison with RFLP-gel electrophoresis, T-RFLP was a more sensitive and precise technique, requiring a lower concentration of template DNA and only 1 μl for analysis. Subassemblage assignment by RFLP was problematic, when diagnostic restriction fragments were of similar base pair size, or when partial digestions produced

ambiguous profiles. Low molecular weight restriction fragments were also not visible following RFLP and gel electrophoresis. In contrast these fragments were clearly identified by fluorescent detection of the T-RF. Digital outputs of fluorescent peaks allowed accurate sizing, and T-RFs of similar base-pair length could be unambiguously differentiated.

For a number of samples in this study, evidence of mixed templates and/or sequence heterogeneity was detected by DNA sequencing and T-RFLP. Mixed templates have been previously identified in human cases; however, the mechanisms that generate heterogeneity are not well understood (Caccio and Ryan, 2008; Lebbad et al., 2008). *Giardia* trophozoites contain two nuclei, and it is unclear whether mixed templates and heterogeneity result from mixed infection or genetic exchange (Caccio and Sprong, 2010; Lebbad et al., 2008).

For this type of detection, we found T-RFLP to be more informative than RFLP and DNA sequencing. Due to the low resolving power of RFLP and gel electrophoresis, we found the higher sensitivity of capillary electrophoresis improved accuracy in detection, and confidence in subassemblage assignment. Capillary electrophoresis has been previously applied to RFLP analyses (Nachamkin et al., 2001) but would produce complex electropherograms for *G. duodenalis*. Digestion of *gdh* sequences with *Nla IV* and *Rsa I* generates numerous restriction fragments, many of which are non-specific. Detection of these fragments by capillary electorphoresis would generate complex banding patterns, and profiles containing mixed templates would be difficult to resolve. Although T-RFLP limits analysis of sequence diversity to the terminal-restriction site, electropherograms are simple to interpret and provide a clear distinction between fluorescent peaks of different subassemblages.

Comparison of T-RFLP to DNA sequencing data identified double peaks in the DNA chromatograms of two clinical samples, which generated additional enzyme restriction

sites, and corresponded to more than one *G. duodenalis* subassemblage. Whilst double peaks in chromatograms can indicate mixed templates or heterogeneity, it can also be the product of PCR artifacts (Lebbad et al., 2008). Analysis by T-RFLP confirms that sequence diversity is present in the sample, through the production of terminal-restriction fragments that would not otherwise appear.

Further modelling of the T-RFLP protocol has shown that whilst subassemblage AII can be differentiated from subassemblages AI, BIII and BIV, different primers and/or restriction enzymes are required to differentiate AII from other *G. duodenalis* assemblages. Whilst we were able to confirm the presence of subassemblage AII using additional molecular data, the protocol must be optimised to investigate zoonotic transmission.

Seventy-two clinical samples were genotyped in this research, and assemblage B was most commonly detected (85% of samples), followed by A (15% of samples). These results provide further evidence that assemblage B is highly prevalent among NSW human cases. Our results correspond with previous research, where prevalence of assemblage B has been documented in Western Australia during separate investigations of sporadic giardiasis (Yang et al., 2010; Read et al., 2002; Hopkins et al., 1997). Based on our analyses of *gdh* sequences, we have identified the presence of four *G. duodenalis* subassemblages causing human infection in NSW. Among all samples, subassemblage BIV was most commonly detected (75% of samples), followed by AII (14% of samples).

In this study, *G. duodenalis* genotyping was conducted at the *gdh* locus. To confirm subassemblage assignment, analyses across several loci is usually required (Caccio and Ryan, 2008). The high cost of DNA sequencing limits the scope of many investigations, (Waldron *at al*. 2009), and a range of diagnostic methods are often used to achieve multilocus genotyping. We have described the first application of T-RFLP, for identification of *G. duodenalis* subtypes that are of importance to human disease. The T-RFLP method requires successful PCR amplification of *Giardia* loci, to generate

fluorescent PCR amplicons for downstream T-RFLP analyses (Waldron et al., 2009). Application of this method to diagnostic loci such as the 18S rRNA would improve T-RFLP detection due to the increased sensitivity of 18S rRNA PCR amplification when compared to the *gdh* (Thompson and Ash, 2015). Due to the high diagnostic performance of the protocol, we expect that T-RFLP can be adapted to detect a wider range of assemblages, and applied to other diagnostic loci. Detection of sequence diversity, and the generation of unique sample profiles in this study have demonstrated potential application as rapid and inexpensive source tracking tool in outbreak situations. The high precision and accuracy of this technique, combined with the low cost and high thoughput nature, means that T-RFLP is an optimal tool for routine identification of *G. duodenalis*. Detection by T-RFLP should improve processing efficiency and provide a cost effective means to screen larger sample sizes, generating molecular information required to understand the epidemiology of human disease.

3.6 Acknowledgments

This research was conducted with financial assistance from Macquarie University. We would like to thank Dr. Jacqui Upcroft of the Queensland Institute of Medical Research for kindly providing trophozoites and also Douglass Hanly Moir Pathology, Symbion Health Laverty Pathology, Hunter Area Pathology Service, Westmead Hospital & John Hunter Hospital for their invaluable help in providing faecal samples

4: Distribution of *Giardia duodenalis* Assemblages A and B among children living in a remote Indigenous community of the Northern Territory, Australia

A J. Asher [a], D.C. Holt [b], R.M. Andrews [b] and M.L. Power [a]

[a] *Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia*

[b] *Menzies School of Health Research, Charles Darwin University, Casuarina, Northern Territory, Australia*

Keywords: *Giardia duodenalis*, assemblages, prevalence, children, Indigenous health

Asher, A.J., Holt, D.C., Andrews, R.M., Power, M.L., 2014. Distribution of *Giardia duodenalis* Assemblages A and B among Children Living in a Remote Indigenous Community of the Northern Territory, Australia. PloS One 9, e112058.

4.1 Abstract

Giardiasis is a communicable gastrointestinal disease caused by *Giardia duodenalis* and two genetic assemblages, A and B, cause human infection. In remote Indigenous communities of Australia, giardiasis is highly prevalent among children but disease transmission is poorly understood. This study investigated the prevalence of *Giardia* and genetic subtypes contributing to human disease in a remote Indigenous community, in the Northern Territory of Australia. Eighty-seven faecal samples were collected from 74 children (< 15 years) over an 18 month period, and the distribution of positive cases relative to participant age and gender were examined. Screening by microscopy and 18S rRNA PCR amplification showed 66.7% (58/87) of faecal samples were positive for *Giardia*. Both males and females were equally affected and high detection rates were obtained for participants aged 0 – <5 years and 5 – <10 years (66.0 and 60.0% respectively). For 58.6% of the positive samples, *Giardia* was only detected by 18S rRNA PCR. Approximately 75% of cases were assemblage B, and subassemblage analyses using terminal-restriction fragment length polymorphism of the glutamate dehydrogenase gene demonstrated that a variety of genetic variants were present. The high proportion of positive cases that were not detectable by microscopy, and dominance of assemblage B cases highlights the need for further research in this community, to assess the contribution of *Giardia* to chronic gastrointestinal disease among children, and to understand conditions conductive to assemblage B transmission.

4.2 Introduction

In Australia, the mortality rates for Indigenous children (0 – 14 years) are more than two times higher than the rates for non-Indigenous children, and overall life expectancy of Indigenous Australians remains 9.7 – 11.5 years below the non-Indigenous population (Australian Institute of Health and Welfare, 2011). In remote Indigenous communities, high rates of gastrointestinal, respiratory, and skin infections are common and children are most at risk from constant infection and re-infection (Holt et al., 2010). Chronic childhood infections are linked to poor physical and cognitive development, poor educational outcomes, socio-economic disadvantage, and poor health status throughout life (McDonald and Bailie, 2010).

Giardia duodenalis, a protozoan parasite, is a significant cause of gastrointestinal disease (giardiasis) and morbidity (Thompson et al., 2001). In remote Indigenous communities giardiasis prevalence is high, ranging from 15 to 36% (Thompson et al., 2001; Jones, 1980), compared to a national prevalence of 2 to 7% (Thompson and Monis, 2012). Among Indigenous children living in remote communities, prevalence of giardiasis is estimated between 32 to 65% and frequency of transmission is comparable to rates observed in developing nations (Thompson et al., 2001; Meloni et al., 1993; Reynoldson et al., 1998). Constant exposure to *Giardia* leads to chronic gastrointestinal disease, malnutrition, and failure to thrive (Thompson, 2008).

In the Northern Territory of Australia approximately 80% of the Indigenous population live in areas classified as remote or very remote (Australian Institute of Health and Welfare, 2011; McDonald et al., 2008). Communities range in size from small groups to a few thousand people, and communities are geographically isolated (McDonald et al., 2008). Overcrowded living conditions, inadequate housing and community sanitation facilities, and poor personal hygiene contribute to the high rates of disease transmission in these communities (McDonald et al., 2008). Previous initiatives to manage infectious

diseases have included improved housing and community wide drug treatment programs; however, many diseases continue to persist with high infection rates (Holt et al., 2010; McDonald and Bailie, 2010) . The high frequency of *Giardia* transmission and continued persistence in remote Indigenous communities is poorly understood.

Giardia duodenalis is a species complex and two genetic assemblages (A and B) infect humans, domestic animals, and wildlife (Caccio and Ryan, 2008; Lasek-Nesselquist et al., 2010). These assemblages are broad clusters of genetically related isolates (Thompson and Monis, 2012) and four human infective subassemblages (AI, AII, BIII and BIV) have been previously described (Ryan and Caccio, 2013; Read et al., 2004). Genetic diversity within assemblage B, however, is higher than assemblage A; assemblage B subgroups are unresolved; and numerous assemblage B genotypes contribute to human and animal infection (Ryan and Caccio, 2008; Feng and Xiao, 2011). Identification of different genetic types that contribute to disease enables differences in host specificities, transmission cycles, and sources of infection to be more closely examined (Feng and Xiao, 2011).

Several epidemiological studies of giardiasis have been conducted in remote Indigenous communities in Australia (Thompson et al., 2001; Meloni et al., 1993; Reynoldson et al., 2008), but few have performed molecular analyses (Hopkins et al., 1999; Hopkins et al., 1997) to identify genetic subtypes contributing to high infection rates. The few molecular epidemiological studies that have been undertaken predate the current understanding of *Giardia*, generated by molecular methods (Hopkins et al., 1999; Hopkins et al., 1997). Subtype identification requires analyses of *G. duodenalis* DNA and are not conducted in routine pathology screenings. The geographic remoteness of communities limits access and feasibility of performing offsite DNA screening for samples. Additionally, in the Northern Territory giardiasis is not listed as a notifiable disease and epidemiological information is not routinely collected for positive cases. It is unclear if

different *G. duodenalis* genetic variants exist in communities and contribute to high reinfections rates among children (Reynoldson et al., 1998).

The purpose of this study was to investigate the prevalence of *G. duodenalis* among children in a remote Indigenous Australian community, and to examine the distribution of genetic assemblages and diversity of subtypes present. Knowledge of *G. duodenalis* assemblages and subtypes is required to understand disease transmission, assess public health risks posed by community conditions, and to understand how these factors contribute to persistent giardiasis infections in these communities.

4.3 Materials and Methods

4.3.1 Ethics statement

Ethical approval for the community based mass drug administration trial, including the collection and analysis of faecal samples, was obtained from the Human Research Ethics Committee of the Menzies School of Health Research and Northern Territory Department of Health. Additional approval for the analysis of the faecal samples was also obtained from the Macquarie University Human Research Ethics Committee. Written informed consent was obtained from the parent or guardian for the collection and analysis of faecal samples from children. Participant age group and gender were used to analyse the distribution of *Giardia* positive cases. No additional or identifying information was provided for participants included in this study.

4.3.2 Faecal samples, microscopy and DNA extraction

To investigate the prevalence of *G. duodenalis*, we examined 87 faecal samples that were collected as part of a mass drug administration trial of ivermectin for the control of scabies and strongyloidiasis in a remote community in the Northern Territory (Kearns et al., 2009). Faecal samples were acquired from children aged less than 15 years as a routine

component of the primary study (Kearns et al., 2009). A total of 87 faecal samples were collected from 74 children at three time points over an 18 month period. Samples were examined by direct smear for the presence of *Giardia* and the remainder stored in ethanol at -20^{o}C until DNA extraction. DNA was extracted from 200mg of faeces using a PowerSoil DNA Isolation Kit (Mo Bio, Calsbad, CA) according to the manufacturer's instructions.

4.3.3 PCR and sequence analyses of the 18S rRNA gene

DNA samples were screened for the presence of *Giardia* DNA by PCR amplification of a fragment of the small subunit ribosomal RNA (18S rRNA) gene, using a two-step nested PCR method and cycling conditions described (Hopkins et al., 1997; Read et al., 2002) with primers RH11, RH4LM, GiAR18SeR and GiAR18SiR. The GC-RICH PCR System, dNTPack (Roche Diagnostics, Indianapolis, IN, USA) was used to prepare PCR reactions as previously described (Sulaiman et al., 2003; Asher et al., 2012a), in a total volume of 25 μl, containing 1 μl of template DNA.

DNA sequencing was performed for 18S rRNA PCR products to identify *Giardia duodenalis* assemblages. Positive secondary PCR products were purified with a QIAquick PCR Purification Kit (Qiagen, Melbourne, Australia), according to the manufacturer's instructions. Purified PCR products were sequenced by Macrogen Inc. (Seoul, Korea) in both the forward and reverse directions, using the secondary PCR primers (GiaR18SeR and GiaR18SiR). Geneious PRO version 5.3.6 (Biomatters Ltd, Auckland, New Zealand) was used to generate a contiguous 18S rRNA sequence (174 - 175 bp) for each sample. To identify *G. duodenalis* assemblages from 18S rRNA contigs, BLASTn sequence searches of the NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html) were performed. 18S rRNA contigs were aligned to previously described 18S rRNA sequences retrieved from the GenBank database (accession numbers AF199446 (assemblage A),

AF199447 (assemblage B), AF199449 (assemblage C), AF199443 (assemblage D), AF199448 (assemblage E), AF199444 (assemblage F), AF199450 (assemblage G) (Thompson et al., 2000). Nucleotide sequence alignments were performed using Clustal W (Thompson et al., 1994) in Geneious PRO. All 18S rRNA contiguous sequences that were generated in this study were submitted to the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena) under accession numbers LN611577-LN611621.

4.3.4 Terminal-restriction fragment length polymorphism (T-RFLP) for gdh products

PCR amplification of the glutamate dehydrogenase (*gdh*) gene was performed for all samples to enable subassemblage identification. A 432bp fragment of the *gdh* gene was amplified following the semi-nested protocol and cycling conditions described (Read et al., 2004) using primers GDHeF, GDHiF and GDHiR. The GC-RICH PCR System, dNTPack was used to prepare the reactions (Asher et al., 2012a) in a total volume of 25 µl, containing 1 µl of template DNA.

To identify *G. duodenalis* subassemblages and examine the population structure of individual infections, *gdh* PCR products were analysed by T-RFLP using the method previously described (Asher et al., 2012b). Secondary fluorescent PCR was conducted on *gdh* PCR positive samples and products were purified as described above. To generate fluorescent terminal-restriction fragments (T-RFs), products were digested with two restriction endonucleases; *Nla IV* (New England BioLabs) and *Rsa I* (Roche Diagnostics, Indianapolis, IN, USA) (Read et al., 2004). Fluorescent T-RFs were detected by capillary electrophoresis at the Macquarie University Sequencing Facility (Waldron et al., 2009). Length in basepairs of T-RFs were estimated using the DNA size standard LIZ 500 (Applied Biosystems) and GeneScan software version 4.0 (Applied Biosystems) (Asher et al., 2012b; Waldron et al., 2009).

4.3.5 Statistical analyses

Differences in *Giardia* positivity and *G. duodenalis* assemblages among males and females, participant age groups, and collection round were investigated using chi-squared statistical analyses. The frequency of males and females among the total number of samples collected were used to determine expected frequencies for *Giardia* positivity, assuming no gender bias for *Giardia* infections, and assuming equal distributions of genetic assemblages. The number of samples collected for participants aged > 10 years, and those collected in round 3 were too small to determine the statistical significance of results, and were recorded as observations.

4. 4 Results

4.4.1 Comparison of *Giardia* screening results over the 18 month collection period

Screening for *Giardia* by microscopy and 18S rRNA PCR produced positive results for a total of 58/87 (66.7%) faecal samples collected over the study period (Table 4.1). A total of 24/87 (27.6%) were positive by microscopy, and 54/87 (62.1%) were positive by 18S rRNA PCR amplification (Table 4.1). Comparison of results showed that of the 58 positive samples, 20/58 (34.5%) were positive by both screening methods, 34/58 (58.6%) were positive by 18S rRNA PCR only, and 4/58 (6.9%) samples were positive by microscopy only. *Giardia* was not detected by either screening method in 29/87 (33.3%) samples.

Detection rates of *Giardia* in faecal samples (combined microscopy and PCR results) were high in all three collection periods. At round 1 (month 0), 54.3% (19/35) of samples were positive, 72.3% (34/47) were positive at round 2 (month 12), and all five of the samples collected during a more limited follow-up at round 3 (month 18) were positive (Table 4.1). The increase in positivity detected between rounds 1 and 2 was almost entirely attributable to an increase in PCR positivity in round 2, compared with round 1.

Table 4.1 Comparison of *Giardia* positive 18S rRNA PCR results to microscopy screenings by participant age, gender, and collection round for 87 faecal samples (74 participants) collected from children living in a remote Indigenous community in the Northern Territory.

		Positive Faecal Samples (*n* = 87)				**Participant Data (*n* = 74)**			
		Collection round				**Age (years)**			
	Gender	Round 1	Round 2	Round 3	Total (%)	0 – <5	5 – <10	10 – <15	Total (%)
Collected	Males	19	24	5	48 (55.2)	28	10	0	38 (51.4)
	Females	16	23	0	39 (44.8)	22	10	4	36 (48.6)
	Total	35	47	5	87	50	20	4	74
Microscopy	Males	7	8	2	17 (70.8)	9	4	0	13 (68.4)
	Females	4	3	0	7 (29.2)	4	2	0	6 (31.6)
	Total	11	11	2	24 (27.6)	13	6	0	19 (25.7)
18S rRNA	Males	10	16	5	31 (57.4)	18	6	0	24 (53.3)
	Females	6	17	0	23 (42.6)	15	3	3	21 (46.7)
	Total	16	33	5	54 (62.1)	33	9	3	45 (60.8)
Total positive	Males	11	17	5	33 (56.9)	18	7	0	25 (52.1)
(combined results)	Females	8	17	0	25 (43.1)	15	5	3	23 (47.9)
	Total	19	34	5	58 (66.7)	33	12	3	48 (64.9)

There was no difference in the percentage of positive samples detected by microscopy at the two collection rounds (11/35 (31.4%) versus 11/47 (23.4%) respectively) (Table 4.1).

4.4.2 *Giardia* positivity for 74 participants by age, gender and collection round

In total there were 74 participants that contributed the 87 faecal samples over the 18 month collection period. Of the 13 participants that contributed samples twice, 11 contributed samples at rounds 1 and 2 of collection, and two participants contributed samples at rounds 2 and 3 of collection. Only the first sample was included for each participant in the following analyses based on total participants. Participants were aged 0 – <5 years (50/74, 67.6%), 5 - <10 years (20/74, 27.0%), and 10 - <15 years (4/74, 5.4%). There were slightly more males (38/74, (51.4%) than females (36/74, 48.6%) included in the study (Table 4.1).

The total number of participants with positive screening results for *Giardia* by microscopy and/or 18S PCR was 48/74 (64.9%) (Table 4.1). There was no difference in the total number of positive males (25/48, 52.1%) and females (23/48, 47.9%), or between those aged 0 - <5 years (33/50, 66.0%) and 5 - <10 years (12/20, 60.0%). A total of 19/74 (25.7%) participants were positive for *Giardia* by microscopy. There was no difference in microscopy positivity between the two age groups (0 - <5 years 13/50, 26.0%; 5 - <10 years 6/20, 30.0%); however, more males were positive (13/19, 68.4%) than females (6/19, 25.7%).

A total of 45/74 participants were 18S rRNA PCR positive and there was no difference in the proportion of positive males (24/45, 53.3%) and females (21/45, 46.7%) (Table 4.1). PCR positivity, however, was 46.7% greater for those aged 0 – <5 years (33/50, 66.0% positive) compared with those aged 5 – <10 years (9/20, 45.0% positive) (Chi-sq = 7.26, df = 1, p-value = 0.01, 95% CI 0.8-6.8). For participants aged 0 - <5 years, 18S rRNA positivity was higher at collection round 2 (23/30, 76.7%) compared with

collection round 1 (14/27,51.9%) (Chi-square = 5.66, df = 1, p-value = 0.02, 95% CI 0.1-1.0), and this was consistent with the increase in 18S rRNA PCR positivity previously described for round 2. Changes in PCR positivity for participants aged 5 - <10 years and 10 - <15 years between collection rounds 1 and 2 could not be determined due to small sample numbers within these groups. The equal distributions of 18S rRNA positivity among males and females, however, were maintained between rounds 1 and 2 (Table 4.1).

4.4.3 DNA sequence analyses of 18S rRNA amplicons and identification of *Giardia* assemblages A and B

DNA sequencing was successful for 45/54 18S rRNA PCR positive samples, and a complete contiguous sequence for the secondary PCR product was assembled for 43 samples. A shorter (158 bp) contiguous 18S rRNA sequence was assembled for two further samples. Polymorphic nucleotide positions for all *Giardia* assemblages were within this 158 bp region, and thus it was possible to genotype the shorter contiguous sequences.

BLASTn sequence searches and nucleotide alignment to reference sequences obtained from GenBank (previously described) identified only assemblages A and B among the samples. A total of 41/45 sequences were 100% identical to the published sequences for either assemblage A or assemblage B. The remaining four sequences contained a novel single nucleotide deletion at position 162 of the 175 bp 18S rRNA contig, but were otherwise identical to either assemblage A or assemblage B.

Both assemblage A and assemblage B were detected in faecal samples from all three collection periods (Table 4.2). For the 45 samples genotyped, the majority were identified as assemblage B (34/45, 75.6%), while assemblage A was identified in less than a quarter of samples (11/45, 24.4%) (Table 4.2). Assemblage B was more commonly found at each of the three collection time periods, in each of the three age groups, and in both males and females (Table 4.2). Despite an increase in *Giardia* 18S rRNA positivity at collection

Table 4.2. *Giardia duodenalis* assemblages identified from children in a remote Indigenous community in the Northern Territory by DNA sequencing of the *Giardia* 18S rRNA locus.

		Faecal samples (n = 45) Collection round				**Participant data (n = 38) Age (years)**			
	Gender	Round 1	Round 2	Round 3	Total (%)	0 – <5	5 – <10	10 – <15	Total (%)
Assemblage A	Males	2	3	1	6 (54.5)	4	2	0	6 (54.5)
	Females	1	4	0	5 (45.5)	4	1	0	5 (45.5)
	Total (%)	3 (21.4)	7 (25.0)	1(33.3)	11 (24.4)	8 (27.6)	3 (42.9)	0 (0.0)	11 (28.9)
Assemblage B	Males	7	10	2	19 (55.9)	12	2	0	14 (51.9)
	Females	4	11	0	15 (44.1)	9	2	2	13 (48.1)
	Total (%)	11 (78.6)	21 (75.0)	2 (66.7)	34 (75.6)	21 (72.4)	4 (57.1)	2 (100.0)	27 (71.1)
Total genotyped		14	28	3	45	29	7	2	38

round 2 (previously described), there was no significant difference in the frequencies of assemblages A and B at collection round 2, compared with round 1 (Table 4.2).

4.4.4 Identification of *Giardia* subassemblages by *gdh* T-RFLP and distribution among samples

A total of 33/87 faecal samples were amplified at the *gdh* locus and genotyping by T-RFLP was successful for 32 samples. *Gdh* T-RFLP analyses identified four different types of genetic profiles present in the samples. Within assemblage A, subassemblage AII was detected in 8/32 (25%) samples (Table 4.3). Subassemblage AI was not identified. Within assemblage B, two genotypes were detected and both were consistent with the previously described BIII and BIV subassemblages at the *gdh* locus (Read et al., 2004). For consistency, the assemblage B genotypes detected in this study have been labelled as BIII or BIV. Both genotypes were detected separately in samples: BIII (3/32 samples, 9.4%); BIV (12/32 samples, 37.5%) and; as mixed samples of BIII and BIV (9/32 samples, 28.1%) (Table 4.3). All four infection profiles (AII, BIII, BIV, mixed BIII/BIV) were observed at rounds 1 and 2 of collection (Table 4.3).

Of the 32 samples genotyped at the *gdh* locus, 28 were also positive by 18S rRNA PCR and of these, 24 were successfully genotyped by 18S rRNA DNA sequencing. Four samples were positive at the *gdh* locus only, but could not be confirmed by 18S rRNA PCR or microscopic examination. Comparison of the *gdh* and 18S rRNA genotyping results showed that assignment of assemblage A and B by *gdh*-T-RFLP and 18S rRNA DNA sequencing was concordant.

4.4.5 *Giardia* positivity and identification of *Giardia* assemblages among participants that contributed samples at two collection rounds

Table 4.3. Distribution of *Giardia duodenalis* assemblages A and B subtypes identified by *gdh* PCR amplification and terminal-RFLP for 32 samples (27 participants) collected from children living in a remote Indigenous community in the Northern Territory.

	Faecal sample data (*n* = 32)				**Participant data (*n* = 27)**					
	Collection round				Age (years)			Gender		
Assemblage A and B subtypes	Round 1	Round 2	Round 3	Total (%)	0 - <5	5 - <10	10 – <15	Males	Females	Total (%)
Subassemblage AII	4	2	2	8 (25.0)	7	1	0	5	3	8 (29.6)
BIII*	1	2	0	3 (9.4)	0	1	2	0	3	3 (11.1)
BIV*	4	8	0	12 (37.5)	10	0	0	7	3	10 (37.0)
Mixed BIII/BIV*	4	5	0	9 (28.1)	5	1	0	3	3	6 (22.2)
Total	13	17	2	32	22	3	2	15	12	27

*BIII and BIV subtypes represent two assemblage B *gdh* genotypes that were identified in this study, and were consistent with the previously described *gdh* BIII and BIV subassemblages Read et al., 2004).

Thirteen participants contributed samples at two collection rounds (previously described). Of these, 8/9 participants who were 18S rRNA positive at the second collection round were also positive at the first round, and may represent ongoing infections. Of these, 4 participants positive at collection rounds 1 and 2, and one participant positive at collection rounds 2 and 3 were successfully genotyped by 18S rRNA sequencing. All 5 participants were genotyped as assemblage B at both collection rounds where samples were contributed. The *gdh* T-RFLP data showed that 3 of these participants had the same assemblage B subtypes (BIV, mixed BIII/BIV) at both rounds 1 and 2. To determine if the presence of ongoing infections contributed to the significant increase in 18S rRNA positivity at collection round 2, these participants were removed from the round 2 data and the above analyses were repeated. The results showed that the increase in *Giardia* 18S rRNA positive samples in round 2 of collection remained statistically supported (data not shown).

4.5 Discussion

In this study we investigated the prevalence of *Giardia*, and genetic subtypes present in children living in a remote Indigenous community in the Northern Territory. Screening by direct microscopy and 18S rRNA PCR amplification showed that *Giardia* was highly prevalent (66.7%) in faecal samples collected over an 18 month period. The high prevalence of *Giardia* detected in this study is similar to high rates (65%) of *Giardia* previously reported for children living in remote Indigenous communities in Australia (Reynoldson et al., 1998). The highest prevalence of *Giardia* was found in 0 - <5 year olds, which is similar to that found in low prevalence regions of Australia, where infants aged 0 – 4 years are the most commonly affected group by sporadic giardiasis (Yang et al., 2010; NSW Health Notifiable Conditions Information Management System).

PCR for the 18S rRNA gene proved to be the most sensitive method to detect *Giardia* in the faecal samples. Of the 58 positive samples, 41.4% of these were detected as positive by microscopy, whilst 93.1% were detected as positive using the 18S rRNA PCR. The sensitivity of PCR detection also differed between the two *Giardia* loci that were examined. Differences in detection rates for microscopy and PCR screenings are expected due to intermittent and/or low parasite shedding, DNA polymerase inhibitors in faecal material, and differences in gene copy number for the *gdh* and 18S rRNA loci (Caccio and Ryan, 2008; Ignatius et al., 2012; Ghosh et al., 2000). Previous studies in remote Indigenous communities of Australia have used microscopy as a preliminary screening tool to determine *Giardia* positivity, and select samples for downstream molecular analyses (Hopkins et al., 1997). The results of the present study, however, demonstrate that a large proportion (58.6%) of positive cases were only detectable by 18S rRNA PCR. Although detection of *Giardia* DNA by PCR is not direct evidence of an established infection, PCR is highly sensitive (Waldron et al., 2011b; Traub et al., 2005a). Similar screening results have been reported from children living in high prevalence regions of Rwanda, and may reflect low parasite shedding due to constant exposure and chronic infections (Ignatius et al., 2012).

Giardia was detected in faecal samples from all three collection periods over the 18 month study and high detection rates were maintained over time. A significant increase in PCR positivity, from 45.7% to 70.2%, was detected between collection rounds 1 and 2, and was consistent with an increase in positivity among the 0 - <5 year age group over this time. No differences in the frequency of positive cases among males and females, and among assemblage A and B between collection rounds 1 and 2 were detected. Our results suggest an overall increase in PCR positivity during the study; however, further sampling in this community would be required to resolve this. Sample sizes for older age groups were small and participants included in this study were self selected. Participants may

have been more likely to provide a faecal sample if gastrointestinal symptoms were present, but additional information pertaining to participant symptoms was not available for this study. It is unknown if PCR positive cases represent established infections, or if the increase in positivity at round 2 represents an increase in *Giardia* infections in this community. In low prevalence regions of Australia, *Giardia* infections fluctuate across demographic groups and with seasonal changes (NSW Health Notifiable Conditions Information Management System) Factors influencing prevalence in remote Indigenous communities in the Northern Territory are largely unknown due to the listing of *Giardia* as non-notifiable. Our results demonstrate that screening by both microscopy and 18S rRNA PCR is beneficial to accurately determine *Giardia* prevalence, the presence of established infections, and to identify demographic groups within the community that are most at risk from giardiasis.

DNA sequencing analyses of the 18S rRNA gene showed that all *Giardia* cases in this remote community were caused by assemblages A or B. *Giardia duodenalis* assemblage B was most commonly identified overall (75%), at all time points, in all age groups, and in both genders. The predominance of assemblage B concords with other studies from Australia (Hopkins et al., 1997; Yang et al., 2010) Further genotyping at the *gdh* locus showed that a diversity of genetic subtypes were present. Within assemblage A, only subassemblage AII was identified. Subassemblage AII is most commonly identified in humans (Yang et al., 2010) and anthroponotic transmission is likely. Subassemblage AI, the most common human subassemblage found in other animals (Ryan and Caccio, 2013; Yang et al., 2010) was not identified in this study. Our results demonstrated that two assemblage B genotypes, consistent with previous *gdh* BIII/BIV descriptions were detected in this community, and both genotypes were detected separately, and as mixed samples. Although designation of isolates to subassemblages BIII or BIV is problematic, an accurate system to classify assemblage B isolates that enables comparison between studies is

currently not available. All four types of *Giardia* (subassemblage AII, BIII, BIV, and mixed samples of BIII/BIV) were identified and persisted in the community over a 12 month period.

Detection of mixed genetic variants in 28% of cases is high when compared to data from low prevalence regions of Australia. Studies of sporadic giardiasis in Western Australia and New South Wales have detected mixed assemblage B samples in less than 6% of samples screened (Asher et al., 2012b; Yang et al., 2010). The larger proportion of mixed samples detected in this study may be indicative of environmental contamination and high frequency transmission of different *G. duodenalis* subtypes, and frequent host contact due to overcrowded living conditions, which may contribute to the higher prevalence of mixed infections. The presence of mixed samples can also be explained by nucleotide sequence heterogeneity among assemblage B isolates (Caccio and Ryan, 2008; Caccio and Sprong, 2010). The mechanisms that produce sequence heterogeneity are unresolved, but genetic diversity may be more prevalent in highly endemic environments due to increased competition and selective pressures (Thompson and Monis, 2012).

The results of this study have demonstrated high detection rates of *Giardia* in children living in a remote Indigenous community in the Northern Territory. Gastrointestinal infections remain a significant cause of morbidity in remote Indigenous communities and the burden of infectious diseases extends beyond childhood (McDonald and Bailie, 2010). The high proportion of positive cases detected in this study among children confirms previous reports from remote Indigenous communities. These results highlight the need for further research to understand the contribution of giardiasis to chronic gastrointestinal disease, and to examine links between PCR positive cases and clinical outcomes.

Giardia duodenalis assemblage B dominates transmission in this community. A variety of *G. duodenalis* subtypes persist, and reinfection of children with different genetic

variants is possible. The similar frequencies of assemblage A and B in this study to other Australian communities warrants further investigation to assess whether disease dynamics are similar between communities, despite differences in prevalence. In contrast, conditions specific to remote Indigenous communities may enhance mixed or heterogeneous infections.

Knowledge of parasite prevalence, infectious subtypes, and community dynamics that enhance transmission are required to address the continuing burden that gastrointestinal diseases impose on children in remote Indigenous communities.

4.6 Acknowledgements

We thank Therese Kearns and Jennifer Shield for providing the samples for this study.

4.7 Supplementary Figure 4.1

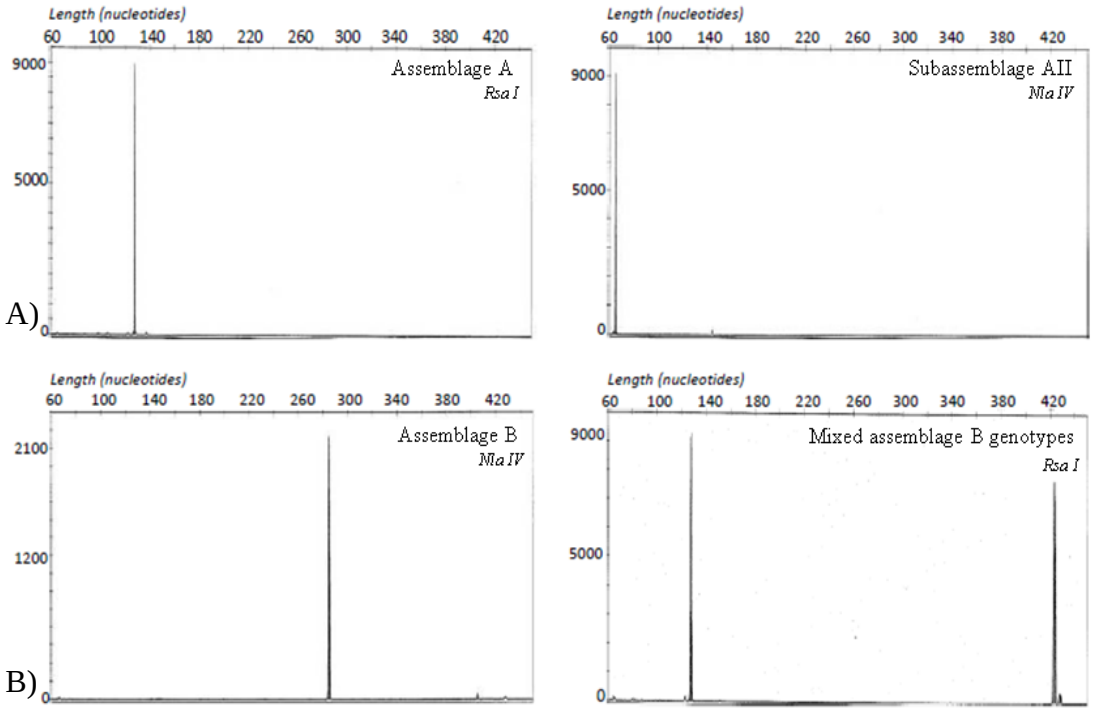

Figure 4.1 Electropherograms showing the two most common profiles detected from remote community DNA samples, a) assemblage A, and subassemblage AII following digestion with *Nla IV* and *Rsa I* digestion; and b) assemblage B, and mixed BIII/BIV genotypes following digestion with *Nla IV* and *Rsa I.*

5: Multilocus next generation sequence typing of *Giardia* from children living in a remote Indigenous community of Australia

A. J. Asher [a], M. L. Power [a]

Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia

Keywords: *Giardia duodenalis*, assemblages, subtypes, next generation sequencing

5.1 Abstract

Giardiasis is a communicable gastrointestinal disease caused by the protozoan parasite, *Giardia duodenalis*. Prevalence of *Giardia* among Indigenous children living in remote communities of Australia is high, and comparable to rates observed in developing nations. In this study we investigated the genetic diversity of *Giardia* contributing to cases among children from a remote Indigenous community of the Northern Territory. Next generation sequencing was applied to a multilocus protocol for four *Giardia* genes (18S rRNA, *gdh*, *tpi*, *β-giardin*), to identify *Giardia* assemblages, subassemblages, and subtypes. Analyses of 17 remote community DNA samples detected assemblages A in 35.3% (6/17), and B in 76.5% (13/17) of cases, including two mixed assemblage samples. Among these, three assemblage A subtypes, including AII-1, AII-2, and AII-3, and five separate assemblage B multilocus genotypes were identified. Our results indicate multiple genetic types of assemblages A and B contribute to cases amongst Indigenous children in this high prevalence community. Further research is required to evaluate subtype prevalence and elucidate disease links and transmission cycles relative to these subtypes, which will be essential for future disease management and control.

5.2 Introduction

Giardiasis is a leading cause of human gastrointestinal disease worldwide, and prevalence is highest in developing nations and disadvantaged regions (Feng and Xiao, 2011; Savioli et al., 2006). The disease is caused by infection of the small intestine with the protozoan parasite *Giardia duodenalis*, which is transmitted between susceptible hosts via the faecal-oral route (Ankarklev et al., 2010). Symptoms of giardiasis are highly variable, ranging in duration and severity, and may present as acute, intermittent, asymptomatic, or chronic gastrointestinal disease (Cotton et al., 2011). Globally, children are the highest risk group for acquiring giardiasis, and are most at risk from the clinical manifestations of the disease (United States Environmental Protection Agency, 1999). In children, acute disease can produce severe diarrhoea requiring hospitalisation, whilst chronic giardiasis is linked to malnutrition, wasting and developmental impairments (Thompson, 2008).

In Australia, prevalence of *Giardia* among Indigenous children living in remote communities is estimated at 32% to 65% (Meloni et al., 1993; Reynoldson et al., 1998; Thompson et al., 2001), and are among the highest childhood prevalence rates in the world (*c.f.* Feng and Xiao, 2011). High rates of disease transmission are attributed to poor personal hygiene practices among children, and environmental conditions in remote communities that are conductive to faecal-oral parasite transmission (McDonald et al., 2008). These conditions contribute to an overall high prevalence rate (15% to 36%) of *Giardia* among Indigenous populations living in remote communities (Jones, 1980; Thompson et al., 2001) which exceeds Australia's national estimates of prevalence of 2% to 7% (Thompson and Monis, 2012).

Previous studies in remote Indigenous communities in Australia have identified genetically diverse types of *G. duodenalis* (assemblages A and B) (Asher et al., 2014; Hopkins et al., 1997). *Giardia duodenalis* is a species complex, consisting of eight genetic

assemblages that infect several mammalian host groups (Lasek-Nesselquist et al., 2010; Monis et al., 2003). Humans are only infected by assemblages A and B, and these assemblages are also infective to domestic animals and wildlife (Monis et al., 2003; Thompson and Monis, 2012).

Assemblages A and B are highly divergent genetic lineages, consisting of numerous clusters of genetically related isolates (Sprong et al., 2009). Several subassemblages within assemblage A (subassemblages AI to AIV), and two genotypes of assemblage B (BIII, BIV) have been identified (Read et al., 2004; Sprong et al., 2009). Subassemblage AI has been detected from human and non-human hosts, whilst subassemblage AII is restricted to humans (Vermeulen et al., 2015; Yang et al., 2010). *Giardia duodenalis* assemblage B has been isolated from numerous mammalian hosts, but genetic sub-structuring, and the extent of genetic diversity within assemblage B is unresolved (Thompson and Monis, 2012). It is unknown if BIII and BIV genotypes represent mixed *Giardia* infections, or heterogeneous *Giardia* infections due to genetic recombination events (*c.f.* Caccio and Sprong, 2010).

Most recently, analysis of *Giardia* isolates by multilocus typing has demonstrated that within subassemblages AI and AII, numerous subtypes exist (AI, subtypes AI-1 to AI-3; AII, subtypes AII-1 to AII-7) (Caccio et al., 2008; Sprong et al., 2009). These subtypes also display varying degrees of host specificity between human and non-human hosts, but knowledge of their occurrence among human populations is limited (Caccio et al., 2008; Sprong et al., 2009). Identification of these subtypes among human populations may elucidate transmission cycles or differences in disease risk factors and aetiology, which, to date, has not been resolved by typing to the assemblage and subassemblage level (Caccio et al., 2008; Monis et al., 2003; Read et al., 2004; Sprong et al., 2009).

The molecular epidemiology of *Giardia* cases, including the diversity of genetic subtypes contributing to high infection rates, among children living in remote Indigenous

communities in Australia is poorly understood. Identification of *Giardia* assemblages/subassemblages is achieved by molecular analyses of *Giardia* genes but few molecular studies have been performed in Australia's Indigenous communities (Asher et al., 2014; Hopkins et al., 1997; Hopkins et al., 1999). Previous studies of *Giardia* in remote communities have identified *Giardia* to the assemblage or subassemblage level (Asher et al., 2014; Hopkins et al., 1997). Identification of subtypes within subassemblages requires analyses of at least three *Giardia* genes (Caccio et al., 2008). Furthermore, understanding the population structure of individual *Giardia* cases requires highly sensitive molecular diagnostic tools, which are still emerging. It is unknown if children simultaneously harbour multiple *Giardia* subtypes, or if the population structure of *Giardia* differs between children, or if heterogeneous *Giardia* infections are common due to increased competitive and selective pressures.

Next generation sequencing (NGS) is a tool for targeted amplicon sequencing (Chen et al., 2015) which provides a potential platform to perform highly sensitive multilocus sequencing of *Giardia* isolates. Use of NGS removes the need for cloning of PCR amplicons prior to sequencing, which is expensive and laborious. Improvements in sequence read lengths, and the ability to generate numerous target sequences from individual samples has facilitated the application of NGS to epidemiological research (Boers et al., 2012; Chen et al., 2015; Grinberg et al., 2013). The technology is commonly applied to characterise microbial diversity in environmental samples (Pilloni et al., 2015); however, its application to examining parasite communities is fairly limited. NGS has been recently used to examine diversity of *Cryptosporidium parvum* isolates (Gringberg et al., 2013), and to examine nematode populations in kangaroos (Lott et al., 2015). Applied to *Giardia*, NGS has the potential to provide an affordable multilocus genotyping tool, and generate comprehensive sequence data, required to understand complexities in *Giardia* subtype diversity and relationships to disease outcomes.

In this study we applied NGS to a multilocus screening protocol, to investigate the genetic diversity of *Giardia* among children living in a high-prevalence, remote Indigenous community of Australia. The study aimed to assess the utility of this method for *Giardia* typing in an epidemiological context; to perform NGS typing to the subassemblage and subtype level, and to assess the diversity of *Giardia* populations contributing to individual cases, and to cases among children in the community.

5.3 Materials and Methods

5.3.1 Ethical approval for sample screenings

DNA samples analysed in this study were extracted from human faecal samples, collected from humans living in a remote Indigenous community of the Northern Territory, Australia (Kearns et al., 2009). Ethics approval for the primary study (Kearns et al., 2009), including the collection and analysis of the faecal samples was provided by the Human Research Ethics Committee, Menzies School of Health Research, with additional approval provided by the Macquarie University Human Research Ethics Committee.

5.3.2 Preparation of *Giardia* positive controls to test multilocus PCR protocols

Prior to screening remote Indigenous community DNA samples, PCR protocols for multilocus amplification were standardised using a set of positive controls that included DNA extracted from a pure culture of *Giardia* trophozoites, kindly provided by the University of Technology, Sydney; and DNA extracted from *Giardia* positive human faecal samples (Asher et al., 2012b). DNA extractions for trophozoites and human faecal samples were performed using an ISOLATE Fecal DNA Kit (Bioline, Sydney, Australia) as previously described (Asher et al., 2012a). Positive controls were deemed PCR competent for *Giardia* by amplification of a 175 base pair region of the small subunit ribosomal rRNA (18S rRNA), and a 430 bp region of the glutamate dehydrogenase (*gdh*)

locus. Both genes were amplified following previously described protocols (18S rRNA, Hopkins et al., 1997; Read et al., 2002; *gdh*, Read et al., 2004), and with previously optimised PCR chemistry (Asher et al., 2012a; Sulaiman et al., 2003). All PCR reactions were performed in an Eppendorf Mastercycler (Eppendorf, North Ryde, Australia), and PCR amplicons analysed by agarose gel electrophoresis (2% w/v, 100V, 30 min), in TBE (Tris, EDTA, boric acid, pH 8.0) using SYBR safe (Invitrogen, Mulgrave, Australia). Results were visualised under UV light, with product sizes estimated using a HyperLadder™ II DNA marker (Bioline, Sydney, Australia).

5.3.3 Multilocus PCR amplification, optimisation of PCR conditions and initial PCR screenings of remote Indigenous community DNA

For multilocus typing by next generation sequencing, four *Giardia* genes were selected for PCR amplification. The 18S rRNA (PCR conditions described above) was chosen for NGS, to compare remote community NGS screenings with Sanger sequencing, which was performed for the same DNA samples during a previously study (Asher et al., 2014). Three additional genes, *gdh*, triosephosphate isomerase (*tpi*), and beta-giardin (*β-giardin*) were chosen to enable subtype identification according to Caccio et al. (2008). For each gene, PCR protocols were chosen to generate forward and reverse NGS reads that targeted the polymorphic positions of the *gdh, tpi,* and *β-giardin* genes described by Caccio et al. (2008). The PCR protocols chosen included; 530 base pair region of the *gdh* and *tpi* genes, and a 511 base pair region of the *β-giardin* gene.

Nested PCR reactions were performed (25 µl primary and secondary reaction volumes) for the *gdh*, *tpi*, and *β-giardin* loci, using positive DNA controls (section 5.3.2), and PCR primers and thermocycling conditions previously described for each gene (listed in Table 5.1a). Due to the high variability in PCR performance of other *Giardia* PCR protocols (Asher et al., 2012a), conditions for the new genes/regions were tested using

positive DNA controls, and by varying concentrations of reaction components and thermocycling conditions (Table 5.1a). All results were analysed by agarose gel electrophoresis as described above (section 5.3.2).

The PCR conditions that most consistently amplified positive DNA controls were used to perform initial PCR screenings of remote community DNA samples (n = 58), to determine PCR positivity at the *gdh*, *tpi*, and *β-giardin* loci. 18S rRNA PCR positivity was already known for remote community DNA samples due to prior screenings (Asher et al., 2014). All PCR reactions for the *gdh*, *tpi*, and *β-giardin* genes were prepared with Platinum DNA polymerase High Fidelity in 25μl reaction volumes, using 20pmol/μl primers, 1.5mM magnesium, 200 nM dNTPs, and using 1μl DNA for the primary reactions followed by 2 - 3μl of primary product in the secondary reactions. For *gdh* PCR, annealing was performed at 62°C (primary reaction) and 58°C (secondary reaction). Amplification of the *tpi* locus was performed at 62°C (primary reaction) and 50°C (secondary reaction) annealing, and 40 cycles for the primary and secondary reactions. *β-giardin* amplification was performed using previously described thermocycling conditions (Lalle et al., 2005), but using the modified primer sequence described by Delport et al. (2014). All amplicons were visualised by agarose-gel electrophoresis as described above (section 5.3.2).

5.3.4 Preparation of the next generation sequencing library

For NGS, a linker sequence, compatible with downstream NGS assays (Research and Testing Laboratory, Texas) was attached to secondary *gdh*, *tpi*, and *β-giardin* primers, and primary 18S rRNA primers. The primary 18S rRNA amplicon (290bp) was selected for NGS due to the short length of the secondary 18S rRNA product (175bp). For all genes (*gdh*, *tpi*, *β-giardin*, 18S rRNA) the linker sequence was attached to the 5-prime end of forward and reverse primers, which generated secondary oligonucleotides, with base pair

Table 5.1 PCR chemistry tested for amplification of three *Giardia duodenalis* genes (*gdh*, *tpi*, *β-giardin*); (A) for regular PCR amplification of genes; and (B) for amplification of a next generation sequencing library.

		***Giardia* genes**		
DNA polymerases	**Conditions tested**	*gdh*	*tpi*	*β-giardin*
A. Regular PCR	DNA (μl)	1-3	1-5	1-3
	Magnesium (mM)	1.5	1.5-4.0	1.5
1. GC-RICH PCR System, dNTPack (Roche Diagnostics, IN, USA)	Resolution Solution (M)	0.5	0.5-2.5	0.5
	Primers (pmol/μl)	20	20	10 &20
2. Platinum DNA polymerase (Life Technologies)	Gradient annealing (°C)	56-64	50-66	Protocol
	Touchdown annealing (°C)	60-50	60-50	-
3. Tth plus DNA polymerase (Fisher Biotec, Wembley, Australia)	Cycles	Protocol	35-40	Protocol
	Primary PCR primers	Gdh1/Gdh2	AL3543/AL3546	G7/G759
4. Platinum DNA polymerase High Fidelity (Life Technologies)	Secondary PCR primers	Gdh3/Gdh4	AL3544/AL3545	*modified β-g F/R
	Protocol reference	Caccio et al., 2008	Sulaiman et al., 2003	Caccio et al., 2002 Lalle et al., 2005 *Delport et al., 2014
B. Secondary PCRs for MiSeq NGS Platinum DNA polymerase High Fidelity & IDT primers	DNA (μl)	3	2-3	2-3
	Dilutions	Neat/1in2	Neat/1in2	Neat/<1in8
	Annealing (°C)	58-60	52-68	55-77

lengths between 40 bp and 45 bp. All modified primers were synthesised by Integrated DNA Technologies (IDT) (Coralville, Iowa, USA).

Amplification of each gene (*gdh*, *tpi*, *β-giardin*, 18S rRNA) with IDT primers was initially tested using positive DNA controls, and the optimal PCR reaction conditions determined above (section 5.3.3). For the *gdh*, *tpi*, *β-giardin* genes, initial PCR results showed amplification with the longer primers produced non-specific high molecular weight PCR products (~ 1000bp), in addition to the gene specific products. Low molecular weight (< 80bp) products were produced during 18S rRNA PCR with IDT primers. PCR annealing temperatures were re-tested for each gene (Table 5.1b), and PCR amplification with IDT primers were performed at 60°C (*gdh*), 64°C (*tpi*), 69°C (*β-giardin*) and 67°C (18S rRNA).

For remote community samples positive at each of the four genes, primary PCR products were re-amplified using the modified IDT primers and optimal conditions described above. All positive PCR products were purified using a QIAquick PCR purification kit or a MiniElute PCR purification kit (Qiagen, Melbourne, Australia). The DNA concentration of purified products was measured using a Qubit 2.0 Fluorometer (Life Technologies, Mulgrave, Australia), and BR dsDNA assay (Life Technologies, Mulgrave, Australia). For all purified products, DNA concentration was normalised to approximately 20ng/µl. Next generation sequencing reactions were performed using an Illumina MiSeq platform and 20K assay, which were performed by Research and Testing Laboratory, Texas.

5.3.5 NGS: raw data analyses, evaluation of read coverage and assembly method

All raw next generation sequence reads were analysed in Geneious Pro, version 8.1.5 (Biomatters Ltd, Auckland, New Zealand). For each sample, forward and reverse raw sequence reads were first identified and separated by *Giardia* gene (*gdh*, *tpi*, *β-giardin*, 18S rRNA), using the barcode separation function in Geneious Pro. Using this function, the nucleotide sequence of the forward and reverse primers up until the first degenerate base, were entered as barcodes, allowing one mismatch per barcode. Raw sequence reads were then trimmed to an error probability limit of 0.01, and all reads with lengths below 99 base pairs were removed from further analyses.

To perform all genotyping analyses of NGS reads, previously described *G. duodenalis* nucleotide sequences (n = 57) were retrieved from the NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html), and included sequences representing *G. duodenalis* assemblages, subassemblages, and subtypes (Table 5.2). Representative sequences for subtypes A1 to A6 (Table 5.2a) were aligned by ClustalW (Thompson et al., 1994) in Geneious Pro to determine the minimum base pair length of forward and reverse NGS reads required to identify samples to the subtype level at the *gdh*, *tpi, β-giardin* genes. For each gene, previously described nucleotide positions (Caccio et al., 2008), specific for subtypes A1 to A6, were mapped to the secondary PCR amplicon region, and the minimum length of NGS reads (after trimming) required for subtyping was determined as follows; *gdh* full length 517 bp with reads of 259 bp each; *tpi*, full length 508bp with reads of 254bp each; *β-giardin* full length 488bp with reads of 244bp each.

To determine the optimal method for read assembly all forward and reverse NGS reads were paired and read coverage (read base pair length and number of reads) was assessed for each gene and sample. Paired reads were assessed in three groups, including; forward and reverse pairs with minimum base pair lengths of 259bp, 254bp, and 244bp, for each gene as described above; one read per pair with minimum base pair lengths of 259bp,

254bp and 244bp; and all forward and reverse pairs with minimum base pair length of 200bp. Due to low average read lengths of *gdh*, *tpi*, and *β-giardin* reads among all samples, paired reads were selected for assembly if one read (forward or reverse) per pair met the minimum length requirement for each gene. For the 18S rRNA, NGS reads met the minimum length requirement for this gene (reads of 134bp each) due to the shorter base pair length of the target PCR amplicon (268bp after trimming).

5.3.6 Construction of read assemblies, evaluation of read diversity and identification of *Giardia* assemblages, subassemblages and subtypes

To identify *Giardia* assemblages for the *gdh*, *tpi*, and *β-giardin* genes, paired read files (for each gene within each individual sample) were assembled by directly mapping reads to reference sequences obtained from the NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html) (Table 5.2b). Due to short base pair lengths of *gdh*, *tpi*, and *β-giardin* reads, it was not possible to merge reads or construct contiguous sequences from each of the paired reads. To increase coverage and accuracy of the analyses, mapping of *gdh*, *tpi*, and *β-giardin* reads was limited to one representative sequence for each *Giardia* assemblage (Table 5.2b).

Table 5.2 Accession numbers and references for representative sequences of *Giardia duodenalis* assemblages, subassemblages, and subtypes that were obtained from the NCBI GenBank database (http://www.ncbi.nlm.gov/genbank/index.html), and used for a) assemblage A subtyping; b) mapping of next generation sequences; and c) additional assemblage B sequences used to assess assemblage B consensus sequences generated in the study.

GenBank accession numbers for representative nucleotide sequences for *Giardia* assemblages, subassemblages and subtypes								
a) Subtyping		**Accession numbers**			**References**[a]			
Subas.	Subtype	*gdh*	*tpi*	*β-giardin*	*gdh*	*tpi*	*β-giardin*	
AI	A1	AY178735	L02120	X85958	Ey et al., unpub.	Mowat et al., 1994	Holberton & Marshall, 1995	
	A5	M84604	AB509383	AB469365	Yee & Dennis, 1992	Abe et al., 2010	Abe et al., 2010	
AII	A2	AY178737	U57897	FJ971422	Ey et al., unpub.	Baruch & Adam, 1996	Kosuwin et al., 2010	
	A3	EU278608	EU041754	EU188635	Lalle et al., 2009	Lalle et al., 2009	Cooper et al., 2007	
	A4	EF507657	AB509382	-	Souza et al., 2007	Abe & Tanoue, unpub.	-	
AIII	A6	EU637582	DQ650648	DQ650649	Caccio et al., 2008	Lalle et al., 2007	Lalle et al., 2007	
b) Mapping		**Accession numbers**			**References**			
Assem.	18S	*gdh*	*tpi*	*β-giardin*	18S rRNA	*gdh*	*tpi*	*β-giardin*
A	AF199446	AY178737	U57897	AY072724	Thompson et al, 2000	As above	As above	Caccio et al., 2002
B	AF199447	AY178739	AY368169	AY072727	Thompson et al, 2000	Ey et al., unpub.	Sulaiman et al., 2003	Caccio et al., 2002
C	AF199449	EF507623	AY228641	JF422719	Thompson et al, 2000	Souza et al., 2007	Sulaiman et al., 2003	Paz e Silva et al., 2012
D	AF199443	EF507622	DQ220289	AY545648	Thompson et al, 2000	Souza et al., 2007	Weilinga et al., unpub.	Lalle et al., 2005
E	AF199448	AY178740	AY655705	AY655703	Thompson et al, 2000	Ey et al., unpub.	Trout et al., 2004	Trout et al., 2004
F	AF199444	EF507593	AF069558	AY647264	Thompson et al, 2000	Souza et al., 2007	Monis et al., 1999	Lalle et al., 2005
G	AF199450	AY178745	-	-	Thompson et al, 2000	Monis et al., 1999	-	-
H	-	GU176081	-	-	-	Lasek-Nesselquist et al., 2010		-
G. ardeae & *G. muris*	AF113895	AF069060	AF069564	EF455599	Monis et al., 1999	Monis et al., 1999	Monis et al., 1999	Lebbad et al., 2008
c) Additional		**Accession numbers**			**References**			
		gdh	*tpi*	*β-giardin*	*gdh*	*tpi*	*β-giardin*	
G. duodenalis Assem. B		AY178756	AY228628	AY647266	Ey et al., unpub.	Sulaiman et al., 2003	Lalle et al., 2005	
		AY178754	AY228636	DQ090526	Ey et al., unpub.	Sulaiman et al., 2003	Robertson et al., 2006	
		AY178750	AY268630	AY072728	Ey et al., unpub.	Sulaiman et al., 2003	Caccio et al., 2002	
		AY178738	-	-	Ey et al., unpub.	-	-	

[a] 'unpub.' refers to sequences deposited in the NCBI database from unpublished observations.

For the 18S rRNA, paired reads were merged in Geneious Pro using the plugin FLASH (Magoc and Salzberg, 2011), and contiguous 18S rRNA sequences were assembled by mapping to 18S rRNA reference sequences (Table 5.2b). All assemblies (*gdh*, *tpi*, and *β-giardin*, 18S rRNA) were constructed using Geneious mapper, with up to five iterations of assembles, and using the highest sensitivity settings.

To construct consensus sequences from each assembly (*gdh*, *tpi*, and *β-giardin*, 18S rRNA), and distinguish diversity from read errors, low coverage regions within each assembly were identified as areas where coverage was below two standard deviations of the mean, and were excluded from calling polymorphic nucleotide positions. Polymorphic nucleotide positions outside of low coverage regions were called using a minimum variant frequency of 0.25, a maximum variant p-value of 10^{-6}, and a minimum strand bias p-value of 10^{-5} (>65% bases). Consensus sequences were then called from the paired reads in each assembly and the analyses excluded the reference sequence that was used for mapping. Only the highest quality bases for each nucleotide position were accepted in the final consensus sequence. For *gdh*, *tpi*, and *β-giardin* genes, constructed assemblies containing fewer than 100 reads were removed from further analyses due to large sequence gaps, which could not be accurately resolved in small assemblies. *Giardia* assemblages, subassemblages, and subtypes were identified from all consensus sequences by BLASTn sequence searches of the NCBI database. Sequence homology and variation among assemblage B samples were investigated by ClustalW (Thompson et al., 1994) nucleotide sequence alignment of consensus sequences and representative sequences using Geneious Pro. Additional representative sequences (Table 5.2c) for assemblage B were retrieved from GenBank and included in the analyses.

5.3.7 Phylogenetic analyses of 18S rRNA, *gdh*, *tpi* and *β-giardin* consensus sequences

Phylogenetic analyses were performed for all consensus sequences generated in the study. Consensus sequences and reference sequences obtained from GenBank were re-aligned by Clustal W in *MEGA* version 5 (Tamura et al., 2011). *Giardia ardeae* and *Giardia muris* nucleotide sequences for each of the four loci were included in the analyses as outgroups (Table 5.2b).

Phylogenetic analyses was performed separately for a total of five groups of samples, including; all 18S rRNA sequences; *gdh* (assemblage A); *tpi* (assemblage A); *β-giardin* (assemblage A). Due mixed genotyping results for assemblage B samples, *gdh*, *tpi,* and *β-giardin* sequences were concatenated, to assess the diversity of samples over three loci. All consensus sequences and reference sequences were concatenated in Geneious Pro and exported to *MEGA5* for re-alignment by Clustal W. For all five groups, nucleotide substitution models were tested in *MEGA5*, and the optimal parameters for phylogenetic inference were determined from Akaike Information Criterion corrected (AICc) values. All phylogenetic trees were constructed by maximum likelihood in *MEGA5* with bootstrap analyses (1000 replicates). Distance models used for each group were as follows; 18S rRNA, Tamura-Nei (TN93); assemblage A *gdh*, Tamura-Nei with gamma distribution (TN93+G); assemblage A *tpi*, General Time Reversible model and gamma distribution with invariant sites (GTR+G+I); assemblage B concatenated *gdh, tpi,* and *β-giardin*, GTR+G+I. Representative nucleotide sequences for *Giardia ardeae* and *Giardia muris* across all three loci (*gdh, tpi,* and *β-giardin*) were not available from the NCBI GenBank database, and as such analyses of concatenated sequences at three loci did not include an outgroup. Similarly, representative sequences for *Giardia duodenalis* assemblage H at the 18S rRNA, *tpi,* and *β-giardin* loci, and *G. duodenalis* assemblage G at the *tpi,* and *β-giardin* loci, were not available and could not be included in the analyses.

5.4 Results

5.4.1 Optimisation of multilocus PCRs and PCR screenings for remote community DNA samples

PCR reaction conditions for three *Giardia* genes (*gdh*, *tpi*, and *β-giardin*) were tested using positive DNA controls and four DNA polymerases. For the *β-giardin* gene, a 511 bp PCR product was successfully amplified with all four DNA polymerases and reaction conditions that were tested (Table 5.1a). Amplification of the *gdh* (530 bp) and *tpi* (530 bp) genes required further optimisation of the thermocycling protocols. For *gdh* and *tpi* amplification, initial tests of DNA controls showed amplification of specific and non-specific products. Improved *gdh* yields and single bands were achieved by increasing annealing temperatures (section 5.3.3). For *tpi*, specific PCR products corresponding to the expected amplicon size were amplified from positive DNA controls, using the optimal conditions described in section 5.3.3 and by preparing PCR reactions at room temperature. *Tpi* PCR amplification under optimal conditions, however, remained problematic, with smearing and multiple non-specific products intermittently amplifying with the target region. Results of other assays tested for *tpi* PCR amplification (Table 5.1a) were less consistent, and positive DNA controls frequently failed to amplify under repeat analyses.

A total of 58 remote community DNA samples were screened at the *gdh*, *tpi*, and *β-giardin* loci, using the optimal PCR conditions described above (section 5.3.3). Of these, 27/58 (46.6%) were positive at the *gdh* locus, and 26/58 (44.8%) were positive at the *β-giardin* locus. For *tpi* PCR amplification, only 7/58 (12.1%) could be confirmed as PCR positive due to multiple non-specific products that amplified for a further 22 samples. DNA samples had been previously screened at the 18S rRNA locus as part of a previous study (Asher et al., 2014) and 18S rRNA PCR positivity among the samples was already known.

Re-amplification of primary PCR products using IDT linker primers and PCR conditions described in section 5.3.4 was successful for a total of 17 samples. These

samples were successfully amplified at all four loci, producing a total of 68 PCR amplicons that were analysed by next generation sequencing.

5.4.2 Analyses of next generation sequencing data

A total of 3.4 million sequence reads were obtained by Illumina MiSeq sequencing. After identification of genes, quality control and filtering of reads, there were approximately two million reads with minimum sequence lengths of 100bp. The average base pair length for forward and reverse reads per gene ranged between 130-173bp. The majority of forward and reverse paired reads, for each gene, did not span full base pair length (after trimming) of the *gdh* (517bp), *tpi* (508bp), or *β-giardin* (488 bp) genes. For these genes, average sequence gaps between forward and reverse reads ranged between165bp to 270bp.

A suitable assembly method for *gdh*, *tpi*, and *β-giardin* reads was determined by assessing paired reads by base pair length and coverage of the gene (Table 5.3). Extracting paired reads based on one read per pair meeting the minimum length requirement (section 5.3.5) produced the highest number of reads, which was on average between 2000 and 4000 reads per gene and for each sample (Table 5.3). This method was deemed most suitable for assembly due to the high number of reads isolated. A high number of reads per assembly was required due to the short base pair length of all forward and reverse reads, and large sequence gaps which could not be resolved in small assemblies (<100 reads).

For the 18S rRNA, construction of contiguous sequences from each pair of forward and reverse reads was possible due to the overall shorter length of the target PCR region (268bp after trimming). Merging of 18S rRNA reads, however, failed for the majority of paired reads. Further investigation by high sensitivity mapping showed that most 18S rRNA reads were mapped to areas outside of the target region, extending in the 3' to 5' direction from forward and reverse 18S rRNA primer sites, and these reads were excluded

Table 5.3 Number of next generation sequencing (NGS) reads that were recovered by three different read extraction methods, for the glutamate dehydrogenase (*gdh*), triosephosphate isomerase (*tpi*) and beta giardin (*β-giardin*) genes. The table shows the total reads recovered and mean number of reads recovered for the 17 remote community DNA samples that were screened by NGS.

	Read extraction method:		**Both paired reads meeting minimum half length/gene**			**One read/pair meeting minimum half length/gene**			**Both paired reads >200 bp**		
Gene	Full length[a] (bp)	Half length[b] (bp)	Total reads	% total	mean/ sample	Total reads	% total	mean/ sample	Total reads	% total	mean/ sample
Gdh	517	259	4	2.7	<1	64078	37.0	3769	9380	25.8	552
Tpi	508	254	142	95.9	8	39466	22.8	2322	26353	72.4	1550
β-giardin	488	244	2	1.4	<1	69716	40.2	4101	662	1.8	39
Total			148			173260			36395		

[a]Full length (bp) refers to the expected base pair length of the mapping region for each gene, after barcode separation which partially removed primer sequences.
[b]Half length (bp) refers to the minimum base pair length of forward and/or reverse reads required to cover half of the full base pair length determined for each gene.

from all analyses. For one sample (m14), no 18S rRNA sequences were recovered.

5.4.3 NGS: sequence mapping to *Giardia* reference sequences and analyses of sequence diversity

A total of 1989 contiguous 18S rRNA sequences, and a total of 174 262 reads/87131 paired *gdh*, *tpi*, and *β-giardin* reads were mapped to *Giardia* reference sequences representing assemblages A to H. Mapping was successful for 17 samples at three loci (*gdh*, *tpi*, and *β-giardin*), and 16/17 (94.1%) samples were successfully mapped at all four loci. For all samples, 18S rRNA *gdh*, *tpi*, and *β-giardin* sequences were mapped to *Giardia* assemblages A or B only, and no other *Giardia* assemblages were identified by sequence mapping. For 15/17 (88.2%) samples, 100% of the sequences from each sample were mapped to either assemblage A (4/17, 23.5% samples) or assemblage B (11/17, 64.7% samples), and two samples (2/17, 11.8%) were identified as containing both assemblages A and B sequences. For one mixed sample, 3.9% and 19.8% of the detected 18S rRNA and *tpi* reads were mapped as assemblage B, whilst all other reads for this sample were assemblage A. The second mixed sample was identified as containing assemblage B reads at three loci (18S rRNA, *gdh*, *tpi*), whilst all *β-giardin* reads were assemblage A.

A total of 69 assemblies were produced from sequence mapping for each gene and sample, and for each assembly, polymorphic nucleotide positions were examined to determine diversity among the reads. For assemblage A *β-giardin*, and *tpi* assemblies, no variation was detected among the reads within each assembly. For assemblage A *gdh*, three polymorphic sites were detected among all assemblies, and all assemblies contained the following polymorphic nucleotides; position 1 (C/T), 7 (C/T), and 517 (C/T) of the consensus sequence. These positions corresponded to the site of degenerate bases contained within the secondary *gdh* PCR primer sequences, and no other variable sites

were identified. Removal of primer sequences from read assemblies could not be performed for assemblage A samples, due to one *gdh* subtyping position (Caccio et al., 2008) occurring within the reverse primer sequence (513bp).

For assemblage B, further removal of primer sequences was performed for each gene and assembly, and final base pair lengths for each gene were 481bp *(gdh*), 490bp *(tpi*), and 475bp (*β-giardin*). Consensus sequences for each sample differed at 24 positions, 23 positions, and 13 positions of the *gdh*, *tpi*, and *β-giardin* loci respectively (Table 5.4a,b,c). Among these samples, polymorphic nucleotide positions were identified for 11/12 (91.7%) *gdh* samples, 8/13 (61.5%) *tpi* samples, and 7/11 (63.6%) *β-giardin* samples (Table 5.4a,b,c). Two types of polymorphic positions were identified among the assemblies, including: positions where two or more bases contributed equally to the position; and positions where a low frequency nucleotide (minimum variant frequency of 0.25) was detected in addition to a high frequency nucleotide. Further mapping of reads for each gene and sample to multiple assemblage B reference sequences produced small assemblies, and low quality consensus sequences which could not be used for further analyses. It was not possible to identify discrete populations within each assembly, or determine which populations corresponded to each individual polymorphic site in the consensus sequence. For this reason, only the highest quality and highest frequency nucleotide at each of the polymorphic positions were accepted for genotyping. Where nucleotides contributed equally to a position, both nucleotides were represented by a degenerate base code in the consensus sequence.

Table 5.4a) Diversity of assemblage B consensus sequences at the glutamate dehydrogenase gene (*gdh*) that were generated from next generation sequencing of remote community DNA samples. For each sample (m1 – m15), the table shows the base pair (*Bp*) position of variable nucleotides that were identified among the 481bp sequences. Positions with a 'y', or 'r' denote contribution of both c/t (y), or a/g (r) to that position. Nucleotides represented in brackets, are low frequency polymorphic nucleotides that were identified at each site, and were not included in the final consensus sequence. The consensus between samples at each position is denoted by '*C.*'

Bp	4	55	76	82	97	106	142	148	163	164	220	229
C	**G**	**C**	**T**	**C**	**T**	**C**	**C**	**A**	**G**	**T**	**G**	**C**
m1	.	y	c	c (t)	.	.	.	r	.	.	.	.
m2	.	.	t (c)	c (t)	.	.	.	.	.	t (a)	.	.
m3	g (a)	y	y	.	.	.	c (t)	.	g (a)	t (a)	.	.
m4	.	y	y	.	.	.	.	r	.	.	.	y
m6	.	t	c	.	.	.	.	g	.	.	.	.
m7	.	.	.	.	.	.	.	.	.	t (a)	.	.
m8	.	.	.	.	t	.	.	.	.	t (a)	.	.
m10	.	.	.	.	y	.	.	.	.	.	.	.
m12	.	.	.	.	.	.	t	.	.	t (a)	.	.
m13	.	.	t (c)	.	.	.	.	r	.	t (a)	.	.
m14	.	.	.	.	.	.	.	.	.	t (a)	a	.
m15	.	.	.	.	.	.	.	.	.	t (a)	.	.

Bp	235	259	319	343	350	370	403	412	427	445	457	469
C	**C**	**T**	**C**	**T**	**A**	**C**	**C**	**G**	**T**	**C**	**A**	**C**
m1	.	.	.	.	a	y	.	.	y	y	g	.
m2	.	.	.	.	g (a)	.	.	.	.	.	g (a)	.
m3	.	.	.	.	.	.	.	.	.	.	g	.
m4	.	c (t)	c (a)	.	r	.	.	.	.	.	r	y
m6	.	c	.	.	a	t	.	.	c	t	g	.
m7	.	.	a (c)	.	.	.	.	.	.	.	.	.
m8	.	.	.	.	.	.	.	.	.	.	.	.
m10	.	y	.	.	.	.	.	.	.	.	g (a)	y
m12	.	c	.	.	a	.	.	.	.	.	g	t
m13	c (t)	t (c)	.	t (g)	g (a)	.	.	.	.	.	g (a)	y
m14	.	.	.	.	.	.	.	a	.	.	g	.
m15	.	.	.	.	a	.	.	.	.	.	.	.

Table 5.4b) Diversity of assemblage B consensus sequences at the triosephosphate isomerase gene (*tpi*) that were generated from next generation sequencing of remote community DNA samples. For each sample (m1 – m15), the table shows the base pair (*Bp*) position of variable nucleotides that were identified among the 490bp sequences. Positions with a 'y', 'r', or 'm'denote contribution of both c/t (y), a/g (r), or a/c (m) to that position. Nucleotides represented in brackets, are low frequency polymorphic nucleotides identified that were at each site, and were not included in the final consensus sequence. The consensus between samples at each position is denoted by '*C.*

Bp	4	10	13	16	56	69	72	127	130	133	160	175
C	**A**	**T**	**A**	**G**	**C**	**A**	**C**	**G**	**C**	**C**	**G**	**G**
m1	.	.	.	.	t	.	.	.	t	t	a	.
m2	.	.	.	a	.	.	.	a	.	.	.	.
m3	g (a)	t (c)	.	.	.	a (g)	.	.	c (t)	c (t)	.	.
m4	g	y	.	.	.	.	.	.	y	y	.	.
m6	.	.	.	a	.	.	.	a	.	.	.	.
m7	.	.	.	.	t	.	.	.	t	t	a	.
m8	a (g)	.	.	a (g)	.	.	.	a	.	.	.	.
m10	.	.	.	.	.	.	.	.	.	.	.	.
m11	.	.	.	g (a)	.	.	..	.	.	.	.	.
m12	a (g)	.	.	g (a)	.	.	.	.	.	.	.	.
m13	r	.	r	g (a)	.	.	.	r	y	.	.	.
m14	.	.	.	g	t	.	.	.	t	t	.	a
m15	r	y	.	g	c (t)	r	.	.	t	t	.	g (a)

Bp	196	220	236	262	310	367	382	394	436	469	478
C	**C**	**G**	**C**	**A**	**G**	**A**	**C**	**G**	**A**	**C**	**G**
m1	.	a	.	.	.	.	.	.	.	.	a
m2	.	.	.	.	.	.	.	.	.	.	.
m3	.	.	.	.	.	r	.	.	.	.	.
m4	.	.	.	r	r	.	.	.	m	.	.
m6	.	.	.	.	.	.	.	.	.	.	.
m7	.	a	.	.	.	.	.	.	.	.	.
m8	.	.	.	.	g (a)	.	.	.	.	.	.
m10	.	.	.	.	.	.	.	.	.	.	.
m11	.	.	.	.	.	.	.	.	.	.	.
m12	t	.	.	.	g (a)	.	.	.	.	.	.
m13	.	r	.	.	g (a)	.	y	r	.	y	.
m14	.	.	.	.	.	.	.	a	.	.	.
m15	.	.	y	.	.	r	.	g (a)	.	.	.

Table 5.4c) Diversity of assemblage B consensus sequences at the beta giardin gene (*β-giardin*) that were generated from next generation sequencing of remote community DNA samples. For each sample (m1 – m15), the table shows the base pair (*Bp*) position of variable nucleotides that were identified among the 475bp sequences. Positions with a 'y', or 'r' denote contribution of both c/t (y), or a/g (r) to that position. Nucleotides represented in brackets, are low frequency polymorphic nucleotides identified that were at each site, and were not included in the final consensus sequence. The consensus between samples at each position is denoted by '*C.*'

Bp	7	46	55	73	79	94	118	160	169	199	246	325	409
C	**C**	**C**	**C**	**A**	**G**	**C**	**A**	**G**	**C**	**C**	**A**	**C**	**T**
m1	.	.	.	g	.	.	.	.	t	.	.	.	.
m2	.	.	t	.	.	.	.	r	.	y	.	.	.
m3	.	y	.	.	r	.	a (g)	.	.	y	.	c (t)	y
m4	.	.	.	.	.	.	.	.	.	t	.	.	.
m7	.	.	.	.	.	.	.	.	.	t	a (g)	.	.
m8	.	.	.	.	.	.	.	g (a)	.	c (t)	.	.	.
m10	c (t)	.	.	a (g)	.	.	.	.	.	t	.	.	.
m12	.	.	.	.	.	t	.	.	.	.	.	.	.
m13	c (t)	.	.	r	.	.	.	.	.	y	.	.	.
m14	.	.	.	.	.	.	.	.	.	.	.	.	.
m15	.	.	.	.	.	.	r	.	.	t	.	y	.

5.4.4 Identification of *Giardia* assemblages, subassemblages and subtypes from consensus sequences

Identification of *Giardia* assemblages, subassemblages, and subtypes from consensus sequences for each locus was achieved by Blastn sequence searches of the NCBI database. Search results corresponded to mapping results, and 4/17 (23.5%) samples were identified as assemblage A, 11/17 (64.7%) were assemblage B, and 2/17 (11.8%) samples were identified as assemblages A and B (Table 5.5 & Table 5.6). All assemblage A samples were subassemblage AII, and within this subassemblage, subtypes A2 and A3 were identified among *gdh*, *tpi*, and *β-giardin* sequences (Table 5.5). The overall subtype for

each assemblage A sample was determined by comparing genotyping results for 4/5 assemblage A samples that were typed across all three loci (*gdh*, *tpi*, and *β-giardin*). Two samples were identified as subtype AII-1, and AII-3, whilst profiles for three samples were consistent with subtype AII-2 (Lebbad et al., 2011).

Table 5.5 *Giardia* subassemblages and subtypes of assemblage A samples that were identified by multilocus next generation sequencing (NGS) of remote community DNA samples. The table shows genotyping data for the 18S rRNA, beta giardin (*β-giardin*), glutamate dehydrogenase *(gdh*), and triose phosphate isomerase (*tpi*) genes. All data presented are for consensus sequences (*gdh, tpi, β-giardin*) and contiguous sequences (18S rRNA) generated from NGS reads. A total of 4 samples were typed as containing only assemblage A reads. Sample m11 contained mixed reads of assemblages A and B, and genotyping data for this sample represents one (1) of the consensus sequences generated from this sample. Sample m6 was typed as assemblage B at the 18S rRNA, *gdh*, and *tpi* loci, and as assemblage A at the *β-giardin* locus.

Assemblage A	**Subtypes at each loci**				**Subtypes determined by three *Giardia* loci**	
Samples	18S rRNA	*gdh*	*tpi*	*β-giardin*	Subas.	Subtype
m5	A	A2	A2	A3	AII	AII-2
m6 (1)	-	-	-	A3	AII	-
m9	A	A3	A2	A2	AII	AII-3
m11 (1)	A	A2	A2	A3	AII	AII-2
m16	A	A2	A2	A3	AII	AII-2
m17	A	A2	A2	A2	AII	AII-1

For assemblage B, BLASTn results for 18S rRNA and *β-giardin* identified samples as assemblage B (Table 5.6). Analyses of *gdh*, and *tpi* sequences showed high sequence homology to GenBank sequences that have previously been identified as genotypes BIII, BIV, and BIV-like (Weilinga and Thompson, 2007). A total of 12/13 (92.3%) samples were typed at the *gdh*, and *tpi* loci, and among these samples, genotyping results across the two loci were consistent for 10/12 (88.3%). These samples that were identifiable (m6(2),

m15) were BIII (n = 1/12) and BIV-like (n = 1/12) (Table 5.6). Nucleotide sequence alignment of consensus sequences to additional assemblage B reference sequences (Table 5.2c) showed that the BIV-like genotypes consisted of a mixture of BIII and BIV specific nucleotides at typing positions for the *gdh* and *tpi* loci.

A total of 11 assemblage B samples were typed at three loci (*gdh*, *tpi*, and *β-giardin*), and of these, multilocus genotypes (MLGs) could be determined for five samples (m2, m7, m8, m12, m14). Polymorphic sites, where more than one nucleotide contributed to a single position occurred in the consensus sequences for six assemblage B samples, and it was not possible to accurately call MLGs from these samples. Analyses of these polymorphic sites showed that for the majority of assemblage B samples, the sites did not occur in typing positions (Table 5.6), but it was not possible to determine subpopulations from NGS reads. For the five MLG samples, nucleotide sequence alignment showed they differed from each other by 4 nucleotide positions (m2 and m8), and by 10 to 14 nucleotide positions (m7, m12, m14), and represented five separate assemblage B MLGs.

5.4.5 Phylogenetic analyses of assemblage A and B sequences

Phylogenetic analyses of all consensus sequences at the 18S rRNA locus was consistent with typing results (Figure 5.1), as determined by sequence mapping & BLASTn sequence searches. All consensus sequences were grouped with GenBank reference sequences for *G. duodenalis* assemblages A and B. For assemblage A samples typed at the *gdh*, *tpi*, and *β-giardin* loci, results of phylogenetic analyses were also consistent (Figure 5.2 – 5.4). Assemblage A samples were grouped with either subtypes A2 or A3 across the three loci (Figure 5.2 – 5.4). The *tpi* phylogeny for assemblage A samples showed the reference sequence for subtype A5 clustered with subtype A2 (Figure 5.3).

Table 5.6 *Giardia* assemblage B identified by multilocus next generation sequencing (NGS) of remote community DNA samples. The table shows genotyping data for the 18S rRNA, beta giardin (*β-giardin*), glutamate dehydrogenase *(gdh*), and triose phosphate isomerase (*tpi*) genes. All data presented are for consensus sequences (*gdh, tpi, β-giardin*), and contiguous sequences (18S rRNA) generated from NGS reads. Included in the table are BlastN sequence search results, including the GenBank accession number, percentage (%) of pairwise identity, and assemblage B genotype, for the *gdh*, and *tpi* loci. The second (2) consensus sequences obtained for mixed assemblage samples, m6 and m11, are included in the table. Degenerate bases refer to ambiguous base calls in each consensus sequence due to nucleotide variants.

		Gdh* BlastN search results *Gdh			***Tpi* BlastN search results**			**Degenerate bases at typing positions**	
Sample	18S/ *β-giardin* BlastN	Accession number of match & similarity (%)		Identity of accession numbers[a]	Accession number of match & similarity (%)		Identity of accession numbers[a]	*Gdh*	*Tpi*
m1	B	AY178756	99.0	BIII	AY228635/34	99.1	BIV-like	1	None
m2	B	AY178754	99.7	BIV-like	AY228630	100	BIII	None	None
m3	B	AY178754	99.5	BIV-like	AY228631	99.9	BIII	None	None
m4	B	AY178756/54	98.9	BIII/BIV-like	AY228631	99.4	BIII	6	3
m6 (2)	B (18S)	AY178756	99.5	BIII	AY228630	100	BIII	None	None
m7	B	AY178750	99.8	BIV	AY22834/35	99.6	BIV-like	None	None
m8	B	AY178738	100	BIV	AY22830	100	BIII	None	None
m10	B	AY178750	99.6	BIV	AY228631	100	BIII	2	None
m11 (2)	B	-	-		AY228634	97.8	BIV-like	-	None
m12	B	AY178754	99.1	BIV-like	AY228631	99.8	BIII	None	None
m13	B	AY178754	99.5	BIV-like	AY228631/630	99.3	BIII	1	3
m14	B	AY178754	99.1	BIV-like	AY228636/38	99.8	BIV	None	None
m15	B	AY178754	99.1	BIV-like	AY228634	99.4	BIV-like	None	None

[a]Identity of the sequence for each accession number, as classified by Weilinga and Thompson, 2007

Nucleotide sequence alignment confirmed assemblage A samples were subtype A2, and showed that the reference sequences for both subtypes (A2, A5) differed by only three positions in the 508 bp product that was assessed.

Phylogenetic analyses of assemblage B sequences that were concatenated across three loci (*gdh*, *tpi*, *β-giardin*) grouped all sequences with assemblage B reference sequences, and bootstrap analyses showed strong support (99%) for this grouping (Figure 5.5). Samples m2 and m8 which were identified as containing similar mulitlocus sequences (section 5.4.4), were grouped together as an internal clade within the assemblage B branch, whilst more diverse MLGs (m7, m12, m14) were placed distantly to one another within the branch. The remaining consensus sequences, which contained ambiguous base calls, were placed reference sequences for BIII and BIV (Figure 5.5). Bootstrap support of samples placed within the assemblage B group was low (18% - 66%), and most likely represented the ambiguity of these genotypes, particularly BIV-like, which was similar to both BIII and BIV at typing positions (section 5.4.4).

The topologies of all five phylogenetic trees generated in this study were similar, consisting of distinct and separate branches for each of the *Giardia* assemblages that were included for each gene (Figures 5.1 – 5.5). In all cases *Giardia* assemblages A and B formed two highly divergent groups, as depicted by the phylogenetic distance between assemblages A and B in each tree (Figures 5.1 - 5.5).

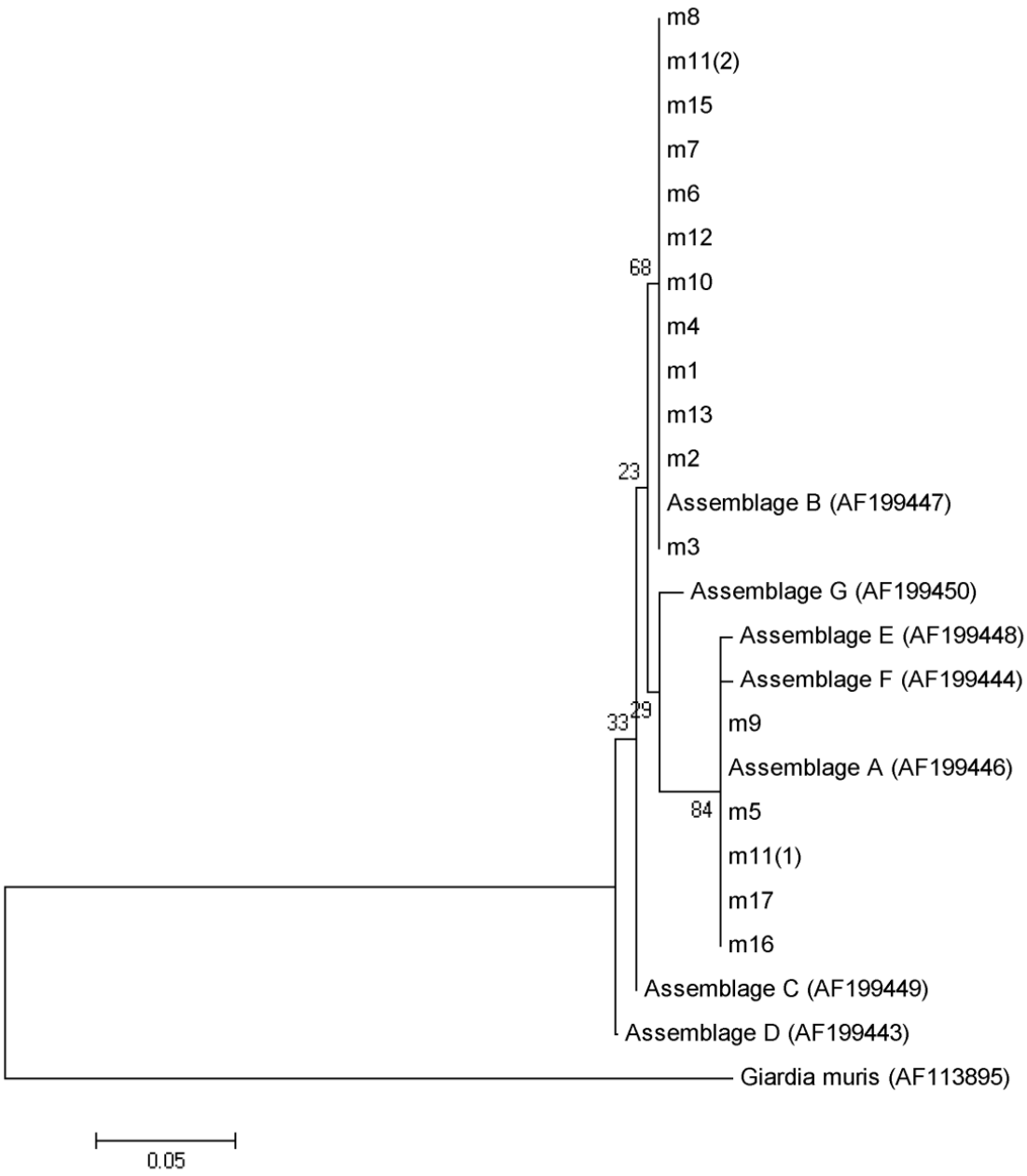

Figure 5.1 Phylogenetic tree of *Giardia duodenalis* 18S rRNA sequences generated from next generation sequencing of remote community DNA samples. The tree shows consensus sequences for 17 samples (m1 to m17). Two consensus sequences (1) and (2) were obtained from sample m11 at the 18S rRNA locus. The tree was inferred by maximum likelihood and using distance model, Tamura-Nei with gamma distribution. A total of eight representative nucleotide sequences obtained from GenBank are included in the tree and represent *G. duodenalis* assemblages A to G. Bootstrap values (%) supporting the topology of the tree are placed to the left of each node.

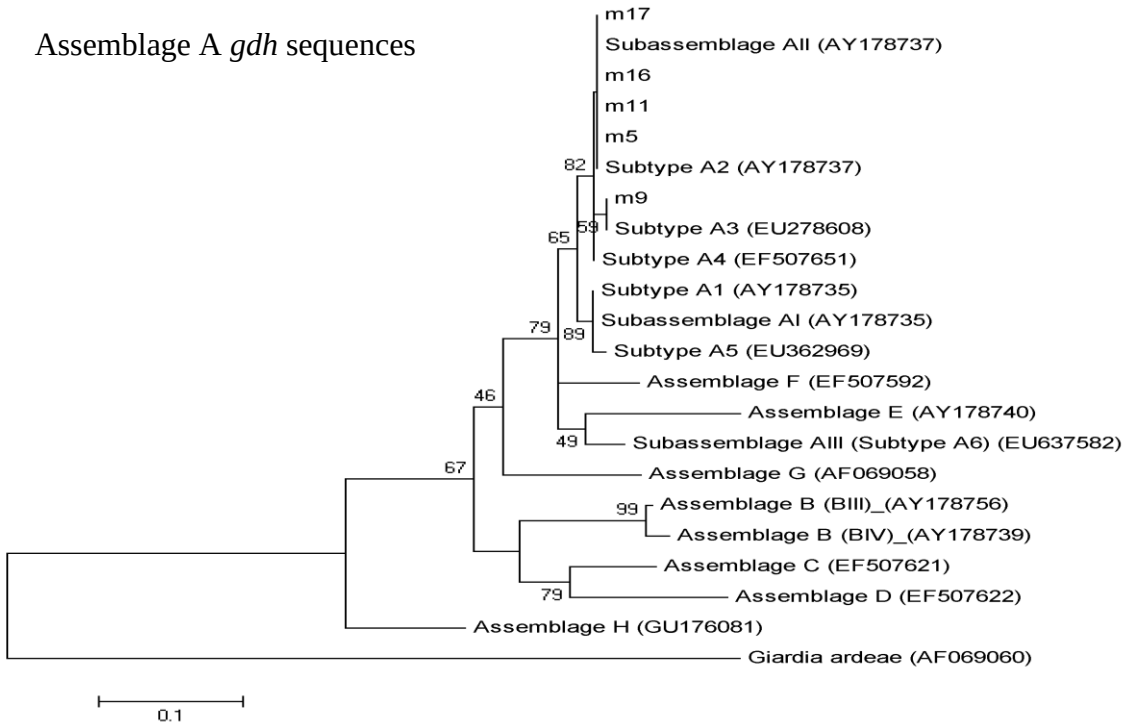

Figure 5.2 Phylogenetic tree of *Giardia duodenalis* assemblage A *gdh* sequences generated from next generation sequencing of remote community DNA.

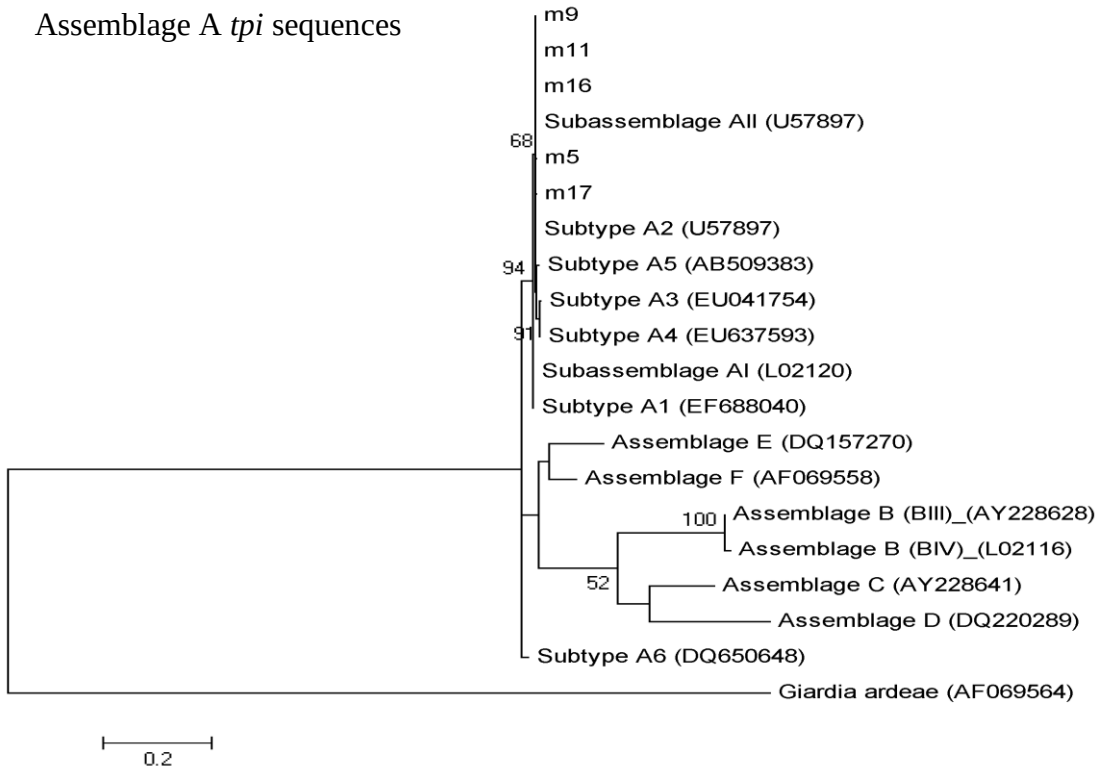

Figure 5.3 Phylogenetic tree of *Giardia duodenalis* assemblage A *tpi* sequences generated from next generation sequencing of remote community DNA.

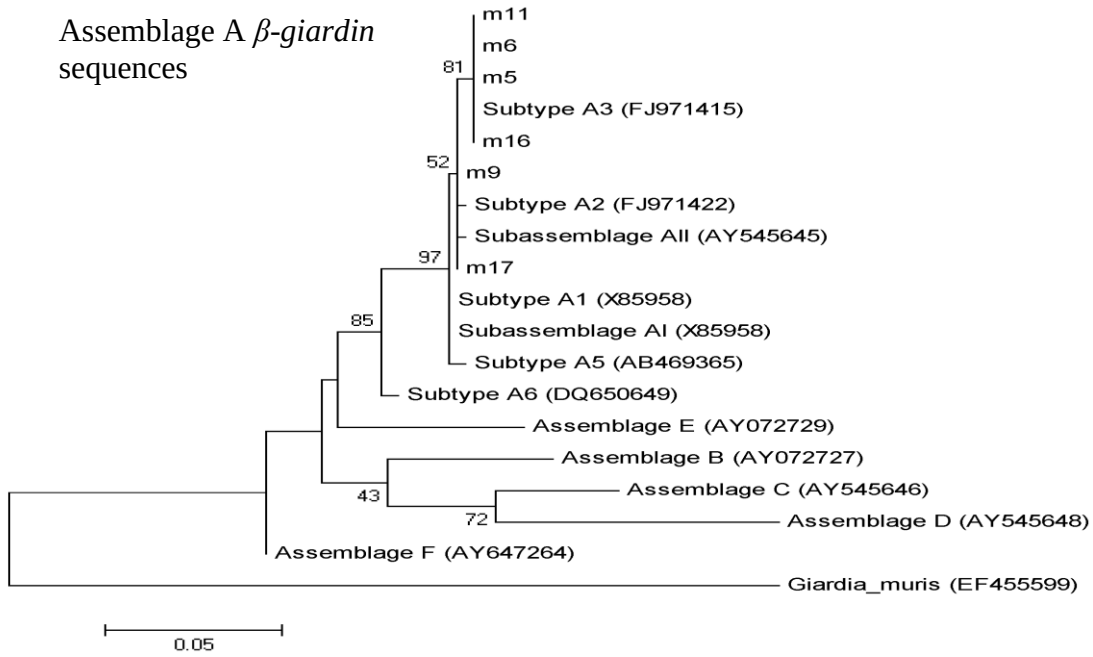

Figure 5.4 Phylogenetic tree of *Giardia duodenalis* assemblage A *β-giardin* sequences generated from next generation sequencing of remote community DNA

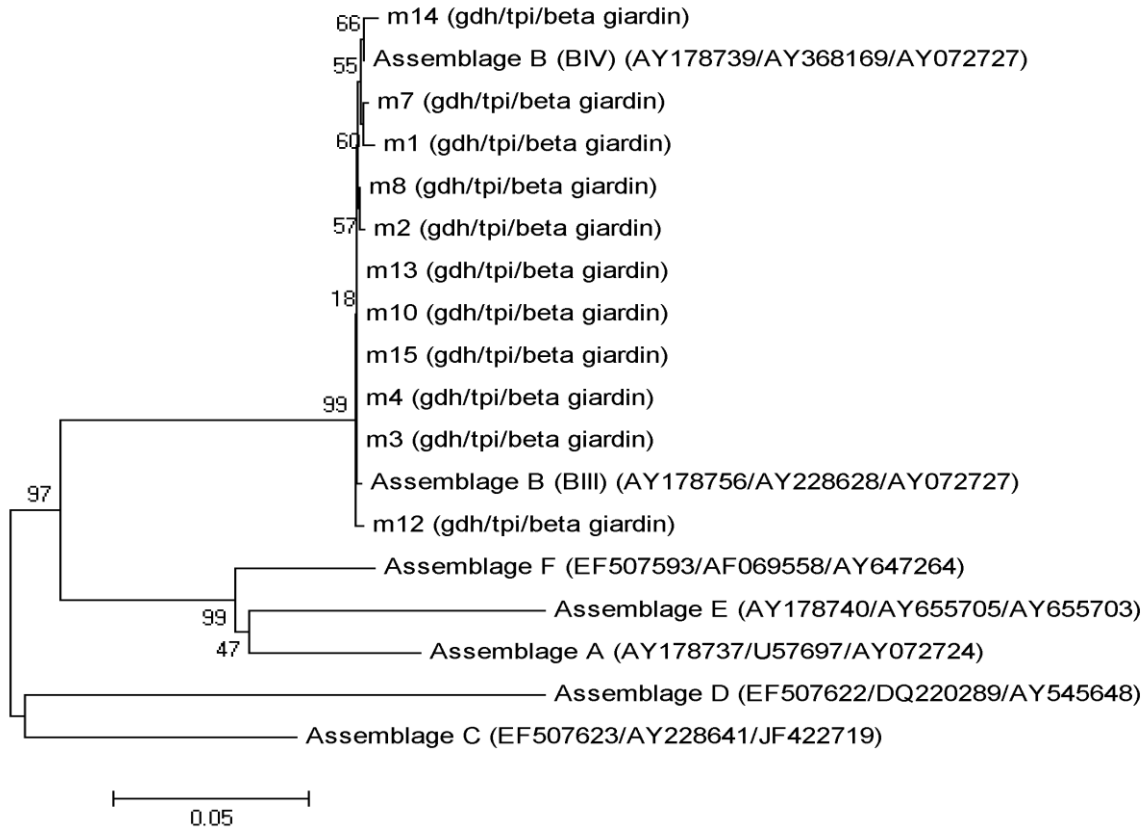

Figure 5.5 Phylogenetic tree of *Giardia duodenalis* assemblage B concatenated *gdh*, *tpi* and *β-giardin* sequences generated from next generation sequencing of remote community DNA.

5.5 Discussion

In this study we investigated NGS as a multilocus genotyping tool for *Giardia*, and to assess the genetic diversity of *Giardia* among children living in a high prevalence community. Four *Giardia* genes (18S rRNA, *gdh*, *tpi*, and *β-giardin*) which are commonly used in epidemiological research (Caccio et al., 2008; Hopkins et al., 1997; Lebbad et al., 2011; Read et al., 2002; Read et al., 2004; Sulaiman et al., 2003) were selected for typing by NGS. Application of a multilocus protocol to NGS first required optimisation of published *gdh* and *tpi* PCR protocols, and was successful for the *gdh* locus. *Tpi* PCR amplification remained difficult, and we were unable to resolve non-specific product amplification, which frequently occurred for positive DNA controls and human samples. Amplification of non-specific PCR products has been previously reported for the *tpi* locus (Breathnach et al., 2010), and inconsistent PCR amplification of isolates across several *Giardia* loci is also common among studies (Breathnach et al., 2010; Caccio et al., 2008; Nantavisai et al., 2007; Traub et al., 2005b). In the present study, typing was limited to samples that were PCR positive at all four of the selected genes. Of the 58 18S rRNA positive samples, only 17 (29.3%) were PCR positive at the *gdh*, *tpi*, and *β-giardin* loci. This result was largely due to many samples (70.7%) failing to amplify at the *tpi* locus. Although the genes selected for this study are frequently used for multilocus genotyping of *Giardia* (Durigan et al., 2014; Lebbad et al., 2011), standardised multilocus PCR protocols have not been developed, and the efficiency and reproducibility of these methods between laboratories is unknown (Lebbad et al., 2011). Amplification of all four genes, using linker PCR primer sets for NGS, required further optimisation of annealing temperatures among all protocols. This result was expected due to the longer base pair length of the linker primers (Bybee et al., 2011) but demonstrates that the application of NGS to *Giardia* typing will require standardised PCR protocols, to streamline sample preparation and increase sample numbers.

We evaluated NGS typing for *Giardia* using paired end sequencing and the Illumina MiSeq platform. Application of NGS was successful at all loci and following quality control, NGS reads were obtained for 16/17 samples at the 18S rRNA locus, and for all 17 samples at the *gdh*, *tpi*, and *β-giardin* loci. Read coverage, however, was limited by the short base pair lengths of forward and reverse reads. Of the two million reads that were recovered, only 148 paired reads met the minimum length requirements (244bp to 258bp) required to construct contiguous *gdh*, *tpi*, and *β-giardin* sequences. Recent application of this paired end MiSeq assay to the stongylid internal spacer transcribed region 2 (ITS2), has demonstrated successful recovery of reads averaging 246bp each (Lott et al., 2015). Due to the recent emergence of this technology, limitations of NGS for diverse applications are unclear (Schirmer et al., 2015). For single loci analyses on the MiSeq platform, library preparation, DNA concentration, and PCR conditions/chemistry have been flagged as potential causes of variation in the quality of reads between different sequencing runs (Schirmer et al., 2015). For multilocus typing, variation in read recovery due to variable sequence lengths, and an over representation of reads from shorter target loci has been reported from other NGS platforms (Bybee et al., 2011). The results of the present study are consistent with this, approximately 36% of all raw reads obtained before quality control were 18S rRNA sequences, whilst 16% were *β-giardin* sequences. The over representation of the shorter 18S rRNA amplicon in our sequencing data most likely resulted in reduced amplification efficiency of the sequencing assay for other loci (Bybee et al., 2011). Additionally, the use of linker PCR primer sets brought target *gdh*, *tpi*, and *β-giardin* regions within 50bp of the allowable length (2 x 300bp) for paired end MiSeq sequencing (www.illumina.com). Shorter amplicon lengths may have been more suitable for multilocus sequencing with the current platform.

All 17 samples included in this study were successfully genotyped to the *G. duodenalis* assemblage, subassemblage, and subtype level, using combined NGS sequence

data from the18S rRNA, *gdh*, *tpi*, and *β-giardin* loci. Our results showed three different assemblage A subtypes and five different assemblage B multilocus genotypes contributed to cases among children in this remote Indigenous community. High nucleotide sequence diversity was identified among all assemblage B cases, which was absent in all assemblage A cases. Subtypes within assemblage B have yet not been defined (Sprong et al., 2009), and multilocus genotyping across all three variable genes (*gdh*, *tpi*, and *β-giardin*) is required to determine the diversity of assemblage B cases within a population (*c.f.* Feng and Xiao, 2011).

Among the total samples screened, assemblage A cases were detected in a total of 6/17 (35.3%) samples, including two mixed assemblage A and B samples. Assemblage B was detected in a total of 13/17 (76.5%) samples, including mixed samples. The three subtypes detected from assemblage A samples (AII-1, AII-3 AII-3) have been previously described (Caccio et al., 2008; Lebbad et al., 2011), but very few human studies have been performed using multilocus genotyping using three *Giardia* genes, and the diversity of these subtypes among human populations is unknown. Subtype AII-1 has been identified from axenic strains of human isolates (Caccio et al., 2008) and Swedish children (Lebbad et al., 2011), whilst subtype AII-3 has been detected among human cases in Italy (Caccio et al., 2008). Subtype AII-2 has also been reported from Swedish children (Lebbad et al., 2011). All assemblage A subtypes belonged to subassemblage AII, which has been associated with anthroponotic transmission of human giardiasis, due to its limited host distribution (Yang et al., 2010).

The five assemblage B MLGs identified in this study were isolated from five different samples (m2, m7, m8, m12, m14), and each MLG was different. Phylogenetic analyses showed MLGs for two samples (m2, m8) were closely related, suggesting a potential transmission path between these two cases. Assemblage swapping, where isolates are assigned to different assemblages (A or B) (Sprong et al., 2009) was detected

for one of two mixed A and B samples in this study. This also occurred within assemblage B, when samples were typed individual loci of the *gdh* and *tpi*; 83% of assemblage B samples could not be assigned to a genotype (BIII, BIV, or BIV-like), due to inconsistent typing results. Similar inconsistencies and assemblage swapping have been reported from other studies, among assemblage A and B isolates, and is frequently documented for isolates within assemblage B (Amar et al., 2002; Berrili et al., 2012; *c.f.* Caccio and Sprong, 2010; Lalle et al., 2009; Yang et al., 2010). *Giardia* assemblages A and B are most likely separate species of *Giardia* and genetic interactions, including meiotic recombination between assemblages is unclear (Caccio and Sprong, 2010). Similarly, heterogeneity and recombination among assemblage B isolates remains unresolved (Berrili et al., 2012; Lalle et al., 2009). Intra-assemblage and inter-assemblage recombination, and mixed template infections are likely to be more common in high prevalence communities, where increased frequency of transmission, host contact, and competitive parasite interactions are more frequent (Thompson and Monis, 2012). For a second mixed sample identified in this study, detection was enhanced by NGS, by the ability to separate populations of assemblage A and B reads, which would not be possible with direct Sanger sequencing of *Giardia* DNA (Asher et al., 2012b; Voelkerding et al., 2010). This enabled detection of mixed templates, and distinguished the result from sequence heterogeneity (Caccio and Sprong, 2010). For all other mixed cases in this study, we were unable to evaluate the presence of subpopulations further, due to short read coverage and limitations in constructing small assemblies. Our results, however, demonstrate that carriage of mixed assemblage A and B cases occur in this community, which were not detected during our previous study (Asher et al., 2014), in which the same samples were screened by 18S rRNA Sanger sequencing, and terminal-restriction fragment length polymorphism (T-RFLP) of a smaller region of the *gdh* gene. With improved NGS protocols, providing increased read quality and length, NGS typing would be a highly sensitive tool for sub-population analyses.

The three assemblage B genotypes (BIII, BIV, BIV-like) that were detected among assemblage B samples, at either the *gdh* or *tpi* loci, have been previously reported from human cases in Western Australia and New South Wales (Asher et al., 2012b; Yang et al., 2010). Assemblage B isolates have also been detected among domestic animals and wildlife in New Zealand and Australia (Learmonth et al., 2003; Nolan et al., 2010); and anthroponotic and/or zoonotic transmission cycles for assemblage B genotypes are unknown (*c.f.* Feng and Xiao, 2011). Genotype 'BIV-like' like has been identified as a variant of BIV (Weilinga and Thompson, 2007), and analyses of sequences generated in this study showed BIV-like sequences consisted of BIII and BIV specific nucleotides at typing positions for both loci.

Assemblage B samples were genetically diverse, MLGs for 6/11 samples with *gdh*, *tpi*, and *β-giardin* typing results could not been determined due to polymorphic nucleotide positions and ambiguous base calls among the consensus sequences. These results correspond to the findings of Lebbad et al. (2011) in which 52% to 64% of paired end multilocus Sanger sequences for *Giardia* isolates could not be resolved due to variant sites. Lower frequency variants, which were not included in consensus calling, were also common among assemblage B samples. In this study, low frequency variants were determined from a minimum variant frequency of 25%, and were called from assemblies containing on average 2000 to 4000 reads. Genomic analyses of *Giardia* using next generation sequencing has demonstrated polymorphic nucleotide detection at sites with 10X coverage, and with minimum variant frequencies of 15% (Adam et al., 2013). In the present, we chose a highly conservative approach to variant detection, given the short base pair lengths of NGS reads, and the high number of reads required in to construct consensus sequences from each assembly. Due to this, subpopulation analyses could not be performed. The large number of low frequency polymorphic sites detected among assemblage B samples may be attributed to the high sensitivity of detection by NGS

(Voelkerding et al., 2010), which has not been applied to *Giardia* typing in an epidemiological setting.

Comparison of assemblage assignment by 18S rRNA NGS to 18S rRNA Sanger sequencing of the same samples (Asher et al., 2014) showed results were consistent between the two methods, with the exception of one sample, where low frequency assemblage B reads were detected among assemblage A reads at the 18S rRNA by NGS. Although *gdh* screening performed in this study differed from previous screenings of these samples by *gdh* T-RFLP, results were similar to previous findings, with many mixed assemblage B cases.

To our knowledge this is the first application of NGS to multilocus typing of *Giardia* in an epidemiological context, and from a high prevalence community. Typing of *Giardia* by NGS to the assemblage level, and within assemblage A, to the subassemblage and subtype level was successful. In this application of NGS, our results were limited by the low sensitivity of the multilocus PCR approach used in this study, and by the low read coverage obtained for PCR positive samples. The diversity of assemblage B MLGs may be underestimated due to the small number of samples which could be assessed. Assemblage B cases caused by related *Giardia* isolates were identified by multilocus NGS, and suggests potential application of this method to disease source tracking. Detection of mixed assemblage A and B templates, and the ability to distinguish A and B populations from individual samples, demonstrates the high sensitivity of detection by NGS and suitability for investigating *Giardia* populations contributing to individual infections. Our results have highlighted improved genotyping accuracy with screening of multiple *Giardia* loci, however future application of this method will require the development of robust multilocus PCR screening protocols that couple high sensitivity screening with informative target loci and PCR amplicon lengths suitable for NGS platforms. Our results provide the first description of several *Giardia* assemblage A subtypes, and assemblage B multilocus

genotypes contributing to childhood cases in a remote Indigenous community of Australia. Further research in this community is essential, to determine the extent of this diversity and prevalence of subtypes, and to examine disease links between subtypes and high *Giardia* carriage rates among children of this community.

5.6 Acknowledgements

This research was conducted with financial assistance from Macquarie University, and received funding from the Macquarie University Post Graduate Research Fund. We thank community members and staff involved in the primary study from this remote Indigenous community. We thank Professor John Ellis and Dr Sonja Frolich of the University of Technology, Sydney, for kindly providing trophozoite cultures used in this research.

6. Discussion and conclusions

The objectives of this thesis included two broad aims 1) to examine the molecular epidemiology of *Giardia* contributing giardiasis in Australia, specifically that contribute to sporadic cases in NSW, and cases in children from a high-prevalence, remote Indigenous community; and 2) to develop, evaluate, and apply emerging molecular tools to *Giardia* identification. Prior to this study, few investigations examining the molecular epidemiology of giardiasis in Australia had been undertaken. Two studies identifying *Giardia* assemblages among children in day care centres (Read et al., 2002) and among human cases in rural and urban regions of Western Australia (Yang et al., 2010) have been published. The genetic assemblages/subtypes of *Giardia* that contribute to sporadic human disease in NSW were unknown. Only 12 *Giardia* isolates had been characterised from a remote Indigenous community of Western Australia (Hopkins et al., 1997; Hopkins et al., 1999), but these studies predated the current understanding of *Giardia*, generated since by numerous molecular investigations (Monis et al., 1999; Monis et al., 2003; Read et al., 2004; Caccio et al., 2008; Sprong et al., 2009). Molecular detection of *Giardia* has previously been performed using RFLP and DNA sequencing (Ajjampur et al., 2009; Read et al., 2004; *c.f.* Weilinga and Thompson, 2007), and new technologies developed for high throughput detection and population analyses had not been explored.

In NSW, *Giardia* assemblages A and B were detected among sporadic cases and the prevalence of assemblage B was higher in NSW than has been reported in other low prevalence Australian communities (Yang et al., 2010). *Giardia* assemblages and subtypes identified in sporadic cases from NSW were highly similar, and patients aged 0 - 39 years were most at risk from assemblage B. Subtyping analyses enabled examination of transmission cycles occurring in NSW, and dispersal patterns across NSW. Among children from the remote Indigenous community, we identified high carriage rates of

Giardia, and found a diversity of *Giardia* subassemblages/genotypes occurring among the children over a 12 month period. Application of T-RFLP and multilocus NGS enabled population analyses of *Giardia* contributing to individual and community cases, and demonstrated potiential for future development and application of these tools for population based screenings of *Giardia*.

The NSW investigations significantly advance knowledge of *Giardia* transmission in this Australian State; through the identification of human infective *Giardia* assemblages among NSW residents. Human infective assemblages have been previously documented among domestic and wild animals in NSW (Delport et al., 2014; Ng et al., 2011; Vermeulen et al., 2015). The results of this thesis provide an essential disease link between the occurrence of these assemblages among domestic and wild animals in NSW and NSW residents. A high endemicity of assemblage B among NSW humans was identified, and evidence of seasonal influence of the geographic dispersal of assemblage A occurring among a human population was demonstrated for the first time (Chapter 2). The remote community studies provide the first comprehensive evaluation of *Giardia* assemblages, subassemblages, and subtypes contributing to cases among children in a remote Indigenous community of Australia (Chapter 4, Chapter 5). *Giardia* infections among children of low and high prevalence communities of Australia differed, and in comparison to NSW, children in high prevalence remote Indigenous communities are at an increased risk of assemblage A, and mixed assemblage B infections by comparison (Chapter 2, Chapter 4, Chapter 5). Two new molecular typing tools were developed and evaluated for *Giardia* identification, terminal-restriction fragment length polymorphism (T-RFLP) and multilocus next generation sequencing (NGS). The study demonstrates the first application of these tools in epidemiological research (Chapter 3, Chapter 4, Chapter 5). The molecular techniques developed here provide high resolution, simultaneous detection of *Giardia* populations from individual human cases and from population cases.

Furthermore, application of these tools in this thesis, demonstrated that low abundance, genetically diverse *Giardia* populations contribute human giardiasis; and that high resolution, population based screening tools are required, to both detect these cases and to evaluate the diversity of *Giardia* that occurs in epidemiological foci.

6.1 Evaluation of sporadic giardiasis in NSW and identification of *Giardia* assemblages, subassemblages and subtypes among NSW patients

Two separate investigations of sporadic human giardiasis in NSW, occurring between 2010 and 2013, were performed with *Giardia* assemblages identified from 237 positive human cases (Chapter 2, 3). The combined results of these studies provide one of the largest investigations into sporadic, clinically diagnosed, human giardiasis occurring within a single population (Table 1.3a-c; *c.f.* Feng and Xiao, 2011). During the study period, human giardiasis in NSW was caused by two different *G. duodenalis* assemblages, A and B, which are known to cause human infection, and are globally distributed (Caccio and Ryan, 2008).

A significant finding of the NSW investigations was the high proportion of human assemblage B (85.7%), dominating sporadic transmission in NSW during the study period. This rate significantly exceeds reports of human assemblage B from Western Australia, where a comparable study has been performed. The prevalence of assemblage B in NSW also exceeds rates reported elsewhere in the world. Among clinical studies with larger sample screenings (> 50 samples), high rates of assemblage B have been reported from the United Kingdom , Belguim, Nicaragua, and Sweden, (Breathnach et al., 2010; Geurden et al., 2009; Lebbad et al., 2008; Lebbad et al., 2011). Among these countries, assemblage B prevalence ranges from 61.8% to 75% in the United Kingdom, Belgium and Sweden, whilst prevalence of 79% is reported from Nicaragua (Breathnach et al., 2010; Geurden et al., 2009; Lebbad et al., 2008; Lebbad et al 2011). Globally, the distributions of *Giardia*

assemblages A and B varies among human populations, and is most likely linked to demographic, socio-economic, and climatic conditions (Savioli et al., 2006). Disease factors contributing to the dispersal of assemblages A and B remain unclear (*c.f.* Feng and Xiao, 2011) and larger population studies are required to enable global comparisons.

Another significant finding of the NSW analyses (Chapter 2) was the large geographic dispersal of assemblages A and B among human populations in NSW and the seasonally-linked geographical shifts in the occurrence of assemblage A. Mapping of patient cases (n = 243) throughout NSW demonstrated geographical clusters of human giardiasis consistently occurring in the Newcastle and lower Hunter region throughout the study, and identified this region as a disease hotspot in NSW. Mapping of assemblage A and B cases (n = 165) showed both assemblages were widely dispersed throughout rural/regional NSW, but seasonal variation in the dispersal of assemblage A was evident. During summer, assemblage A cases were more frequent in North-Western, and Western NSW regions, including the towns of Tamworth and Dubbo. During Autumn, geographic dispersal of assemblage A occurred in the South-Western communities of Dubbo, Orange, and Wagga Wagga. In disease hotspot regions of Newcastle and the lower Hunter, no assemblage A cases were detected during Autumn, however, during Winter and Spring, all assemblage A cases were confined to the Hunter New England health district, which includes Newcastle and the lower Hunter.

Spatial mapping has been previously performed to investigate prevalence of giardiasis in Canada (Odoi et al., 2003; Odoi et al., 2004) but this is the first time this approach has been applied in the Australian context. Overall, this study provides strong evidence of anthroponotic transmission occurring among AII cases in NSW, particularly among patients aged 40 years and over, and suggests the risk of acquiring AII type infections in NSW varies seasonally and geographically. Subassemblage AII has a restricted host range and has been most commonly identified from human populations

(Yang et al., 2010). Previous NSW studies have identified assemblage A among non-human hosts (cattle, deer, kangaroos, captive and wild marsupials, wild dogs, foxes, dingo, horse) but subtyping of these cases has only identified subassemblage AI (Ng et al., 2011; Vermeulen et al., 2015), which further supports anthroponotic transmission of AII human cases in NSW. Patient data for the single human AI case identified in this study was not available, and further analysis of this case was not possible. The infrequency in which subassemblage AI was detected among humans in NSW indicates very limited transmission of AI between human and animal hosts. Furthermore, the results of this thesis provide the first evidence that human populations in NSW are unlikely sources of AI infections among non-human hosts in NSW.

Assemblage B cases were identified from 5.1% of NSW cattle DNA samples (*n* = 175) that were screened for this study (Chapter 2), and these cases occurred in areas proximal to human populations. In previous studies, assemblage B has been identified from cattle, deer, kangaroos, captive and wild marsupials, and from captive marine mammals in NSW (Delport et al., 2014; Ng et al., 2011; Vermeulen et al., 2015). Identification of assemblage B among the NSW human population supports transmission of this assemblage among human and non-human hosts. The prevalence of AII and high rate of assemblage B among humans, and occurrence of AI and B that has been documented among non-human hosts (Delport et al., 2014; Ng et al., 2011; Vermeulen et al., 2015) supports separate transmission cycles occurring in NSW between humans (AII & B), between humans, domestic, captive, and wild animals (B), and between domestic, captive, and wild animals exclusively (AI) (Thompson, 2013; Vermeulen et al., 2015). The zoonotic potential of subassemblage AI and limited cases among NSW humans suggests several potential disease mechanisms that limit establishment of AI among the human population, including; limited host contact and directional transmission of assemblage B from humans to non-human hosts; and increased host specificity and

adaptation of assemblage B among the human population, which prevents establishment of AI infections due to competitive parasite interactions (Hopkins et al., 1997; Thompson and Monis, 2012). The high endemicity of assemblage B detected among sporadic human cases in NSW would support both disease mechanisms.

Risk factors associated with anthroponotic transmission, and human subassemblage AII or assemblage B infections in NSW are many. Contact between infants at day care centres, poor hygiene practices among infants at day cares, and secondary transmission from infants to familial contacts, carers, and staff are a primary source of anthroponotic giardiasis in NSW (Hoque et al., 2002; Laupland et al., 2005; United States Environmental Protection Agency (U.S. EPA), 1999). The high prevalence of assemblage B among children and young adults in NSW may be indicative of transmission occurring among these sources. Seasonal increases in the overall prevalence of sporadic giardiasis in NSW are documented (NSW Health Notifiable Conditions Information Management System (NCIMS), 2015) and are most likely linked to increased water activities and contact among children enrolling in day care/school during the Summer and Autumn months each year (Hoque et al, 2002; Odoi et al., 2003). Small familial outbreaks of human giardiasis have been linked to grey water usage in rural/regional NSW (NSW Department of Health, 2006). Additionally, human excrement, and contamination of public swimming pools and drinking water sources is a major concern for NSW public health authorities (NSW Department of Health, 1998; Stein, 2000). Faecal run off from untreated human sewage, and form non human hosts inhabiting water catchment areas in NSW is a primary source of contamination in the NSW water supply (Ng et al., 2011; NSW Department of Health, 1998). During the study period, heavy rainfall events were reported in NSW during the months of Autumn (Australian Government Bureau of Meteorology 2011 and 2012). These events are not uncommon during the early weeks of March each year (Australian Government Bureau of Meteorology, 2011 and 2012) and may contribute to increased

Giardia transmission during Autumn, through increased faecal-run off and potential for water contamination. It is estimated that up to 75% of all giardiasis infections occurring within a population go undetected, due to high rates of asymptomatic cases (Hoque et al., 2002; U.S. EPA, 1999). In this study, samples were acquired from pathology companies, and it is reasonable to assume that a pathology request for these patients was triggered by clinical manifestations associated with giardiasis. The high rate of human assemblage B identified in NSW, and lower prevalence of AII cases, may represent an association to clinical features of giardiasis, which are not yet known, or are specific to this population. Assemblage linked differences to manifestations of human giardiasis have been reported from different countries (Homan and Mank, 2001; Ignatius et al., 2012; Lebbad et al., 2011; Read et al., 2002), however, these studies have investigated different target populations, and overall disease links are unclear.

The prevalence of assemblages A and B among human NSW cases was confirmed by two separate investigations. For both studies, 86% (n = 165), and 85% (n = 72) of samples typed by RFLP (Chapter 2), T-RFLP and DNA sequencing (Chapter 3) were assemblage B. The acquisition of patient data during the first study enabled duplicate patient samples to be removed from analyses, and these results represent separate human cases in NSW. An important limitation of the NSW investigations, however, is evident, relating to typing of isolates at the *gdh* locus. Assemblage swapping, which occurs when a single isolate is typed as different assemblages or subassemblages across two or more loci, have been reported from studies of human and non-human hosts (Bonhomme et al., 2011; Caccio and Sprong, 2010; Delport et al., 2014). A recent survey of publically available *Giardia* sequence data generated from human European studies, demonstrated that 12% of human cases are typed as mixed assemblages (A and B) when screened at two loci (Sprong et al., 2009). In Australia, mixed assemblage A and B cases have been reported from one human in a remote Indigenous community (Hopkins et al., 1997; Hopkins et al., 1999) and

from five (4%) human cases in Western Australia (Yang et al., 2010). Multilocus genotyping results presented in this thesis showed that among 74 participants from the Northern Territory studies (Chapter 4, Chapter 5), 2.7% were detected as mixed A and B cases. The underlying mechanisms that result in mixed genotyping results, including the presence of mixed *Giardia* templates, or heterogeneous *Giardia* isolates, are still unresolved (Caccio and Sprong, 2010; Sprong et al., 2009). The low prevalence of these cases from Australian communities may be associated with the overall high rates of assemblage B (75% to 86.5%), reported in this thesis and from a previous study (Yang et al., 2010), and fewer human assemblage A cases (13.5% to 25%), which differ from the frequencies of A (43%) and B (56%) cases reported from Europe (Sprong et al., 2009). The results of the NSW investigations would support this, mixed A and B NSW human cases occurred among patients aged 40 years and over, and among this group, the frequency of assemblage A infections were comparatively higher than patients less than 40 years of age. Detection of 72 cases by fluorescence, using T-RFLP, demonstrated high resolution detection of mixed cases at the *gdh* locus (section 6.3). Additionally, 17/72 (23.6%) BIV NSW human cases screened by T-RFLP contained secondary unique fluorescent peaks that were not specific to assemblage B genotypes, and indicated low frequency variant assemblage B populations among these cases (Chapter 3). The potential for assemblage swapping, however, and thus potentially underestimating the proportion of mixed assemblage cases occurring in NSW must be considered in the interpretation of these results.

Difficulties in amplifying *Giardia* DNA from positive human cases were experienced during this study. Additional samples were screened for the NSW investigations but could not be included in the analyses, due to PCR results which could not be confirmed as positive or negative. Difficulties in amplifying *Giardia* DNA are reported by numerous investigators (Breathnach et al., 2010; Caccio et al., 2008; Delport et

al., 2014; Nantavisai et al., 2007) and these difficulties are emerging as a primary limiting factor in many *Giardia* studies. These results demonstrate that improved and standardised PCR protocols will be essential to future *Giardia* research, to reduce costly and labour intensive PCR screenings, which limit the feasibility of sample screenings in current research.

Knowledge of *Giardia* subtypes, and potential for transmission among hosts, to disease management and control is critical and can only be obtained via molecular testing (Koehler et al., 2014). The importance of molecular data to disease management is evident through the Sydney Water Crisis of 1998 (Stein, 2000). At the time, routine molecular typing tools for identification of *Giardia* and *Cryptosporidium* had not been developed, and the source of the contamination, and potential for human infection were unknown (NSW Department of Health, 1998; Stein, 2000). The Sydney Water crisis lasted nine weeks, resulted in Sydney-wide boil water alerts, including to schools, and hospitals, and was estimated to directly cost Sydney Water $33 million at the time (Stein et al., 1998). Although contamination sources were not determined, the crisis was linked to a prolonged drought, followed by heavy rainfall events that occurred just prior to the contamination (Stein, 2000).

Data on molecular epidemiology generated here for NSW provides a basis in which to progress knowledge of *Giardia* transmission in NSW. These data indicate that two separate species of *Giardia* (assemblages A and B) were contributing to NSW infections during the study period. Assemblages A and B are considered to be separate species of *Giardia*, due to the large genetic and biotypic differences between these assemblages, which are documented by numerous studies (Adam et al., 2013; Jerlstrom-Huitqvist et al., 2010; Monis et al., 2003; Sprong et al., 2009). The findings reported here provide evidence of disease links between *Giardia* occurring in human populations and non-human hosts. These data also contribute to the global understanding of giardiasis transmission, by

providing potential avenues to investigate spatial, seasonal, and demographic differences among assemblage A and B human infections, which have so far been elusive. Knowledge of transmission cycles and demographic distributions of the human infective assemblages among global and local populations will be increasingly important, to understand giardiasis transmission, with increasing population sizes and population densities (Waldron et al., 2011a). Knowledge of disease hotspot regions, and spatio-temporal distributions of assemblages will be increasingly important, to understand potential for environmental transmission of *Giardia*, as the frequency of extreme weather events, and seasonal climate variations due to climate change become more apparent (McMichael et al., 2006).

6.2 Prevalence of *Giardia* and genetic subtypes contributing to high prevalence cases among children of a remote Indigenous community of Australia

Globally, diarrhoeal disease is a leading cause of death among children under five years, second only to deaths caused by respiratory infections (Lanata et al., 2013; World Health Organisation, 2013). In 2010, an estimated 700 000 of 1.7 billion cases of diarrhoea led to death in children under the age of five (Walker et al., 2013). Childhood deaths caused by diarrheal disease are linked to underlying medical conditions, including malnutrition, which is a significant risk factor for severe diarrhoea (Grimwood and Forbes, 2009; World Health Organisation, 2013). Poor nutrition and chronic enteric infections contribute to malnutrition among children living in developing regions and socially disadvantaged communities (Holt et al., 2010; Grimwood and Forbes, 2009). Poor sanitary and hygiene infrastructure, and overcrowded living conditions contribute to high frequency transmission of infectious diseases (McDonald et al., 2008).

Globally, an estimated 280 million cases of symptomatic human giardiasis occurs each year (*c.f.* Caccio and Ryan) but overall prevalence of giardiasis may be as high as one billion in developing regions (Wright et al., 2003). Children living in high prevalence

regions are most at risk from the chronic manifestations of giardiasis, which leads to malnutrition, and impairments to growth and cognitive development (Thompson, 2008). Approximately 3.1% of child deaths due to diarrhoea are attributed to giardiasis (Lanata et al., 2013), but the contribution of giardiasis to enhanced clinical manifestations of acute and severe diarrhoea is likely to be much higher.

In Australia, Indigenous health is a primary concern to health authorities and governments, due to the large gap in overall life expectancy, which is 11.5 years below the non-Indigenous population, and infant mortality rates which are two times higher than non-Indigenous children (Australian Institute of Health and Welfare, 2011; CDC Northern Territory, 2008). Hospital admissions due to acute gastroenteritis are up to eight times higher for Indigenous children, than they are for non-Indigenous children of the same age (Gracey and Cullinana, 2003; Grimwood and Forbes, 2009). Malnutrition, and failure to thrive are common among children living in remote Indigenous communities (Australian Institute of Health and Welfare, 2011; Grimwood and Forbes, 2009; Holt et al., 2010; Mc Donald and Bailie, 2010) and pose a significant health threat to these children. Indigenous populations living in remote communities of the Northern Territory are among the highest risk groups for communicable gastrointestinal infections in Australia (Grimwood and Forbes, 2009).

In this thesis, the prevalence of *Giardia*, and the genetic diversity of *Giardia* contributing to cases among children in the community were examined in order to determine the frequency and prevalence of genetic subtypes in this community, and to allow comparison of these findings to children from low prevalence Australian communities (Chapter 4, Chapter 5).

A significant finding of these investigations was the high rate of *Giardia* detection (64.9%) among children of this community and the large proportion (58.6%) of cases that were not detectable by microscopy. These findings provide an important insight into the

high, and ongoing prevalence of *Giardia* documented from Australia's remote Indigenous communities, particularly among children (Jones, 1980; Meloni et al., 1993; Reynoldson et al., 1998; Thompson et al., 2001). Intermittent parasite shedding, low parasite shedding, and the high sensitivity of screening by PCR are expected to produce variations in the positivity of children screened by these methods (Caccio and Ryan, 2008; Ghosh et al., 2000; Ignatius et al., 2012). Although microscopy has low resolution, it is required to determine the presence of an established infection from the host (Traub et al., 2005a). The high rate of PCR positive detection reported here suggests high carriage rates among children of this community, however, the presence of an established *Giardia* infection among these children is unknown. These data contribute to growing evidence that *Giardia* and other enteric pathogens are difficult to detect among children of high prevalence communities (Ajjampur et al., 2008; Cook et al., 2009; Ignatius et al., 2012). Recent studies have documented increased *Giardia* detection rates by PCR screenings for children in high prevalence regions of Rwanda (Ignatius et al., 2012), and failure of treatment programs to reduce *Giardia* prevalence among Guatemalan children based on microscopy screenings (Cook et al., 2009). Increased detection rates (35%) of enteric pathogens among hospitalised Indian children with diarrhoea have also been demonstrated by combining PCR screenings with conventional diagnostic techniques, including microscopy, culture, and immunoassays (Ajjampur et al., 2008). Previous initiatives in remote Indigenous communities of Australia have included improved housing and sanitation facilities, and community wide drug treatment programs, but the prevalence of infectious diseases remains high in many communities (Grimwood and Forbes, 2009; Holt et al., 2010; McDonald and Bailie, 2010). The results of this thesis indicate that children may go undetected, enabling continued transmission of *Giardia* which contributes to high prevalence rates, and limits the value of community health programs.

The highest prevalence of *Giardia* was found among infants and children, aged 0 - <5 years, however, the combined results from both screening approaches, and including microscopy (Table 4.1), demonstrated no significant difference in detection among children aged 0 -< 5 years, or 0 -< 10 years. Overall, high rates of positivity (60% - 66%) were identified from both groups. In NSW, approximately 65% of positive children were < 5years, and only 35% were 5 - <10 years, which concords with prevalence rates reported by NSW Health (NSW Health, NCIMS, 2015). Among children of high prevalence communities, carriage of *Giardia* is thought to occur among older children with more developed immunity (Lanata et al., 2013). From the PCR results alone, the study demonstrated carriage rates were significantly higher among younger children, and positivity increased among these children during the study period.

Children under five are at the highest risk of diarrhoea related illness and death (World Health Organisation, 2013). Similar to global studies, *Giardia* has only been directly linked to a small number of diarrhoea related hospital admissions among Indigenous children in the Northern Territory (Kukuruzovic et al., 2002). These data would suggest *Giardia* carriage and/or discrete infections among young children may contribute to the increased risk of acute and severe diarrhoea among this group. Chronic giardiasis has been linked to epithelial dysfunction among patients (Troeger et al., 2007) and to increased intestinal permeability among Nepali children (Goto et al., 2002). Furthermore, global and local estimates of *Giardia*'s contribution to diarrhoeal disease have been based on stool microscopy alone, or microscopy and immuno-assays (Lanata et al., 2013; Kukuruzovic et al., 2002), which may not have detected *Giardia* as a concurrent infection, contributing to the overall disease manifestations. Up to 45% of asymptomatic children in remote Indigenous communities have concurrent infections (Grimwood and Forbes, 2009). In this study repeated PCR positivity was demonstrated for eight children in the community over a 12 month period, providing evidence for either a long period of

carriage, or re-colonisation by *Giardia* among these children. Repeated and prolonged exposure to *Giardia* is associated with chronic disease outcomes among remote Indigenous children (Holt et al., 2010). In high prevalence communities, the burden of these chronic infections, including stunted growth and development, are linked to poor educational and health outcomes throughout life (Holt et al., 2010; McDonald et al., 2010).

Detection of *Giardia* by 18S rRNA PCR and *gdh* T-RFLP identified a diversity of *Giardia* assemblages and subtypes contributing to cases in this community (Chapter 4). The last studies to examine *Giardia* assemblages among Australia's remote Indigenous populations were performed 16 to 18 years prior to this thesis, and contribution of *Giardia* assemblages/subassemblages among children were unknown (Hopkins et al., 1997; Hopkins et al., 1999). It has been unclear, as to whether the high prevalence of *Giardia* reported from remote Indigenous children is caused by diverse assemblages, and whether re-infection with the same/different assemblages occurs (Thompson and Monis, 2012). The results of this thesis demonstrated wide dispersal of *Giardia* assemblages A and B, and within these, subassemblage AII, and genotypes BIII and BIV among all children of the community. The study demonstrated children were most at risk from assemblage B (75%), and that the frequency of assemblages A (25%) and assemblages B (75%) were maintained over time. Furthermore, the study showed that *Giardia* subassemblages/genotypes (AII, BIII, BIV) persisted in the community over a 12 month period. Among five children who were repeatedly positive, only assemblage B was detected among the samples, which were collected 12 months apart. For three of these cases, the same assemblage B subtypes (BIII, mixed BIII/BIV) were identified at the two collection intervals. Overall, these results demonstrate that a diversity of human infective assemblages/genotypes contribute to community cases, and persist in the community over time, suggesting potential for re-infection with different types of *Giardia* in this community.

In comparison to children of the same age in NSW, remote community children were at a higher risk of assemblage A (NSW, 5.3%, remote community, 25%). However, the frequencies of assemblages A and B detected in this remote community are comparable to rates documented among children in low prevalence communities of Western Australia (Read et al., 2002; Yang et al., 2010). The proportion (28%) of mixed assemblage B populations among remote community children was high, when compared to NSW and Western Australia. Mixed assemblage B populations have not been reported among children from Western Australia (Yang et al., 2010) and were not detected amongst NSW children. The overall rate of mixed assemblage B cases in NSW was 4.9%. Detection of subassemblage AII (25%) in this community suggests anthroponotic transmission occurring among these cases (Yang et al., 2010). Sources of assemblage B are unclear due to the large host range of this assemblage (Monis et al., 1999; Monis et al., 2003; *c.f.* Monis and Thompson, 2003), but genetic diversity of *Giardia* isolates is expected to be higher, in regions of high prevalence due to increased transmission, competitive parasite interactions, and selective pressures (Thompson and Monis, 2012). Environmental conditions in remote Indigenous communities, such as overcrowded living conditions, frequent host contact, and environmental contamination, may contribute to the higher proportion of mixed assemblage B populations that were detected among children of this community. A limitation of comparing these results is that both NSW and Western Australian studies have included clinically confirmed cases only, where as all samples (microscopically positive/negative) were included in the analyses of remote community children. The distributions demonstrated by these studies may be relative to the inclusion of mostly symptomatic or asymptomatic cases from each community.

In this thesis, the diversity of *Giardia* among the children of this remote community, was investigated further my multilocus NGS (Chapter 5), and is the first application of multilocus NGS to *Giardia* typing in an epidemiological context. Although the study was

limited by sample numbers and sequence quality, ten cases from this community were successfully genotyped. Our results demonstrated eight different *Giardia* multilocus genotypes, including three subassemblage AII MLGs from five children, and five assemblage B MLGs from five children. The number of MLGs detected among children in this community was high, however, due to the limited sample numbers it is unknown how many cases in the community would be represented by these MLGs. A recent large study of Swedish patients identified 19 different A or B MLGs among 98 patients (Lebbad et al., 2011), and our results suggest a high level of diversity exists among *Giardia* cases from this remote community. The results provide the first report of AII subtypes associated with human cases in Australia. Subtypes AII-1 (n = 1), AII-2 (n = 3), and AII-3 (n = 1) were identified, and have been previously reported from human cases in Sweden, Belgium, and Italy (Caccio et al., 2008; Geurden et al., 2009; Lebbad et al., 2011). The large number of AII-1 and AII-2 cases that were detected among 57 patients in Sweden (Lebbad et al., 2011) may indicate they are common among human populations. Furthermore, detection of these subtypes in Australia demonstrates large geographical dispersal among human populations.

Among assemblage B cases, phylogenetic analyses demonstrated a potential disease link between two assemblage B MLGs, which were closely related, whilst the remaining three MLGs were genetically diverse. Unfortunately, our analyses by NGS was limited by both sample number and sequencing quality, and we were unable to differentiate variant subpopulations, or determine the the prevalence of these MLGs among the samples. Of the assemblage B cases in which genotypes could not be resolved, mixed population detection for the same samples by *gdh* T-RFLP would support these cases being mixed templates. Additionally, evidence of concurrent assemblage A and B populations detected from one case in this community was provided by NGS detection, by identification of different assemblage A and B read populations from a single sample. Few studies have

identified human cases of *Giardia* using three loci, and the prevalence of assemblage A subtypes among human populations, and their global distributions are not yet known.

6.3 Application of new tools for *Giardia* identification

Development and application of T-RFLP to *Giardia* typing (Chapter 3, Chapter 4) in this study demonstrated improved sensitivity and simplicity by comparison to RFLP, and improved discrimination of mixed populations by comparison to DNA sequencing. These studies describe the first application of T-RFLP to *Giardia* identification, and the first application of T-RFLP to epidemiological surveys. Use of this method provided four significant findings that were related to the high sensitivity of fluorescent detection by T-RFLP, including; 1) detection of separate populations of assemblage B genotypes (BIII and BIV) contributing to a single infection, through the generation of BIII and BIV-specific T-RFs, which would not appear among homogenous populations (Asher et al., 2012b); 2) detection of low level and unique genetic variants; 3) the ability to compare unique infection profiles across cases in each community application; and 4) unambiguous detection of mixed BIII/BIV templates, which, by comparison to DNA sequencing can be problematic due to PCR artefacts in DNA chromatograms (Lebbad et al., 2008).

Application of NGS to multilocus typing using four *Giardia* loci (Chapter 5) demonstrated the high diagnostic utility of this tool for *Giardia* identification. Population analyses by next generation sequencing provided strong evidence that assemblage A subtypes are highly similar, contrasting the diversity found among assemblage B isolates. Mixed assemblage A and B populations contributing to individual cases were only detectable by NGS analyses, and demonstrated the high sensitivity of this technology.

Several limitations of the T-RFLP and NGS method are evident; for TRFLP, the lack of DNA sequence information and variant detection beyond the position of the terminal restriction site (Koehler et al., 2014); and the lack of informative sequence data that is

generated for comparative analyses among studies (Koehler et al., 2014), whilst for NGS data analyses remains complex and labour intensive. In this study, the application of T-RFLP and NGS to the *gdh* (T-RFLP, NGS), *β-giardin* and *tpi* loci (NGS) resulted in fewer positive samples available for analyses by T-RFLP or NGS, due to the lower sensitivity of these PCR assays when compared to 18S rRNA PCR. Application of these protocols to the 18S rRNA locus alone, however, would limit the interpretation of results, due to the lack of informative sequence information generated from the 18S rRNA, which does not allow detection to the subassemblage level (Traub et al., 2009). The molecular typing techniques used in this thesis were chosen specifically to evaluate the sensitivity and discriminatory power of new and existing screening methods for *Giardia*, and to assess the utility of these methods for *Giardia* identification in diverse epidemiological foci. Although the use of different typing techniques limits comparison between studies, the development of new molecular tools in this thesis has facilitated new avenues for future research, provided a framework in which to further examine molecular screening approaches, and has identified highly sensitive population based screening approaches as key to understanding the epidemiology of *Giardia*. The results of this thesis have also highlighted typing loci and the inefficiency of the PCR assays as key limitations in *Giardia* research. Currently, molecular analyses of more variable, single copy loci, are required to identify isolates to the subassemblage and subtype level and to facilitate comparison of genotyping results between studies. Discrimination of *Giardia* subassemblages/subtypes is essential for identifying genetic variants from diverse host species, and to elucidate anthroponotic, zoonotic, and environmental disease sources and transmission cycles (Prystajecky et al., 2015; Thompson and Ash, 2015). Inconsistent amplification of typing loci using established PCR protocols for *Giardia* are also commonly reported (Caccio and Ryan, 2008), and was demonstrated by sample screenings performed for this thesis (Chapter 2), and by previous research conducted using the same PCR protocols (Asher

2012a). Several future research avenues are suggested to address this; 1) a comprehensive evaluation of DNA extraction and PCR screening approaches for *Giardia*, including new target loci and/or multicopy genes, which demonstrate enhanced PCR screening sensitivity (Thompson and Ash, 2015); 2) targeted screening of assemblage B isolates, including whole genome sequencing studies and epidemiological studies of assemblage B, to evaluate genetic variability of this assemblage, mixed templates and assemblage swapping, and the presence of genetic subtypes among diffent host species in epidemiological foci; and 3) application of highly sensitive, population based molecular detection tools for analyses of PCR amplicons, to evaluate the presence of mixed templates which contribute to intermittent and inconsistent screening results. The multicopy ITS region would be an optimal target for PCR amplification and genotyping of *Giardia* isolates, due to higher sequence variability than the 18S rRNA, which facilitates subassemblage typing (*c.f.* Thompson and Ash, 2015). A recent whole genome study has identified 70 new genotyping genes with potential use in epidemiological studies, specifically to discriminate assemblage B isolates (Weilinga et al., 2015) and these targets should be considered in future studies. Additional targets include genes encoding variable surface proteins (VSPs), which have been identified as regions of genetic diversity among *Giardia* isolates of subassemblage AII and assemblage B, and among isolates identified as subtype AII-1 and AII-2 (Adam et al., 2013; Ankarklev et al., 2015). These approaches will enhance future understanding of *Giardia* prevalence, genetic subtypes that infect different host species, and improve our understanding of the epidemiology of giardiasis, through increased of subassemblage/subtype prevalence and abundance, and knowledge of potential disease sources and transmission cycles between host species.

This thesis describes the first application of multilocus NGS to *Giardia* identification, and with further development of the protocol, this technology has the potential to facilitate genetic characterisation of assemblage B subtypes, required to progress molecular based

studies of *Giardia*. With increased knowledge of *Giardia* diversity and genetic substructuring within each assemblage, T-RFLP would be highly adaptable for subtype detection. A multilocus protocol for T-RFLP is recommended and would provide a highly sensitive and simplistic approach for routine epidemiological surveys, and to disease source tracking for giardiasis.

6.4 Conclusions and future directions

The results of this thesis demonstrate complex disease links between the diversity of *Giardia*, prevalence and frequency of human giardiasis transmission, and environmental conditions in Australian communities that contribute to the dispersal of different types of human infective *Giardia*. The findings of the NSW investigations show demographic and spatio-temporal dispersal of human assemblage A occurring in the Australian environment, and have demonstrated important disease links among human infective assemblages and transmission to human and non-human hosts. These results provide several key avenues in which to pursue research into the dispersal of *Giardia* subtypes among human populations.

Additional zoonotic sources of giardiasis in NSW and remote Indigneous communities should be investigated. Limited evidence of zoonotic transmission between dogs and humans of a remote Indigenous community of Western Australia has been provided (Hopkins et al., 1997), and recent studies suggest that host contact influences zoonotic transmission of *Giardia* assemblages between humans and non-human hosts (Delport et al., 2014; *c.f* Feng and Xiao, 2011). Directional transmission, between humans and non-human hosts in Australian communities warrants further investigation. Furthermore, the results of this thesis highlights seasonal distributions of *Giardia* assemblages contributing to human infections in NSW, and these distributions should be examined among animal hosts in future investigations.

Essential to *Giardia* epidemiological research is the ability to detect and identify *Giardia* isolates from environmental sources. Future research should be directed towards

standardising molecular detection protocols, and developing systematic approaches to assemblage B subtype identification.

The results of the remote community investigations have provided several significant insights into the contribution of *Giardia* to Indigenous child health issues in remote communities of Australia. Our results have demonstrated high carriages among children of this community and suggest both molecular based, and conventional screening approaches, will be needed to accurately determine *Giardia* prevalence among remote Indigenous populations. Molecular based detection should also be considered in evaluating clinical manifestations and causative agents, particularly among young children who are most at risk from enhanced symptoms. This is particularly applicable to wider global child health issues, where underlying *Giardia* infections may be difficult to detect, but increase the risk of acute and serious gastroenteritis (Grimwood and Forbes, 2009).

Disease management in remote communities is key to improving the health of Indigenous children, and Indigenous adults living in the communities. These data suggest cases may go undetected, and have provided a potential avenue in which to address *Giardia* transmission in remote Indigenous communities. The results of this thesis progress global knowledge of *Giardia*, by demonstrating large differences in *Giardia* cases that occur among children of low and high prevalence communities of Australia. The application of multilocus NGS has demonstrated potential for further development and application of this tool, for population-based screening of *Giardia* in epidemiological research.

6.5 References

Abe, N., Tanoue, T., Noguchi, E., Ohta, G., Sakai, H., 2010. Molecular characterization of *Giardia duodenalis* isolates from domestic ferrets. Parasitol Res 106, 733-736.

Abeywardena, H., Jex, A.R., Koehler, A.V., Rajapakse, R.P.V.J., Udayawarna, K., Haydon, S.R., Stevens, M.A., Gasser, R.B., 2014. First molecular characterization of *Cryptosporidium* and *Giardia* from bovines (*Bos taurus* and *Bubalus bubalis*) in Sri Lanka: unexpected absence of *C. parvum* from pre-weaned calves. Parasit Vectors 7:75.

ACT Government Health. 2015. Available from: <http://www.health.act.gov.au/public-information/public-health/disease-surveillance>. [20 March 2015].

Adam, R.D., 2001. Biology of *Giardia lamblia.* Clin Microbiol Rev 14, 447-475.

Adam, R.D., Dahlstrom, E.W., Martens, C.A., Bruno, D.P., Barbian, K.D., Ricklefs, S.M., Hernandez, M.M., Narla, N.P., Patel, R.B., Porcella, S.F., Nash, T.E., 2013. Genome Sequencing of *Giardia lamblia* Genotypes A2 and B Isolates (DH and GS) and Comparative Analysis with the Genomes of Genotypes A1 and E (WB and Pig). Genome Biol Evol 5, 2498-2511.

Ajjampur, S.S., Sankaran, P., Kanna, A., Sathyakumar, K., Sarkar, R., Gladstone, B.P., Kang, G., 2009. *Giardia duodenalis* assemblages associated with diarrhea in children in South India identified by PCR-RFLP. Am J Trop Med Hyg 80, 16-19.

Ajjampur, S.S.R., Rajendran, P., Ramani, S., Banerjee, I., Monica, B., Sankaran, P., Rosario, V., Arumugam, R., Sarkar, R., Ward, H., Kang, G., 2008. Closing the diarrhoea diagnostic gap in Indian children by the application of molecular techniques. J Med Microbiol 57, 1364-1368.

Amar, C.F.L., Dear, P.H., Pedraza-Diaz, S., Looker, N., Linnane, E., McLauchlin, J., 2002. Sensitive PCR-Restriction Fragment Length Polymorphism Assay for Detection and Genotyping of *Giardia duodenalis* in Human Feces. J Clin Microbiol 40, 446-452.

Andersson, J.O., 2012. Double peaks reveal rare diplomonad sex. Trends Parasitol 28, 46-52.

Andrews, R. H., Adams, M., Boreham, P.F.L., Mayrhofer, G., Meloni, B.P., 1989. *Giardia intestinalis*: electrophoretic evidence for a species complex. Int J Parasitol 19, 183-190.

Ankarklev, J., Franzen, O., Peirasmaki, D., Jerlstrom-Hultqvist, J., Lebbad, M., Andersson, J., Andersson, B., Svard, S.G., 2015. Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. BMC Genomics 16:697.

Ankarklev, J., Jerlstrom-Hultqvist, J.J., Ringqvist, E., Troell, K., Svard, S., 2010. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat Rev Microbiol 8, 413-422.

Ankarklev, J., Svard, S.G., Lebbad, M., 2012. Allelic sequence heterozygosity in single *Giardia* parasites. BMC Microbiol 2012, 12:65.

Aronson, N.E., Cheney, C., Rholl, V., Burris, D., Hadro, N., 2001. Biliary Giardiasis in a Patient With Human Immunodeficiency Virus. J Clin Gastroenterol 33, 167-170.

Asher, A.J., Holt, D.C., Andrews, R.M., Power, M.L., 2014. Distribution of *Giardia duodenalis* Assemblages A and B among Children Living in a Remote Indigenous Community of the Northern Territory, Australia. PloS One 9, e112058.

Asher, A.J., Waldron, L.S., Power, M.L., 2012a Evaluation of a PCR protocol for sensitive detection of *Giardia intestinalis* in human faeces. Parasitol Res 110, 853-858.

Asher, A.J., Waldron, L.S., Power, M.L. 2012b. Rapid identification of *Giardia duodenalis* assemblages in NSW using terminal-restriction fragment length polymorphism. Parasitology 139, 1005-1013.

Australian Government Bureau of Meteorology, 2011. Seasonal Climate Summary for New South Wales 'NSW in autumn 2011: 6th coldest Autumn on record.' Product code IDCKGC15R0. Available from; <http://www.bom.gov.au/climate/current/season/nsw/archive/201105.summary.shtml > [29 November 2015].

Australian Government Bureau of Meteorology, 2012. Seasonal Climate Summary for New South Wales 'NSW in autumn 2012: starts with floods, ends with dry.' Product code IDCKGC15R0. Available from; <http://www.bom.gov.au/climate/current/season/nsw/archive/201205.summary.shtml > [29 November 2015].

Australian Bureau of Statistics, 2012. 1301.0-Year Book Australia, 2012. Available from:<http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1301.0~2012~Main%20Features~Geographic%20distribution%20of%20the%20population~49>. [30 November 2014].

Australian Institute of Health and Welfare, 2011. The health and welfare of Australia's Aboriginal and Torres Strait Islander people, an overview 2011. Cat. no. IHW 42. Canberra: AIHW. Available from: <http://www.aihw.gov.au/publication-detail/?id=10737418989>. [20 May 2014].

Baruch, A.C., Isaac-Renton, J., Adam, R.D., 1996. The molecular epidemiology of *Giardia lamblia*: a sequence-based approach. J Infect Dis 174, 233-236.

Bertrand, I., Albertini, L., Schwartzbrod, J., 2005. Comparison of Two Target Genes for Detection and Genotyping of *Giardia lamblia* in Human Feces by PCR and PCR-Restriction Fragment Length Polymorphism. J Clin Microbiol 43, 5940-5944.

Berrilli, F., D'Alfonso, R., Giangaspero, A., Marangi, M., Brandonisio, O., et al., 2012. *Giardia duodenalis* genotypes and *Cryptosporidium* species in humans and domestic

animals in Côte d'Ivoire: occurrence and evidence for environmental contamination. Trans R Soc Trop Med Hyg 106, 191-195.

Boers, S.A., van der Reijden, W.A., Jansen, R., 2012. High-Throughput Multilocus Sequence Typing: Bringing Molecular Typing to the Next Level. PloS One 7:e39630.

Bonhomme, J., Le Goff, L., Lemee, V., Gargala, G., Ballet, J.J., Favennec, L., 2011. Limitations of *tpi* and *bg* genes sub-genotyping for characterization of human *Giardia duodenalis* isolates. Parasitol Int 60, 327-330.

Breathnach, A.S., McHugh, T.D., 2010. Prevalence and clinical correlations of genetic subtypes of *Giardia lamblia* in an urban setting. Epidemiol Infect 138, 1459-1467.

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., 2011. Targeted Amplicon Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa Phylogenetics. Genome Biol Evol 3, 1312-1323.

Caccio, S.M., Beck, R., Almeida, A., Bajer, A., Pozio, E., 2010. Identification of *Giardia* species and *Giardia duodenalis* assemblages by sequence analysis of the 5.8S rDNA gene and internal transcribed spacers. Parasitology 137, 919-925.

Caccio, S.M., Beck, R., Lalle, M., Marinculic, A., Pozio, E., 2008. Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B. Int J Parasitol 38, 1523-1531.

Caccio, S.M., Ryan, U., 2008. Molecular epidemiology of giardiasis. Molec Biochem Parasitol 160, 75-80.

Caccio, S.M., Sprong, H., 2010. *Giardia duodenalis*: Genetic recombination and its implications for taxonomy and molecular epidemiology. Exp Parasitol 124, 107-112.

Caccio, S.M., Thompson, R.C.A., McLauchlin, Smith, H.V., 2005. Unravelling *Cryptosporidium* and *Giardia* epidemiology. Trends Parasitol 21, 430-437.

Centre for Disease Control, Northern Territory, 2008. The Northern Territory Disease Control Bulletin 15, 1-38.

Centers for Disease Control and Prevention, 2012. Morbidity and Mortality Weekly Report (MMWR) Cryptosporidiosis Surveillance – United States, 2009 - 2010 and Giardiasis Surveillance – United States, 2009 - 2010. Surveillance Summaries 61, 12-24.

Chen, Y., Frazzitta, A.E., Litvintseva, A.P., Fang, C., Mitchell, T.G., Springer, D.J., Ding, Y., Yuan, G., Perfect, J.R., 2015. Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequencing typing. Fungal Genet Biol 75, 64-71.

Choy, S.H., Al-Mekhlafi, H.M., Mahdy, M.A.K., Nasr, N.N., Sulaiman, M., Lim, Y.A.L., Surin, J.,2014. Prevalence and Associated Risk Factors of *Giardia* Infection among Indigenous Communities in Rural Malaysia. Sci Rep 4:6909, doi: 10.1038/srep06909

Choy, S. H.,Mahdy, M.A.K., Al-Mekhlafi, H.M., Nasr, N.A., Lim, Y.A.L., Mahmud, R., Surin, J., 2013. Multilocus genotyping of *Giardia* duodenalis in Malaysia, Infect Genet Evol 17, 269-276

Cifuentes, E., Gomez, M., Blumenthal, U., Tellez-Rojo, M.M., Romieu, I., Ruiz-Palacios, G., Ruiz-Velazco, S., 2000. Risk factors for *Giardia intestinalis* infection in agricultural villiages practicing wastewater irrigation in Mexico. Am J Trop Med Hyg 62, 388-392.

Coklin, T., Farber, J., Parrington, L., Dixon, B., 2007. Prevalence and molecular characterization of *Giardia duodenalis* and *Cryptosporidium* spp. in dairy cattle in Ontario, Canada. Vet Parasitol 150, 297-305.

Cook, D.M., Swanson, C., Eggett, D.L., Booth, G.M., 2009. A Retrospective Analysis of Prevalence of Gastrointestinal Parasites among School Children in the Palajunoj Valley of Guatemala. J Health Popul Nutr 27, 31-40.

Cooper, M.A., Adam, R.D., Worobey, M., Sterling, C.R., 2007. Population genetics provides evidence for recombination in *Giardia*. Curr Biol 17, 1984-1988.

Cotton, J.A., Beatty, J.K., Buret, A.G., 2011. Host parasite interactions and pathophysiology in *Giardia* infections. Int J Parasitol 41, 925-933.

Cox, F.E.G., 2004. Parasitic Protozoa, I. In Cox, F.E.G., (Ed.), *Modern Parasitology: A text book of Parasitology*. Blackwell Science Ltd. Oxford. pp10.

Craun, G.F., Brunkard, J.M., Yoder, J.S., Roberts, V.A., Carpenter, J., Wade, T., Calderon, R.L., Roberts, J.M., Beach, M.J., Roy, S.L., 2010. Causes of Outbreaks Associated with Drinking Water in the United States from 1971 to 2006. Clin Microbiol Rev 23, 507-528.

Delport, T.C., Asher, A.J., Beaumont, L.J., Webster, K.N., Harcourt, R.G., Power, M.L., 2014. *Giardia duodenalis* and *Cryptosporidium* occurrence in Australian sea lions (*Neophoca cinerea*) exposed to varied levels of human interaction. Int J Parasitol Parasites Wildl 3, 269-275.

Department of Health and Human Services, Tasmania, 2010. Guidelines for Notifiable Diseases, Human Pathogenic Organisms and Contaminants. Available from: <http://www.dhhs.tas.gov.au/peh/communicable_diseases_prevention_unit/infectious_diseases>. [20 March 2015].

Dobell, C., 1920. The Discovery of the Intestinal Protozoa of Man. Proc R Soc Med 13, 1-15.

Durigan, M., Abreu, A.G., Zucchi, M.I., Franco, R.M.B., de Souza, A.P., 2014. Genetic Diversity of *Giardia duodenalis*: Multilocus Genotyping Reveals Zoonotic Potential between Clinical and Environmental Sources in a Metropolitan Region of Brazil. PLoS One 9: e115489. doi:10.1371/ journal.pone.0115489.

ElBakri, A., Samie, A., Bessong, P., Potgieter, N., Abu Odeh, R., 2014. Detection and molecular characterisation of *Giardia lamblia* genotypes in Sharjah, United Arab Emirates. Trans R Soc Trop Med Hyg 108, 466-473.

Escobedo, A.A., Cimerman, S., 2007. Giardiasis: a pharmacotherapy review. Expert Opin Pharmacother 8, 1885-1902.

Espelage, W., an der Heiden, M., Stark, K., Alpers, K., 2010. Characteristics and risk factors for symptomatic *Giardia lamblia* infections in Germany. BMC Public Health 10:41.

European Centre for Disease Prevention and Control, Annual epidemiological report 2014 – food- and waterborne diseases and zoonoses. Stockholm: ECDC; 2014.

Ey, P., Mansour, M., Kulda, J., Nohynkova, E., Monis, P.T., Andrews, R.H., Mayrhofer, G., 1997. Genetic Analysis of *Giardia* from Hoofed Farm Animals Reveals Artiodactyl-Specific and Potentially Zoonotic Genotypes. J Eukaryot Microbiol 44, 626-635.

Feasey, N., Wansbrough-Jones, M., Mabey, D.C.W., Soloman, A.W., 2010. Neglected tropical diseases. Br Med Bull 93, 179-200.

Feng, Y., Xiao, L., 2011. Zoonotic Potential and Molecular Epidemiology of *Giardia* Species and Giardiasis. Clin Microbiol Rev 24, 110-140.

Foronda, P., Bargues, M.D., Abreu-Acosta, N., Periago, M.V., Valero, M.A., et al., 2008. Identification of genotypes of *Giardia intestinalis* of human isolates in Egypt. Parasitol Res 103, 1177-1181.

Geurden, T., Levecke, B., Caccio, S.M., Visser, A., De Groote, G., Casaert, S., Vercruysse, J., Claerebout, E., 2009. Multilocus genotyping of *Cryptosporidium* and *Giardia* in non-outbreak related cases of diarrhoea in human patients in Belgium. Parasitology 136, 1161-1168.

Ghosh, S., Debnath. A., Sil, A., De, S., Chattopadhyay, D.J., et al., 2000. PCR detection of *Giardia lamblia* in stool: targeting intergenic spacer region of multicopy rRNA gene. Mol Cell Probes 14: 181-189.

Goto, R., Panter-Brick, C., Northrop-Clewes, C.A., Manahdhar, R., Tuladha, N.R., 2002. Poor intestinal permeability in mildly stunted Nepali children: associations with weaning practices and *Giardia lamblia* infection. B J Nutr 88, 141-149.

Gracey, M., Cullinane, J., 2003. Gastroenteritis and environmental health among Aboriginal infants and children in Western Australia. J Paediatr Child Health 39, 427–31.

Grimwood, K., Forbes, D.A., 2009. Acute and Persistent Diarrhea. Pediatr Clin N Am 56, 1343-1361.

Grinberg, A., Biggs, P.J., Dukkipati, V.S., George, T.T., 2013. Extensive intra-host genetic diversity uncovered in *Cryptosporidium parvum* using Next Generation Sequencing. Infect Genet Evol 15, 18-24.

Hanevik, K., Dizdar, V., Langeland, N., Hausken, T., 2009. Development of functional gastrointestinal disorders after *Giardia lamblia* infection. BMC Gastroenterol 9:27, doi:10.1186/1471-230X-9-27.

Hanevik, K., Hausken, T., Morken, M.H., Strand, E.A., Morch, K., Coll, P., Helgeland, L., Langeland, N., 2007. Persisting symptoms and duodenal inflammation related to *Giardia duodenalis* infection. J Infect 55, 524-530.

Holberton, D.V., Marshall, J., 1995. Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. Nucleic Acids Res 23, 2945-2953.

Holt, D.C., McCarthy, J.S., Carapetis, J.R., 2010. Parasitic diseases of remote Indigenous communities in Australia. Int J Parasitol 40, 1119-1126.

Homan, W.L., Mank, T.G., 2001. Human giardiasis: genotype linked difference in clinical symptomatology. Int J Parasitol 31, 822-826.

Homan, W.L., van Enckevort, F.H.J., Limper, L., van Eys, G.J.J.M., Schoone, G.J., Kasprzak, W., Majewska, A.C., van Knapen, F., 1992. Comparison of *Giardia* isolates from different laboratories by isoenzyme analysis and recombinant DNA probles. Parasitol Res 78, 316-323.

Hopkins, R.M., Constantine, C.C., Groth, D.A., Wetherall, J.D., Reynoldson, J.A., et al., 1999. PCR-based DNA fingerprinting of *Giardia duodenalis* isolates using the intergenic rDNA spacer. Parasitology 118, 531-539.

Hopkins, R.M., Meloni, B.P., Groth, D.M., Wetherall, J.D., Reynoldson, J.A., et al., 1997 Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. J Parasitol 83, 44-51.

Hoque, M.E., Hope, V.T., Kjellstrom, T., Scragg, R., Lay-Yee, R., 2002. Risk of giardiasis in Aucklanders: a case-control study. Int J Infect Dis 6, 191-197.

Kearns, T., Andrews, R., Speare, R., Cheng, A., McCarthy, J., et al., 2009. Beating scabies and strongyloidiasis in the Northern Territory, Australia with an ivermectin MDA. Trop Med Int Health 14(S2): 196.

Kosuwin, R., Putaporntip, C., Pattanawong, U., Jongwutiwes, S., 2010. Clonal diversity in *Giardia duodenalis* isolates from Thailand: evidences for intragenic recombination and purifying selection at the beta giardin locus. Gene 449, 1-8.

Kukuruzovic, R., Robins-Browne, R.M., Anstey, N.M., Brewster, D.R., 2002. Enteric pathogens, intestinal permeability and nitric oxide production in acute gastroenteritis. Pediatr Infect Dis J 21, 730-739.

Ignatius, R., Gahutu, J.B., Klotz, C., Steininger, C., Shyirambere, C., et al., 2012. High Prevalence of *Giardia duodenalis* Assemblage B Infection and Association with Underweight in Rwandan Children. PLoS Negl Trop Dis 6: e1677.

Inpankaew, T., Schar, F., Odermatt, P., Dalsgaard, A., Chimnoi, W., Khieu, V., Muth, S., Traub, R., 2014. Low risk for transmission of zoonotic *Giardia duodenalis* from dogs to humans in rural Cambodia. Parasites and Vectors 7:412.

Jerlstrom-Hultqvist, J., Franzen, O., Ankarklev, J., Xu, F., Nohynkova, E., Andersson, J.O., Svard, S.G., Andersson, B., 2010. Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. BMC Genomics 11:543.

Johnston, A.R., Gillespie, T.R., Rwego, I.B., McLachlan, T.L.T., Kent, A.D., Goldberg, T.L., 2010. Molecular Epidemiology of Cross-Species *Giardia duodenalis* Transmission in Western Uganda. PloS Negl Trop Dis 4, e683.

Jones, H.I., 1980. Intestinal parasite infections in Western Australian Aborigines. Med J Aust 2, 375-380.

Kang, G., Mathew, M.S., Rajan, D.P., Daniel, J.D., Mathan, M.M., Mathan, V.I., Muliyil, J.P., 1998. Prevalence of intestinal parasites in rural Southern Indians. Trop Med Int Health 3, 70-75.

Karanis, P., Kourenti, C., Smith, H., 2007. Waterborne transmission of protozoan parasites: a worldwide review of outbreaks and lessons learnt. J Water Health 5, 1-38.

Koehler, A.V., Jex, A.R., Haydon, S.R., Stevens, M.A., Gasser, R.B., 2014. *Giardia*/giardiasis – A perspective on diagnostic and analytical tools. Biotechnol Adv 32, 280-289.

Lalle, M., Bruschi, F., Castagna, B., Campa, M., Pozio, E., Caccio, S.M., 2009. High genetic polymorphism among *Giardia duodenalis* isolates from Sahrawi children. Trans R Soc Trop Med Hyg 103, 834-838.

Lalle, M., Frangipane di Regalbono, A., Poppi, L., Nobili, G., Tonanzi, D., Pozio, E., Caccio, S.M., 2007. A novel *Giardia duodenalis* assemblage A subtype in fallow deer. J Parasitol 93, 426-428.

Lalle, M., Pozio, E., Capelli, G., Bruschi, F., Crotti, D., Caccio, S.M., 2005. Genetic heterogeneity at the *β-giardin* locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes. Int J Parasitol 35, 207-213.

Lanata, C.F., Fishcer-Walker, C.L., Olascoaga, A.C., Torres, C.X., Aryee, M.J., Black, R.E., 2013. Global Causes of Diarrheal Disease Mortality in Children, 5 Years of Age: A Systematic Review. PloS One 8:e72788.

Lane, S., Lloyd, D., 2002. Current Trends in Research into the Waterborne Parasite *Giardia.* Crit Rev Microbiol 28, 123-147.

Lasek-Nesselquist, E., Welch, D.M., Sogin, M.L., 2010. The identification of a new *Giardia duodenalis* assemblage in marine vertebrates and a preliminary analysis of *G. duodenalis* population biology in marine systems. J Parasitol 40, 1063-1074.

Laupland, K.B., Church, D.L., 2005. Population-based laboratory surveillance for *Giardia* sp. and *Cryptosporidium* sp. infections in a large Canadian health region. BMC Infect Dis 5:72, doi: 10.1186/1471-2334-5-72.

Learmonth, J.J., Ionas, G., Pita, A.B., Cowie, R.S., 2003. Identification and genetic characterisation of *Giardia* and *Cryptosporidium* strains in humans and dairy cattle in the Waikato Region of New Zealand. Water Sci Technol 47, 21-26.

Lebbad, M., Ankarklev, J., Tellez, A., Leiva, B., Andersson, J.O., Svard, S., 2008. Dominance of *Giardia* assemblage B in Leon, Nicaragua. Acta Trop 106, 44-53.

Lebbad, M., Petersson, I., Karlsson, L., Botero-Kleiven, S., Andersson, J.O., Svenungsson, B., Svard, S.G., 2011. Multilocus Genotyping of Human *Giardia* Isolates Suggests Limited Zoonotic Transmission and Association between Assemblage B and Flatulence in Children. PloS Negl Trop Dis 5:e1262.

Lederberg, J., 1996. Infectious Disease – A Threat to Global health and Security. JAMA 276, 417-419.

Lipoldova, M., 2014. *Giardia* and Vilem Dusan Lambl. PloS Negl Trop Dis 8, e2628.

Lott, M.J., Hose, G.C., Power, M.L., 2015. Parasitic nematode communities of the red kangaroo, *Macropus rufus*: richness and structuring in captive systems. Parasitol Res 114, 2925-2932.

Magnoc, T., Salzberg, S., 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957-2963.

Mahdy, A.K.M., Lim, Y.A.L., Surin, J., Wan, K.L., Al-Mekhlafi, M.S.H., 2008. Risk factors for endemic giardiasis: highlighting the possible association of contaminated water and food. Trans R Soc Trop Med Hyg 102, 465—470.

Marsh, T., 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterising diversity among homologous populations of amplification products. Curr Opin Microbiol 2, 323-327.

Mayrhofer, G., Andrews, R.H., Ey, P.L., Chilton, N.B., 1995. Division of *Giardia* isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with *Giardia muris*. Parasitology 111, 11-17.

McDonald, E., Bailie, R., 2010. Hygiene improvement: Essential to improving child health in remote Aboriginal communities. J Paediatr Child Health 46: 491-496.

McDonald, E., Bailie, R., Brewster, D., Morris, P., 2008. Are hygiene and public health interventions likely to improve outcomes for Australian Aboriginal children living in remote communities? A systematic review of the literature. BMC Public Health doi:10.1186/1471-2458-8-153.

McMichael, A.J., Woodruff, R.E., Hales, S., 2006. Climate change and human health: present and future risks. Lancet 367, 859-869.

Meloni, B.P., Thompson, R.C.A., Hopkins, R.M., Reynoldson, J.A., Gracy, M., 1993. The prevalence of *Giardia* and other intestinal parasites in children, dogs and cats from Aboriginal communities in the Kimberley. Med J Aust 158, 157-159.

Minette, C., Taweenan, W., Hogg, R., Featherstone, C., Randle, N., Lathan, S.M., Wastling, J.M., 2014. Occurrence and Diversity of *Giardia duodenalis* Assemblages in Livestock in the UK. Transbound Emerg Dis doi:10.1111/tbed.12075.

Minvielle, M.C., Molina, N.B., Polverino, D., Basualdo., 2008. First genotyping of *Giardia lamblia* from human and animal feces in Argentina, South America. Mem Inst Oswaldo Cruz, Rio de Janeiro 103, 98-103.

Monis, P.T., Andrews, R.H., Mayrhoffer, G., Mackrill, J., Kulda, J., Isaac-Renton, J.L., Ey, P.L., 1998. Novel lineages of *Giardia intestinalis* identified by genetic analysis of organisms isolated from dogs in Australia. Parasitology 116, 7-19.

Monis, P.T., Andrews, R.H., Mayrhofer, G., Ey, P.L., 1999. Molecular Systematics of the parasitic *Giardia intestinalis.* Mol Biol Evol 16, 1135-1144

Monis, P.T., Caccio, S.M., Thompson, R.C.A., 2009. Variation in *Giardia*: towards a taxonomic revision of the genus. Trends Parasitol 25, 93-100.

Monis, P.T., Mayrhofer, G., Andrews, R.H., Homan, W.L., Limper, L., Ey, P.L., 1996. Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus. Parasitology 112, 1-12.

Monis, P.T., Andrews, R.H., Mayrhofer, G., Ey, P.L., 2003. Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. Infect Genet Evol 3, 29-38.

Monis, P.T., Thompson, R.C.A., 2003. *Cryptosporidium* and *Giardia*-zoonoses: fact or fiction? Infect Genet Evol 3, 233-244.

Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., et al., 2007. Genomic Minimalism in the Early Diverging Intestinal Parasite *Giardia lamblia*. Science 317, 1921-1926.

Mowatt, M.R., Weinbach, E.C., Howard, T.C., Nash, T.E., 1994. Complementation of an Escherichia coli glycolysis mutant by *Giardia lamblia* triosephosphate isomerase. Exp Parasitol 78, 85-92.

Mukherjee, A.K., Chowdhury, P., Rajendran, K., Nozaki, T., Ganguly, S., 2014. Association between *Giardia duodenalis* and Coinfections with Other Diarrhea-Causing Pathogens in India. Biomed Res Int doi:10.1155/2014/786480.

Nachamkin, I., Panara, N.J., Li, M., Ung, H., Yuen, P.K., Kricka, L.J., Wilding, P., 2001. Agilent 2100 bioanalyzer for restriction fragment length polymorphism analysis of the Campylobacter jejuni flagellin gene. J Clin Microbiol 39, 754-757.

Nantavisai, K., Mungthin, M., Tan-ariya, P., Rangsin, R., Naaglor, T., Leelayoova, S., 2007. Evaluation of the Sensitivities of DNA Extraction and PCR Methods for Detection of *Giardia duodenalis* in Stool Specimens. J Clin Microbiol 45, 581-583.

Nash, T.E., McCutcheun, T., Keister, D., Dame, J.B., Conrad, J.D., Gillin, F.D., 1985. Restriction endonuclease analysis of DNA from 15 *Giardia* isolates obtained from humans and animals. J Infect Dis 152, 64-73.

Nash, T.E., Mowatt, M.R., 1992. Identification and characterization of *Giardia lamblia* group-specific gene. Exp Parasitol 75, 369-378.

Nash, T.E., Ohl, C.A., Thomas, E., Subramanian, G., Keiser, P., Moore, T.A., 2001. Treatment of Patients with Refractory Giardiasis. Clin Infect Dis 33, 22-28.

Naumova, E.N., Chen, J.T., Griffiths, J.K., Matyas, B.T., Estes-Smargiassi, S.A., Morris, R.D., 200. Use of Passive Surveillance Data to Study Temporal and Spatial Variation in the Incidence of Giardiasis and Cryptosporidiosis. Public Health Rep 115, 436-447.

Ng, C.T., Gilchrist, C.A., Lane, A., Roy, S., Haque, R., Houpt, E.R., 2005. Multiplex Real-Time PCR Assay Using Scorpion Probes and DNA Capture for Genotype-Specific Detection of *Giardia lamblia* on Fecal Samples. J Clin Microbiol 43, 1256-1260.

Ng, J., Yang, R., Whiffin, V., Cox, P., Ryan, U., 2011. Identification of zoonotic *Cryptosporidium* and *Giardia* genotypes infecting animals in Sydney's water catchments. Exp Parasitol 128, 138-144.

Nolan, M.J., Jex, A.R., Koehler, A.V., Haydon, S.R., Stevens, M.A., Gasser, R.B., 2013. Molecular-based investigation of *Cryptosporidium* and *Giardia* from animals in water catchments in southeastern Australia. Water Res 47, 1726-1740.

Nolan, M.J., Jex, A.R., Pangasa, A., Young, N.D., Campbell, A.J., Stevens, M., Gasser, R.B., 2010. Analysis of nucleotide variation within the triose-phosphate isomerase gene of *Giardia duodenalis* from sheep and its zoonotic implications. Electrophoresis 31, 287-298.

NSW Department of Health, 1998. The Sydney Water Incident: July – September 1998. NSW Public Health Bull 9, 91 – 94.

NSW Department of Health, 2006. Hunter New England Population Health, Communicable Diseases Bulletin, 150, 1-4.

NSW Department of Health, 2010. Year in review: health protection in NSW, 2009. Centre for Health Protection. NSW Public Health Bull 22: 156-166.

NSW Department of Health, 2011. Year in review: health protection in NSW, 2010. Centre for Health Protection. NSW Public Health Bull 23: 137-143.

NSW Department of Health, 2013. Year in review: health protection in NSW, 2012. Centre for Health Protection. NSW Public Health Bull Vol. 24: 105-118.

NSW Health Notifiable Conditions Information Management System (NCIMS)(HOIST), Communicable Diseases Branch and Centre for Epidemiology and Evidence, NSW

Ministry of Health., 2015. Available from: <www0.health.nsw.gov.au/data/diseases/giardiasis.asp> [20 March 15].

Odoi, A., Martin, S.W., Michel, P., Holt, J., Middleton, D., Wilson, J., 2003. Geographical and temporal distribution of human giardiasis in Ontario, Canada. Int J Health Geogr 2, August 2003.

Odoi, A., Martin, S.W., Michel, P., Holt, J., Middleton, D., Wilson, J., 2004. Determinants of the geographical distribution of endemic giardiasis in Ontario, Canada: a spatial modelling approach. Epidemiol Infect 132, 967-976.

Olson, M.E., Goh, J., Phillips, M., Guselle, N., McAllister, T.A., 1999. *Giardia* Cyst and *Cryptosporidium* Oocyst Survival in Water, Soil, and Cattle Feces. J Environ Qual 28, 1991-1996.

Paz e Silva, F.M., Monobe, M.M., Lopes, R.S. and Araujo, J.P. Jr., 2012. Molecular characterization of *Giardia duodenalis* in dogs from Brazil. Parasitol Res 110, 325-334.

Pelayo, L., Nunez, F.A., Rojas, L., et al., 2008 *Giardia* infections in Cuban children: the genotypes circulating in a rural population. Annals Trop Med Parasitol 102, 585-595.

Pilloni, G., Granitsiotis, M.S., Engel, M., Leuders, T., 2012. Testing the Limits of 454 Pyrotag Sequencing: Reproducibility, Quantitative Assessment and Comparison to T-RFLP Fingerprinting of Aquifer Microbes. PloS One doi:10.1371/journal.pone.0040467.

Poxleitner, M.K., Carpenter, M.L., Mancuso, J.J., Wang, C.R., Dawson, S.C., Cande, W.Z., 2008. Evidence for Karyogamy and Exchange of Genetic Material in the Binucleate Intestinal Parasite *Giardia intestinalis*. Science 319, 1530-1533.

Prucca, C.G., Lujan, H.D., 2009. Antigenic variation in *Giardia lamblia*. Cell Microbiol 11, 1706-1715.

Prystajecky, N., Tsui, C.K.M., Hsiao, W.W.L., Uyaguari-Diaz, M.K., Ho, J., Tang, P., Isaac-Renton, J., 2015. Giardia spp. Are Commonly Found in Mixed Assemblages in Surface Water, as Revealed by Molecular and Whole-Genome Characterization. Appl Environ Microbiol 81:4827-4834.

Ramesh, M.A., Malik, S.B., Logsdon, J.M., 2005. A Phylogenomic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis. Curr Biol 15, 185-191.

Read, C.M., Monis, P.T., Thompson, R.C.A., 2004. Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP. Infect Genet Evol 4, 125-130.

Read, C., Walters, J., Robertson, I.D., Thompson, R.C.A., 2002 Correlation between genotype of *Giardia duodenalis* and diarrhoea. Int J Parasitol 32: 229-231.

Rendtorff, R.C., 1954. The experimental transmission of human intestinal protozoan parasites. II. *Giardia lamblia* cysts given in capsules. Am J Hyg 59, 209-220.

Reynoldson, J.A., Behnke, J.M., Gracey, M., Horton, R.J., Spargo, R., et al., 1998. Efficacy of albendazole against *Giardia* and hookworm in a remote Aboriginal community in the north of Western Australia. Acta Trop 71, 27-44.

Robertson, L.J., Hanevik, K., Escobedo, A.A., Morch, K., Langeland, N., 2010. Giardiasis – why do the symptoms sometimes never stop? Trends Parasitol 26, 75-82.

Robertson, L.J., Hermansen, L., Gjerde, B.K., Strand, E., Alvsvag, J.O., Langeland, N., 2006. Application of Genotyping during an Extensive Outbreak of Waterborne Giardiasis in Bergen, Norway, during Autumn and Winter 2004. Appl. Environ. Microbiol. 72, 2212-2217.

Roxstrom-Lindquist, K., Palm, D., Reiner, D., Ringqvist, E., Svard, S.G., 2006. *Giardia* immunity – an update. Trends Parasitol 22, 26-31.

Ryan, U., Caccio, S.M., 2013. Zoonotic potential of *Giardia*. Int J Parasitol 43, 943-956.

Savioli, L., Smith, H., Thompson, A., 2006. *Giardia* and *Cryptosporidium* join the 'Neglected Diseases Initiative.' Trends Parasitol 22, 203-208.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., Quince, C., 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res doi: 10.1093/nar/gku1341

Schutte, U.M.E., Abdo, Z., Bent, S.J., Shyu, C., Williams, C.J., Pierson, J.D., Forney, L.J., 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Appl Microbiol Biotechnol 80, 365-380.

Singh, A., Janaki, L., Petri Jr, W.A., Houpt, E.R., 2009. Short Report: *Giardia intestinalis* Assemblages A and B Infections in Nepal. Am J Trop Med Hyg 81, 538-539.

Smith, H.V., Caccio, S.M., Tait, A., McLauchlin, J., Thompson, R.C., 2006. Tools for investigating the environmental transmission of *Cryptosporidium* and *Giardia* infections in humans. Trends Parasitol 22 160-167.

Souza, S.L.P., Gennari, S.M., Richtzenhain, L.J., Pena, H.F.J., Funada, M.R., et al., 2007. Molecular identification of *Giardia duodenalis* isolates from humans, dogs, cats and cattle from the state of Sa˜o Paulo, Brazil, by sequence analysis of fragments of glutamate dehydrogenase (*gdh*) coding gene. Vet Parasitol 149, 258-264.

Sprong, H., Caccio, S.M., van der Giessen, J.W.B., 2009. Identification of Zoonotic Genotypes of *Giardia duodenalis*. PloS Negl Trop Dis 3:e558.

Stein, P.L., 2000. The Great Sydney Water Crisis of 1998. Water Air Soil Pollut 123, 419-436.

Sulaiman, I.M., Fayer, R., Bern, C., Gilman, R.H., Trout, J.M., Schantz, P.M., Das, P., Lal, A.A., Xiao, L., 2003. Triosephosphate isomerase gene characterization and potential zoonotic transmission of *Giardia duodenalis*. Emerg Infect Dis 9, 1444-1452.

Svard, S.G., Hagblom, P., Palm, J.E., 2003. *Giardia lamblia* – a model organism for eukaryotic cell differentiation. Microbiol Lett 218, 3-7.

Takumi, K., Swart, A., Mank, T., Lasek-Nesselquist, E., Lebbad, M., Caccio, S.M., Sprong, H., 2012. Population-based analyses of *Giardia duodenalis* is consistent with the clonal assemblage structure. Parasit Vectors, 5:168.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 10, 2731-2739.

Teodorovic, S., Braverman, J.M., Elmendorf, H.G., 2007. Unusually Low Levels of Genetic Variation among *Giardia lamblia* Isolates. Eukaryot Cell 6, 1421-1430.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, population-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673-4680.

Thompson, R.C.A., 2000. Giardiasis as a re-emerging infectious disease and its zoonotic potential. Int J Parasitol 30, 1259-1267.

Thompson, R.C.A., 2008. Giardiasis: Modern Concepts in Control and Management. Ann Nestle 66, 23-29.

Thompson, R.C.A., 2013. Parasite zoonoses and wildlife: One health, spillover and human activity. Int J Parasitol 43, 1079-1088.

Thompson, R.C.A., Ash, A., 2015. Molecular epidemiology of Giardia and Cryptosporidium infections. Infect Genet Evol http://dx.doi.org/10.1016/j.meegid.2015.09.028.

Thompson, R.C.A., Hopkins, R.M., Homan, W.L., 2000. Nomenclature and Genetic Groupings of *Giardia* Infecting Mammals. Parasitol Today 16, 210-213.

Thompson, R.C.A., Monis, P., 2012. *Giardia* – From Genome to Proteome. Adv Parasitol doi: 10.1016/B978-0-12-394303-3.00003-7.

Thompson, R.C.A., Reynoldson, J.A., Garrow, S.C., McCarthy, J.S., Behnke, J.M., 2001 Towards the eradication of hookworm in an isolated Australian community. Lancet 357, 770-771.

Traub, R.J., Inpankaew, T., Reid, S.A., Sutthikornchai, C., Sukthana, Y., Robertson, I.D., Thompson, R.C.A., 2009. Transmission cycles of *Giardia duodenalis* in dogs and humans in Temple communities in Bangkok—A critical evaluation of its prevalence using three diagnostic tests in the field in the absence of a gold standard. Acta Trop 111, 125-132.

Traub, R.J., Monis, P.T., Robertson, I.D., 2005a. Molecular epidemiology: a multidisciplinary approach to understanding parasite zoonoses. Int J Parasitol 35, 1295-1307.

Traub, R.J., Monis, P.T., Robertson, I., Irwin, P., Mencke, N., Thompson, R.C.A., 2004. Epidemiological and molecular evidence supports the zoonotic transmission of *Giardia* among humans and dogs living in the same community. Parasitology 128, 253-262.

Traub, R., Wade, S., Read, C., Thompson, A., Mohammed, H., 2005b. Molecular characterization of potentially zoonotic isolates of *Giardia duodenalis* in horses. Vet Parasitol 130, 317-321.

Troeger, H., Epple, H.J., Schneider, T., Wahnschaffe, U., Ullrich, R., Burchard, G.D., Jelinek, T., Zeitz, M., Fromm, M., Schulzke, J.D., 2007. Effect of chronic *Giardia lamblia* infection on epithelial transport and barrier function in human duodenum. Gut 56, 328-335.

Trout, J.M., Santin, M., Greiner, E. Fayer, R., 2004. Prevalence of *Giardia duodenalis* genotypes in pre-weaned dairy calves. Vet Parasitol 124, 179-186.

United States Environmental Protection Agency, 1999. *Giardia*: Risk for Infants and Children. Addendum to the U.S. Environmental Protection Agency's (EPA) Drinking Water Criteria Document on *Giardia* (U.S. EPA,1998). EPA-823-R-99-011.

Upcroft, P., Upcroft, J.A., 2001. Drug Targets and Mechanisms of Resistance in the Anaerobic Protozoa. Clin Microbiol Rev 14, 150-164.

van Keulen, H., Macechko, P.T., Wade, S., et. Al., 2002. Presence of human *Giardia* in domestic, farm and wild animals, and environmental samples suggests a zoonotic potential for giardiasis. Vet Parasitol 108, 97-107.

Voelkerding, K.V., Dames, S., Durtschi, J.D., 2010. Next Generation Sequencing for Clinical Diagnostics-Principles and Application to Targeted Resequencing for Hypertrophic Cardiomyopathy. J Mol Diagn 12, 539-551.

Vermeulen, E.T., Ashworth, D.L., Eldridge, M.D.B., Power, M.L., 2015. Investigation into potential transmission sources of *Giardia duodenalis* in a threatened marsupial (*Petrogale penicillata*). Infect Genet Evol 33, 277-280.

Waldron, L.S., Dimeski, B., Beggs, P.J., Ferrari, B.C., Power, M.L., 2011a. Molecular Epidemiology, Spatiotemporal Analysis, and Ecology of Sporadic Human Cryptosporidiosis in Australia. Appl Environ Microbiol 77, 7757-7765.

Waldron, L.S., Ferrari, B.C., Gillings, M.R., Power, M.L., 2009. Terminal Restriction Fragment Length Polymorphism for Identification of *Cryptosporidium* Species in Human Faces. Appl Environ Microbiol 75, 108-112.

Waldron, L.S., Power, M.L., 2011b. Fluorescence analysis detects *gp60* subtype diversity in *Cryptosporidium* infections. Infect Genet Evolut 11, 1388-1395.

Walker, C.L., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Butta, Z.A., O'Brien, K.L., Campbell, H., Black, R.E., 2013. Global burden of childhood pneumonia and diarrhoea. Lancet 381, 1405-1416.

Wang, H., Zhao, G., Chen, G., Jian, F., Zhang, S., Feng, C., Wang, R., Zhu, J., Dong, H., Hua, J., Wang, M., Zhang, L., 2014. Multilocus Genotyping of *Giardia duodenalis* in Dairy Cattle in Henan, China. PloS One 9, e100453.

Wegayehu, T., Adamu, H., Petros, B., 2013. Prevalence of *Giardia duodenalis* and *Cryptosporidium* species infections among children and cattle in North Shewa Zone, Ethiopia. BMC Infect Dis 13:419.

Weiss, J.B., van Keulen, H., Nash, T.C., 1992. Classification of subgroups of *Giardia lamblia* using the polymerase chain reaction. Mol Biochem Parasitol 54, 73-86.

Winkworth, C.L., Learmonth, J.J., Matthaei, C.D., Townsend, C.R., 2008. Molecular Characterization of *Giardia* Isolates from Calves and Humans in a Region in Which Dairy Farming Has Recently Intensified. Appl Environ Microbiol 74, 5100-5105.

Wielinga, C.M., Thompson, R.C.A., 2007. Comparative evaluation of *Giardia duodenalis* data. Parasitology 134, 1795-1821.

Wielinga, C., Thompson, R.C.A., Monis, P., Ryan, U., 2015. Identification of polymorphic genes for use in assemblage B genotyping assays through comparative genomics of multiple assemblage B *Giardia duodenalis* isolates. Mol Biochem Parasitol 201, 1-4.

World Health Organisation, 1987. Prevention and control of intestinal parasitic infections. World Health Organ Tech Rep Ser 749, 1-86

World Health Organisation. 2013. Fact sheet N°330 'Diarrhoeal disease'. Available from: <www.who.int/mediacentre/factsheets/fs330/en/>. [15 July 2015].

Wright, J.M., Dunn, L.A., Upcroft, P., Upcroft, J.A., 2003. Efficacy of antigiardial drugs. Expert Opin Drug Saf 2, 529-541.

Xiao, L., Fayer, R., 2008. Molecular characterisation of species and genotypes of *Cryptosporidium* and *Giardia* and assessment of zoonotic transmission. Int J Parasitol 38, 1239-1255.

Xu, F., Jerlstrom-Hultqvist, J., Andersson, J.O., 2012. Genome-Wide Analyses of Recombination Suggest That *Giardia intestinalis* Assemblages Represent Different Species. Mol Biol Evol 29, 2895-2898.

Yang, R., Lee, J., Ng, J., Ryan, U., 2010. High prevalence of *Giardia duodenalis* assemblage B and potentially zoonotic subytpes in sporadic human cases in Western Australia. Int J Parasitol 40: 293-297.

Yee, J., Dennis, P.P., 1992. Isolation and characterization of a NADP-dependent glutamatedehydrogenase gene from the primitive eucaryote *Giardia lamblia*. J Biol Chem 267 7539-7544.

Yoder, J.S., Gargano, J.W., Wallace, R.M., Beach, M.J., 2012. Giardiasis Surveillance – United States. MMWR Surveill Summ 61, 13-23.

7: Appendices

7.1 Evaluation of a PCR protocol for sensitive detection of *Giardia intestinalis* in human faeces

A.J. Asher [a], L.S Waldron [a], M.L. Power [a]

Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia

Keywords: *Giardia intestinalis*, PCR, 18S rRNA, *gdh*, faeces

Asher, A.J., Waldron, L.S., Power, M.L., 2012. Evaluation of a PCR protocol for sensitive detection of *Giardia intestinalis* in human faeces. Parasitol Res 110, 853-858.

7.1.1 Abstract

Giardia intestinalis is a protozoan parasite and a human pathogen. It is a leading cause of human diarrheal disease and a significant cause of morbidity worldwide. At the molecular level, *G. intestinalis* is a species complex, consisting of genetic assemblages (A to G) and sub-assemblage strains. The genotypes that cause human disease have been characterised to assemblages A and B, and include strains AI, AII, BIII and BIV. PCR amplification of diagnostic loci is used to genotype samples and is required to understand different transmission cycles within communities. A multilocus approach is required for validation of *Giardia* genotyping and molecular diagnostic techniques that are efficient across numerous loci have not been established. This study evaluated several published protocols for the 18S small subunit ribosomal RNA (18S rRNA) and glutamate dehydrogenase genes (*gdh*) genes. Assays were compared using spiked faecal samples and by measuring the concentration of DNA generated following DNA extraction and PCR amplification. An optimal molecular method for *G. intestinalis* identification was established from direct DNA extraction of faecal material and GC-rich PCR chemistry. The protocol was applied to fifty clinical samples and produced PCR success rates of 90% and 94% at the 18S rRNA and *gdh* loci. Cyst concentration prior to DNA extraction was not necessary and the optimal protocol was highly sensitive and an efficient method for testing clinical samples.

7.1.2 Introduction

Giardia intestinalis (syn. *G. duodenalis, G. lamblia*) is a flagellated, protozoan parasite and a leading cause of human diarrheal disease worldwide (Caccio et al., 2005; Read et al., 2004). Molecular data indicates that *G. intestinalis* is a species complex, consisting of seven genetic assemblages (A to G) that have been isolated from a range of mammalian hosts (Caccio and Ryan 2008; Monis et al., 1999). These assemblages also consist of numerous sub-assemblage strains (Caccio and Ryan 2008; Monis et al., 1999). Human disease appears to be only caused by assemblages A and B, including four strains; AI, AII, BIII and BIV (Monis et al., 1999). Transmission cycles for sub-assemblage strains have not been defined and both anthroponotic and zoonotic transmission is likely (Caccio and Ryan 2008). Clinical diagnosis of giardiasis is confirmed through immuno-detection assays and / or microscopic examination of faecal smears (Traub et al., 2005a). These tests do not involve genetic characterisation and as such, strain information is not collected for human cases (Traub et al., 2005a). Molecular analysis is the only means to identify assemblages and strains, and has become a central component of epidemiological research (Traub et al., 2005a). Published protocols to extract and amplify *G. intestinalis* DNA are numerous, however no standard method exists (Caccio et al., 2005; Nantavisai et al., 2007). For clinical samples PCR remains problematic due to the presence of DNA polymerase inhibitors and degraded DNA in faecal material, non-specific binding of primers, and intermittent parasite shedding (Caccio and Ryan 2008; Nantavisai et al., 2007; Traub et al., 2005a). To overcome PCR inhibition, many protocols incorporate cyst concentration prior to DNA extraction, however it is unclear what combination of molecular protocols is optimal for *Giardia* (Hopkins et al., 1997; Pelayo et al., 2008; Read et al., 2004). Robust, reproducible and highly sensitive techniques are essential for comparisons of regional data and for estimating the contribution of different *G. intestinalis* strains to global human infection (Smith et al., 2006).

In this study, published molecular protocols for *G. intestinalis* identification were evaluated for both the small subunit ribosomal RNA (18S rRNA) and the glutamate dehydrogenase (*gdh*) genes. DNA extraction using FTA filter paper (Whatman Bioscience, Cambridge, United Kingdom) and 18S rRNA amplification have been previously reported as the most sensitive method for *G. intestinalis* detection in faecal samples (Nantavisai et al., 2007). The 18S rRNA gene is highly conserved and can be used to assign *G. intestinalis* to the assemblage level (A-G) (Read et al., 2004). Analysis of clinical samples requires more variable diagnostic loci such as the *gdh* to characterise the sub-assemblage strains that infect humans (Read et al., 2004). To validate *G. intestinalis* genotyping, molecular techniques also need to be efficient across numerous loci (Caccio and Ryan 2008; Read et al., 2004). The purpose of the present study was to evaluate current molecular protocols for detection of *G. intestinalis* in human faeces at the 18S rRNA and *gdh* loci, and to determine an efficient and sensitive method, that improves processing of problematic clinical samples.

7.1.3 Materials and methods

Giardia positive faecal samples (n = 50) were obtained from pathology laboratories in Sydney, Australia and assigned a Macquarie University *Giardia* (MQG) number. Pathology companies had deemed samples positive using a Remel ProSpecT (Thermo Fisher Scientific© 2010) *Giardia*/*Cryptosporidium* microplate immunoassay, and confirmed this result by microscopic examination of faecal smears. Samples were irradiated with ^{60}Colbalt exposure, 1500 curies, for 30 minutes (Macquarie University) to enable processing in non-PC2 laboratories and stored at 4°C until further analysis.

The optimal method for extracting and amplifying *G. intestinalis* DNA at two loci was evaluated using two separate approaches and positive DNA controls that included; pure *G. intestinalis* DNA extracted from laboratory cultivated cysts (+ cysts) (Waterborne

Ltd, New Orleans, LA, USA) and laboratory cultivated trophozoites (+ trophozoites). Spiked faecal material from four clinical samples (MQG17, MQG23, MQG24, MQG27) were also included to assess the clinical performance of each assay in the presence of PCR inhibitors in faecal material.

7.1.3.1 Evaluation of PCR protocols

The first approach evaluated several published PCR protocols following direct DNA extraction using an ISOLATE Fecal DNA Kit (Bioline, Sydney, Australia). Faecal material (50 mg) was directly added to a bead lysis tube and samples were spiked with 625 *G. intestinalis* cysts in PBS (5 μl) and DNA extraction was conducted according to the manufacturer's instructions.

To determine the most sensitive PCR method, all positive controls were amplified at the 18S rRNA locus (approximately 180 bp fragment) following the nested PCR protocol (Hopkins et al., 1997; Read et al., 2002) and using primers RH11, RH4LM, GiAR18SiR and GiAR18SeR. Six DNA polymerases; *Tth* plus DNA polymerase (Fisher Biotech, Australia) (Hopkins et al., 1997; Read et al. 2002; Read et al,. 2004), Red Hot *Taq* (ABgene, Surrey, United Kingdom) (Waldron et al., 2009), *Accutaq* (Sigma-Aldrich, Australia) (Waldron et al., 2009), *Sahara* DNA polymerase (Bioline, United Kingdom) (Waldron et al., 2009), Ampli*Taq* 360 DNA polymerase (Invitrogen) (Caccio et al., 2003; van der Giessen et al., 2006), GC-RICH PCR System, dNTPack (Roche Diagnostics, Indianapolis, IN) (Sulaiman et al., 2003) were tested. All reaction mixtures (25 μl) were prepared following the manufacturer's instructions. To optimise PCR chemistries, primer concentration was varied between 12.5 pmol and 20 pmol. For *Tth* plus, Red Hot *Taq*, *Accutaq* and *Sahara* DNA polymerases, $MgCl_2$ was varied between 1.5 and 6 mM, BSA between 0 and 1 μg/μl, and DMSO was trialled at 2% and 5%. For Ampli*Taq*, 360 GC-enhancer (manufacturer supplied) was tested at 0 μl, 5 μl and 10 μl. For the GC-RICH

PCR System, dNTPack, GC-RICH Resolution Solution (manufacturer supplied) was tested at 0 M, 0.5 M, 1 M and 2 M. All reactions were performed in an Eppendorf Mastercycler (Eppendorf, North Ryde, Australia). PCR products were resolved by agarose gel electrophoresis (2% w/v, 110V for 30 min) in TBE (Tris, EDTA, boric acid, pH 8) with SYBR safe and product size was estimated against a HyperLadder™ II DNA marker (Bioline, Sydney, Australia). Secondary PCR products were compared under UV light and products were purified using a QIAquick PCR Purification Kit (Qiagen, Melbourne, Australia), following the manufacturer's instructions for the Microcentrifuge Bench Protocol. To evaluate the efficiency of each assay, the amplicon concentration (ng/μl) (for one PCR amplification per assay) was measured using a NanoDrop® Spectrophotometer (NanoDrop Technologies, Inc., Wilmington DE, USA). To evaluate the differences in DNA concentrations for secondary PCR products, One-Way ANOVA and Tukey post hoc statistical tests were performed. To evaluate *gdh* PCR, positive controls were amplified using the DNA polymerases that generated the highest 18S rRNA PCR yields, and an optimal PCR method for both loci was determined.

7.1.3.2 Evaluation of DNA extraction techniques

The second approach evaluated DNA extraction techniques using Immunomagnetic Separation (IMS) to concentrate cysts from faeces and remove DNA from PCR inhibitors (Power et al., 2003). The procedure for IMS, as outlined by Power et al. (2003) was followed using *Giardia*-specific mouse IgG1 monoclonal antibody, G203 (100 µg; AusFlow®) to prepare *Giardia*-specific paramagnetic beads. Following filtration of the faecal slurry (1 ml), each sample was spiked with 1250 *G. intestinalis* cysts in PBS (10 ul). Cyst capture involved a primary bead volume of 150 µl and antibody disassociation was not conducted (Power et al., 2003). Following cyst capture the supernatant was removed and cysts bound to beads were washed with water (160 µl; sterile) and resuspended by

vortexing. Purified IMS suspensions (160 μl; sterile H_20) were then divided into two aliquots (80 μl), each containing 625 purified cysts (equivalent to spiked cyst loads in direct ISOLATE fecal DNA extractions). The samples were used for two DNA extraction methods, the ISOLATE Fecal DNA Kit (used previously for direct faecal DNA extractions) and PrepGEM (ZyGEM Corporation Ltd., Hamilton, New Zealand). For the ISOLATE Fecal DNA Kit, one aliquot (80 μl) of IMS purified cysts were directly added to a bead lysis tube and processed according to manufacturer's instructions. For PrepGEM DNA extraction, the second aliquot (80 μl) was suspended in 1X Buffer Gold and processed as described by Ferrari et al. (2007). DNA samples were amplified at the *gdh* locus alone using the optimal PCR chemistry (previously determined). The *gdh* gene was favoured as the higher sequence diversity enables all human strains and all major assemblages to be delineated (Read et al., 2004). The *gdh* gene is also more difficult to amplify due to its single copy nature and higher nucleotide diversity, and as such DNA extraction was targeted at increasing PCR sensitivity for this gene (Caccio and Ryan 2008; Nantavisai et al., 2007). DNA concentration was measured for all isolates as previously described. A two tailed t-test for significance was performed on concentration measurements for secondary PCR products and an optimal molecular protocol for *G. intestinalis* identification was determined.

7.1.3.3 Assessment of clinical performance

Clinical performance of the optimal protocol was evaluated using fifty clinical samples and amplification of the 18S rRNA and *gdh* loci. The numbers of *Giardia* cysts per gram of faeces were calculated for each sample using IMS-flow cytometry (IMS-FC) to determine sensitivity of the protocol. IMS-FC was conducted according to the method described by Power et al. (2003) using G203-FITC pre-stain (8 μl; 50μg/ml), G203-FITC post stain (50 μl; 10 μg/ml) and a positive control containing *G. intestinalis* cysts were

used to identify the FITC detection (FL2), compensation and threshold parametres for *Giardia* cysts. FC sorting was conducted on a FACSCALibur sort, flow cytometre (Becton Dickinson Biosciences, Sydney Australia) and analysed by Cellquest software (BD Biosciences, Sydney Australia). Cysts were collected on Isopore™ membranes (13mm, 0.8µM; Millipore, Sydney, Australia) using a Sort Stage (MRL, Sydney, Australia) connected to the flow cytometre, as previously described (Power et al., 2003). Cysts were resolved under epifluorescent microscopy, 40X magnification (Zeiss Axioskop epifluorescent microscope, Carl Zeiss, Sydney Australia) and counted across 20 microscope fields per membrane.

7.1.4. Results

7.1.4.1 Evaluation of PCR protocols

The performance of six published PCR assays were evaluated by comparing the intensity of PCR products under UV light and the concentration of DNA generated by each assay. Amplification of the 18S rRNA gene was only successful for five of the six protocols tested. Ampli*taq* failed to generate any PCR products whilst Red Hot *Taq* and *Sahara* DNA polymerase generated comparatively lower PCR yields, as visualised under UV light and by comparing DNA concentration measurements (Table 7.1.1). 18S rRNA amplification using *Tth* plus DNA polymerase, following the PCR chemistry previously described (Hopkins et al., 1997; Read et al., 2002) and the GC-RICH PCR System, dNTPack (using 20 pmol of each primer and 0.5 M of GC-RICH Resolution Solution) generated the highest DNA concentrations (Table 7.1.1) and clearer and brighter amplicons (Figure 1). For the GC-RICH PCR System, dNTPack DNA concentration measurements were significantly higher than Red Hot *Taq* and *Sahara* (ANOVA main p-value < 0.00 and all Tukey p-values < 0.01), whilst *Tth* plus generated higher concentrations than Red Hot *Taq*, *Sahara* and *Accutaq* (Tukey p-values < 0.01). There was no statistical difference

between concentration measurements for *Tth* plus and the GC-RICH PCR System, dNTPack. Whilst our results indicate both PCR methods were optimal, the GC-RICH PCR System, dNTPack was chosen to evaluate DNA extraction methods, *gdh* amplification and clinical performance. This method was favoured as the GC-RICH PCR System, dNTPack was more user-friendly, and reduced PCR preparation times.

7.1.4.2 Evaluation of DNA extraction techniques

DNA concentration measurements were used to compare the efficiency of ISOLATE Fecal DNA extraction and PrepGEM DNA extraction, following IMS. The ISOLATE Fecal DNA Kit generated the highest DNA concentration measurements in *gdh* amplicons (Table 7.1.1) and the differences between the two techniques were statistically significant (t-test p-value < 0.00). *Gdh* concentrations measurements were also substantially higher than 18S rRNA measurements however product size and number of PCR cycles would account for this difference. The method was trialled on clinical samples, however, conducting IMS prior to DNA extraction resulted in PCR failures and IMS was removed from the protocol. The optimal molecular protocol was established from direct DNA extraction using the ISOLATE Fecal DNA Kit and PCR amplification using the GC-RICH PCR System, dNTPack.

7.1.4.3 Evaluation of clinical and diagnostic performance

Application of the optimal molecular protocol to 50 clinical samples resulted in successful amplification at both loci. For the 18S rRNA gene, 45 clinical samples were amplified and for the *gdh* gene, 47 samples were amplified. The numbers of cysts in clinical samples ranged from 10 to over 10^6 cysts per gram of faeces (Table 7.1.2). One sample that was PCR negative at both the 18S rRNA and *gdh* loci, was also negative by IMS-FC and microscopy, and for this sample, *G. intestinalis* cysts could not be identified

Table 7.1.1 Comparison of DNA concentrations (ng/μl) generated by different molecular protocols for 18S rRNA and *gdh* amplification of samples spiked with 625 *G. intestinalis* cysts

	18S rRNA PCR[a]					***gdh* PCR[b]**	
Samples	*Tth* plus	Red Hot *Taq*	*Sahara*	*Accutaq*	GC-RICH PCR System	IMS & PrepGEM	IMS & ISOLATE
+ cysts	15.5	8.4	9.8	11.5	12.8		
+ trophozoites	14.5	10.2	10.1	10.8	13.5		
MQG17	14.2	7.3	8.4	13	14.9	22.6	39.3
MQG23	12.7	6.2	7.4	11	12.2	11.8	30.1
MQG24	15.1	6.9	8.4	11.4	13.5	15.6	36.0
MQG27	16.7	6.8	6.5	11.6	13	20.3	40.7
Mean concentration	14.78	7.63	8.43	11.55	13.32	17.58	36.53
SD	1.35	1.45	1.38	0.77	0.92	4.83	4.71

[a] 180 bp fragment, 35 PCR cycles
[b] 432 bp fragment, 55 PCR cycles, amplified using GC-RICH PCR System

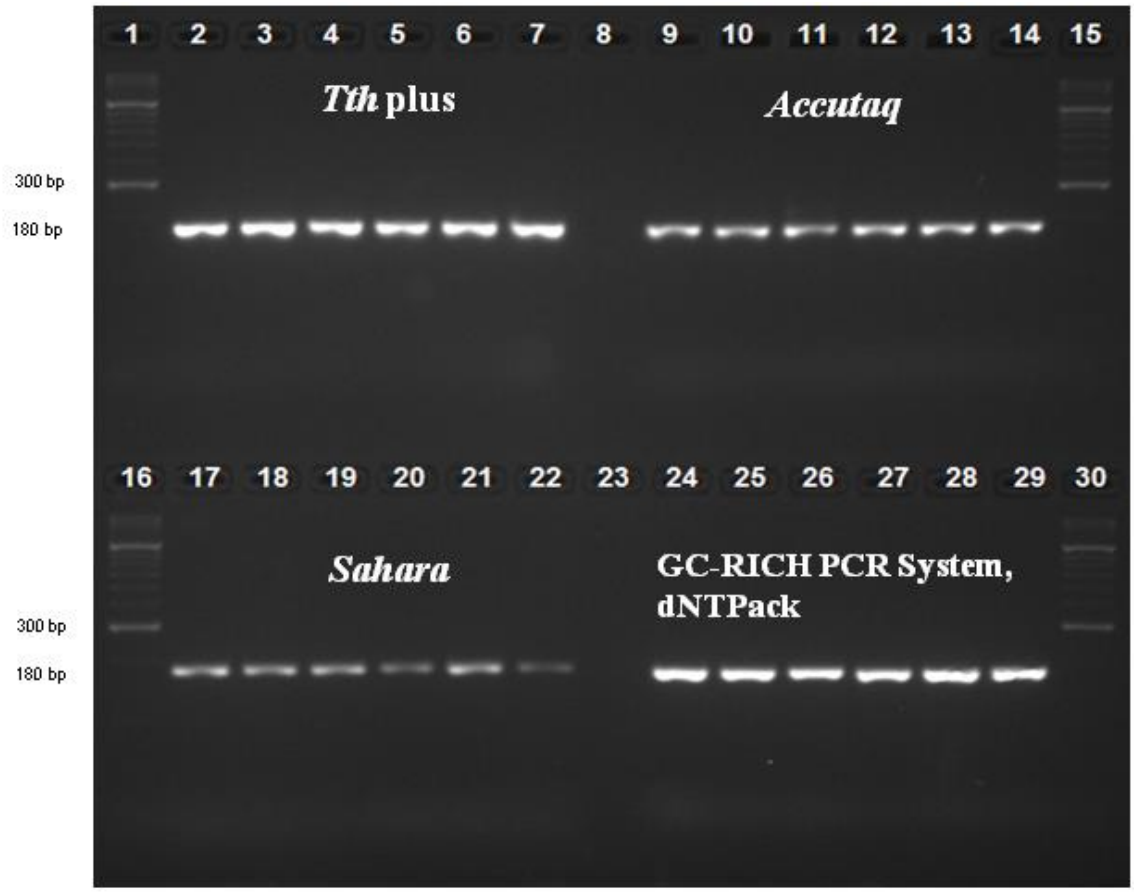

Figure 7.1.1 Comparison of positive DNA controls and DNA polymerases that generated the highest 18S rRNA yields. Lanes 2–7, *Tth* plus DNA polymerase. Lane 2, cyst DNA; lane 3, trophozoite DNA, lanes 4–7, spiked clinical samples MQG17, MQG23, MQG24, MQG27; lane 8, no product added. Lanes 9–14, PCR using *Accutaq* DNA polymerase. Lane 9, cyst DNA; lane 10, trophozoite DNA; lanes 11- 14 spiked samples MQG17, MQG23, MQG24, MQG27; Lanes 17-22, PCR using *Sahara* DNA polymerase. Lane 17, cyst DNA; lane 18, trophozoite DNA; lanes 19-22, spiked samples MQG17, MQG23, MQG24, MQG27; lane 23, no product added. Lanes 24-29, PCR with GC-RICH PCR System, dNTPack. Lane 24, cyst DNA; lane 25, trophozoite DNA; lanes 26-29 spiked samples MQG17, MQG23, MQG24, MQG27.

(Table 7.1.2). Of the six that samples that were PCR negative at one locus, four samples contained *G. intestinalis* cysts, and in two of these cyst loading exceeded 10^4 cysts per gram of faeces (Table 7.1.2). An additional eight samples were negative by IMS-FC.

Table 7.1.2 Comparison of 18S rRNA and *gdh* PCR amplification with the number of cysts isolated from 49 clinical samples.

PCR amplification	> 10^3 cysts/gram faeces	200 – 10^3 cysts/gram faeces	< 200 cysts/gram faeces	IMS-FC negative	**Total**
18S rRNA & *gdh* positive	20	3	13	6	42
18S rRNA positive only			1	1	6
gdh positive only	2		1	1	
Negative both loci				1	1
Total	22	3	15	9	49

7.1.5 Discussion

This study evaluated several published protocols for *G. intestinalis* identification and found that direct DNA extraction using the ISOLATE Fecal DNA Kit followed by amplification of DNA using the GC-RICH PCR System, dNTPack, was optimal for *Giardia* identification. *Tth* plus DNA polymerase was also suitable for PCR amplification. *Tth* plus is commonly used for *Giardia* (Hopkins et al., 1997; Read et al., 2004; Read et al., 2002) however we chose to use the GC-RICH PCR System, dNTPack to optimise a streamlined protocol and improve processing speed. *Tth* plus was difficult to use, reagents were sticky and solutions were difficult to aliquot, and time consuming to use. Although we preferred the GC-RICH PCR System, dNTPack, *Tth* plus is an inexpensive alternative (< 20% of GC-RICH PCR System, dNTPack price on a per unit basis).

Once a sensitive PCR method was determined, our results showed that optimising DNA extraction to increase DNA template was more efficient than purifying parasites to extract DNA and reduce faecal inhibition. For clinical samples, conducting IMS prior to DNA extraction resulted in PCR failures, which may result from several factors including; non-specific binding of the monoclonal antibody; intermittent cyst shedding; and the exclusion of trophozoites through G203 specificity for cyst wall antigens (Ferrari et al., 1999; Power et al., 2003). The IMS recovery rate for *Giardia* was not determined in this

study however our results suggest that trophozoites are likely to be an important source of *Giardia* DNA. Previous studies have reported reduced sensitivity following IMS (Pelayo et al., 2008; Wilke et al., 2009) and higher PCR success rates (>70%) using direct faecal DNA extraction (Wilke et al., 2009). Our analysis of faecal samples by IMS-FC showed that several samples contained low numbers of cysts. Many clinical patients shed trophozoites with the rapid passage of intestinal contents (Nantavisai et al., 2007; Roxstrom-Lindquist et al., 2006) and maximising detection by extracting DNA from both cysts and trophozoites is beneficial.

The optimal molecular protocol for 18S rRNA and *gdh* amplification was a combination of previously described methods (Read et al., 2004; Sulaiman et al., 2003) that demonstrated high diagnostic performance when applied to fifty positive clinical samples; generating PCR success rates of 90% and 94% respectively. DNA extraction was simple and rapid, and PCR amplification was successful in the presence of faecal inhibitors, eliminating the need for cyst concentration. DNA was amplified from as little as 10 cysts per gram of faeces and the high sensitivity of the protocol minimised the need to repeat assays to account for intermittent shedding. Identification of high cyst loads in 18S rRNA negative samples demonstrated that even when cyst shedding is high, other factors such as DNA polymerase inhibition and the quality of intact DNA can prevent PCR amplification. A multilocus approach is required for validation of *Giardia* genotyping and application of this protocol to two loci removes additional PCR optimisation steps that are normally conducted.

7.1.6 Conclusions

This study has determined an optimal molecular method for *G. intestinalis* identification in clinical samples. We view the assay as a tool for epidemiological investigations, which is of clinical and public health importance. Using PCR detection for

routine diagnosis of giardiasis would be favourable and of significant value to understanding the biology of this important parasite. However, due to the expense and expertise required to perform such analyses, application of PCR in clinical diagnostic laboratories is not yet routine (Savioli et al., 2006).

The use of the protocol described here for molecular diagnostics and epidemiological investigations should provide a streamlined, efficient and sensitive method for examining clinical samples. It should offer a specific framework for *G. intestinalis* identification that is reproducible between laboratories. Application of this protocol should allow larger sample screenings and identification of *G. intestinalis* strains, which is essential for understanding transmission cycles, infection sources and disease risks, and developing targeted disease management and surveillance relative to the different human infective *Giardia* strains.

7.1.7 Acknowledgments

This research was conducted with financial assistance from Macquarie University. We would like to thank Dr. Jacqui Upcroft of the Queensland Institute of Medical Research for kindly providing trophozoites and also Douglass Hanly Moir Pathology, Symbion Health Laverty Pathology, Hunter Area Pathology Service, Westmead Hospital & John Hunter Hospital for their invaluable help in providing faecal samples.

7.1.8 References

Caccio, S.M., Ryan, U., 2008. Molecular epidemiology of giardiasis. Mol Biochem Parasitol 160, 75-80.

Caccio, S.M., Thompson, R.C.A., McLauchlin, J., Smith, H.V., 2005. Unravelling *Cryptosporidium* and *Giardia* epidemiology. Trends Parasitol 21, 430-437.

Caccio, S.M., De Giacomo, M., Aulicino, F.A., Pozio, E., 2003. *Giardia* Cysts in Wastewater Treatment Plants in Italy. Appl Environ Microb 69, 3393-3398.

Ferrari, B.C., Power, M.L., Bergquist, P.L., 2007. Closed-tube DNA extraction using a thermostable proteinase is highly sensitive, capable of single parasite detection. Biotechnol Lett 29, 1831-1837.

Ferrari, B.C., Vesey, G., Weir, C., Williams, K.L., Veal, D.A., 1999. Comparison of *Cryptosporidium*-specific and *Giardia*-specific monoclonal antibodies for monitoring water samples. Wat Res 33, 1611-1617.

Hopkins, R.M., Meloni, B.P., Groth, D.M., Wetherall, J.D., Reynoldson, J.A., Thompson, R.C.A., 1997. Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. J Parasitol 83, 44-51.

Monis, P.T., Andrews, R.H., Mayrhofer, G., Ey, P.L., 1999. Molecular Systematics of the Parasitic Protozoan *Giardia intestinalis.* Mol Biol Evol 16, 1135-1144.

Nantavisai, K., Mungthin, M., Tan-ariya, P., Rangsin, R., Naaglor, T., Leelayoova, S., 2007. Evaluation of the Sensitivities of DNA Extraction and PCR Methods for Detection of *Giardia duodenalis* in Stool Specimens. J Clin Microbiol 45, 581-583.

Pelayo, L., Nunez, F.A., Rojas, L., Furuseth Hansen, E., Gjerde, B., Wilke, H., Mulder, B., Robertson, L., 2008. *Giardia* infections in Cuban children: the genotypes circulating in a rural population. Ann Trop Med Parasitol 102, 585-595.

Power, M.L., Shanker, S.R., Sangster, N.C., Veal, D.A., 2003. Evaluation of a combined immunomagnetic separation/flow cytometry technique for epidemiological investigations of *Cryptosporidium* in domestic and Australian native animals. Vet Parasitol 112, 21-31.

Read, C.M., Monis, P.T., Thompson, R.C.A., 2004. Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP. Infect Genet Evol 4, 125-130.

Read, C., Walters, J., Robertson, I.D., Thompson, R.C.A., 2002. Correlation between genotype of *Giardia duodenalis* and diarrhoea. Int J Parasitol 32, 229-231.

Roxstrom-Lindquist, K., Palm, D., Reiner, D., Ringqvist, E., Svard, S.G., 2006. *Giardia* immunity – an update. Trends Parasitol 22, 26-31.

Savioli, L., Smith, H., Thompson, A., 2006. *Giardia* and *Cryptosporidium* join the 'Neglected Diseases Initiative.' Trends Parasitol 22, 203-208.

Smith, H.V., Caccio, S.M., Tait, A., McLauchlin, J., Thompson, R.C., 2006. Tools for investigating the environmental transmission of *Cryptosporidium* and *Giardia* infections in humans. Trends Parasitol 22, 160-167.

Sulaiman, I.M., Fayer, R., Bern, C., Gilman, R.H., Trout, J.M., Schantz, P.M., Das, P., Lal, A.A., Xiao, L., 2003. Triosephosphate Isomerase Gene Characterization and Potential Zoonotic Transmission of *Giardia duodenalis.* Emerg Infect Dis 9, 1444-1452.

Traub, R.J., Monis, P.T., Robertson, I.D., 2005. Molecular epidemiology: A multidisciplinary approach to understanding parasitic zoonoses. Int J Parasitol 35, 1295-1307.

van der Giessen, J.W.B., de Vries, A., Roos, M., Weilinga, P., Kortbeek, L.M., Mank, T.G., 2006. Genotyping of *Giardia* in Dutch patients and animals: A phylogenetic analysis of human and animal isolates. Int J Parasitol 36, 849-858.

Waldron, L.S., Ferrari, B.C., Gillings, M.R., Power, M.L., 2009. Terminal Restriction Fragment Length Polymorphism for Identification of *Cryptosporidium* Species in Human Feces. Appl Environ Microb 75, 108-112.

Wilke, H., Robertson, L.J., 2009. Preservation of *Giardia* cysts in stool samples for subsequent PCR analysis. J Microbiol Meth 78, 292-296.

7.2 *Giardia duodenalis* and *Cryptosporidium* occurrence in Australian sea lions (*Neophoca cincerea*) exposed to varied levels of human interaction.

T.C. Delport [a], A.J.Asher [a], L.J.Beaumont [a], K.N. Webster [a], R.G.Harcourt [a], M.L. Power [a]

[a] *Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia*

Delport, T.C., Asher, A.J., Beaumont, L.J., Webster, K.N., Harcourt, R.G., Power, M.L., 2014. *Giardia duodenalis* and *Cryptosporidium* occurrence in Australian sea lions (*Neophoca cinerea*) exposed to varied levels of human interaction. Int J Parasitol Parasites Wildl 3, 269-275.

7.2.1 Abstract

Giardia and *Cryptosporidium* are amongst the most common protozoan parasites identified as causing enteric disease in pinnipeds. A number of *Giardia* assemblages and *Cryptosporidium* species and genotypes are common in humans and terrestrial mammals and have also been identified in marine mammals. To investigate the occurrence of these parasites in an endangered marine mammal, the Australian sea lion (*Neophoca cinerea*), genomic DNA was extracted from faecal samples collected from wild populations (n= 271) in Southern and Western Australia and three Australian captive populations (n= 19). These were screened using PCR targeting the 18S rRNA of *Giardia* and *Cryptosporidium*. *Giardia duodenalis* was detected in 28 wild sea lions and in seven captive individuals. Successful sequencing of the 18S rRNA gene assigned 27 *Giardia* isolates to assemblage B and one to assemblage A, both assemblages commonly found in humans. Subsequent screening at the *gdh* and β-giardin loci resulted in amplification of only one of the 35 18S rRNA positive samples at the β-giardin locus. Sequencing at the β-giardin locus assigned the assemblage B 18S rRNA confirmed isolate to assemblage AI. The geographic distribution of sea lion populations sampled in relation to human settlements indicated that *Giardia* presence in sea lions was highest in populations less than 25 km from humans. *Cryptosporidium* was not detected by PCR screening in either wild colonies or captive sea lion populations. These data suggest that the presence of *G. duodenalis* in the endangered Australian sea lion is likely the result of dispersal from human sources. Multilocus molecular analyses are essential for the determination of *G. duodenalis* assemblages and subsequent inferences on transmission routes to endangered marine mammal populations.

7.2.2 Introduction

Protozoan parasites are a primary cause of morbidity and mortality in terrestrial and marine mammal populations (Hughes-Hanks et al., 2005; Ryan et al., 2014). Two genera, *Cryptosporidium* and *Giardia*, are amongst the most common organisms identified as causing enteric disease in pinniped species (eg. Appelbee et al., 2010; Bass et al., 2012; Dixon et al., 2008; Olson et al., 1997). Currently, 26 *Cryptosporidium* species are considered valid and over 40 genotypes or cryptic species have been identified (c.f. Ryan et al., 2014). Several *Cryptosporidium* species present in humans and terrestrial mammals have been identified in marine mammals, including *C. parvum, C. muris* and *C. hominis* (eg. Appelbee, 2005; Deng et al., 2000; Santin et al., 2005). In addition, *Cryptosporidium* seal genotypes 1 and 2, thought to be specific to marine mammals, have been identified in the ringed seal (*Phoca hispida*), and seal genotype 2 in the harbor seal (*Phoca vitulina*) and grey seal (*Halichoerus grypus*) (Bogomolni et al., 2008; Dixon et al., 2008; Santin et al., 2005). More recently, novel genotypes of *Cryptosporidium* have been described in the southern elephant seal (*Mirounga leonina*), harp seal (*Pagophilus groenlandicus*) and Weddell seal (*Leptonychotes weddellii*) (Bass et al., 2012; Rengifo-Herrera et al., 2011; Rengifo-Herrera et al., 2013).

Of the six *Giardia* species, *Giardia duodenalis* has the broadest host range (c.f. Feng and Xiao, 2011). Molecular characterization of *G. duodenalis* has revealed significant genetic diversity, an indicator of complexity within this species. Accordingly, *G. duodenalis* is divided into assemblages A to H (c.f. Feng and Xiao, 2011), with assemblages A and B being the most diverse with at least four sub-assemblage types (I-IV) (c.f. Monis and Thompson, 2003). Assemblages A and B are human infective and have also been documented in wildlife and domestic animals (c.f. Feng and Xiao, 2011). Assemblages C and D are found in dogs, assemblages E, F and G in domestic ruminants, cats and rodents respectively, while assemblage H has been described in seals (c.f. Cacciò

et al., 2005; Lasek-Nesselquist et al., 2010).

Infections of *G. duodenalis* are reported with a greater frequency in marine mammals than *Cryptosporidium*. *Giardia* assemblages A and B are the most commonly identified assemblages in pinniped species (eg. Appelbee et al., 2005; Lasek-Nesselquist et al., 2008). Assemblages C and D have been described in harbor seals (*Phoca vitulina*) and the novel seal genotype H identified in grey seals (Gaydos and Miller, 2008; Lasek-Nesselquist et al., 2010). The presence of human host specific *Giardia* assemblages in marine mammals may be an indication of human impacts on the marine environment, which poses potential concerns for conservation of endangered pinniped species.

The Australian sea lion (*Neophoca cinerea*) is one of the rarest seals in the world, with a total population of less than 15,000 (Shaughnessy et al., 2011). Colonies of this endangered pinniped are distributed on coastal islands and the mainland of Western and South Australia, many within close proximity to human settlements (Goldsworthy and Gales, 2008; Goldsworthy et al., 2007). As a tourist icon, some South Australian colonies, in particular Seal Bay on Kangaroo Island, experience frequent human visitation and habitat disturbance. Interactions with, and proximity to, humans and wastewater run-off increases the likelihood of transmission of *Giardia* and *Cryptosporidium* from humans and domestic animals to seal populations. The aim of this study was to detect and characterize *Giardia* and *Cryptosporidium* in wild and captive Australian sea lions within a phylogenetic framework, and to determine if proximity to human settlements was related to parasite detection. We hypothesized that if transmitted through human influences, protozoal strains would be more likely to be detected in seal colonies in close proximity to human settlements and would be similar to those found in domestic animals and human populations.

7.2.3 Materials and Methods

7.2.3.1 Sample collection

Fresh wild faecal samples (n= 271) were collected over a range of seasons during a 2 year period from coastal and island colonies in Western Australia (Figure 7A) and South Australia (Figure 7B and 7C). For captive animals, freshly passed faecal samples (*n* = 19) were collected from animal housing over a period of 4 months from the resident populations held at Dolphin Marine Magic and Taronga Zoo, New South Wales, and Sea World, Queensland. Faecal samples were transported to the laboratory and stored at 4° C until processing for genomic DNA extraction.

7.2.3.2 DNA extraction

Genomic DNA was extracted from sea lion faecal samples (*n* = 290) using the ISOLATE Fecal DNA Kit (Bioline, Sydney, Australia). Faecal samples (approximately ~150 mg) were aliquoted into lysis bead tubes and DNA extraction performed as per the manufacturers protocol. Eluted DNA was stored at -20° C until further analysis.

7.2.3.3 PCR screening for characterization of *Giardia duodenalis*

The presence of *Giardia* isolates was determined using the protocol targeting the 18S rRNA gene described in Hopkins et al. (1997) and Read et al. (2004).

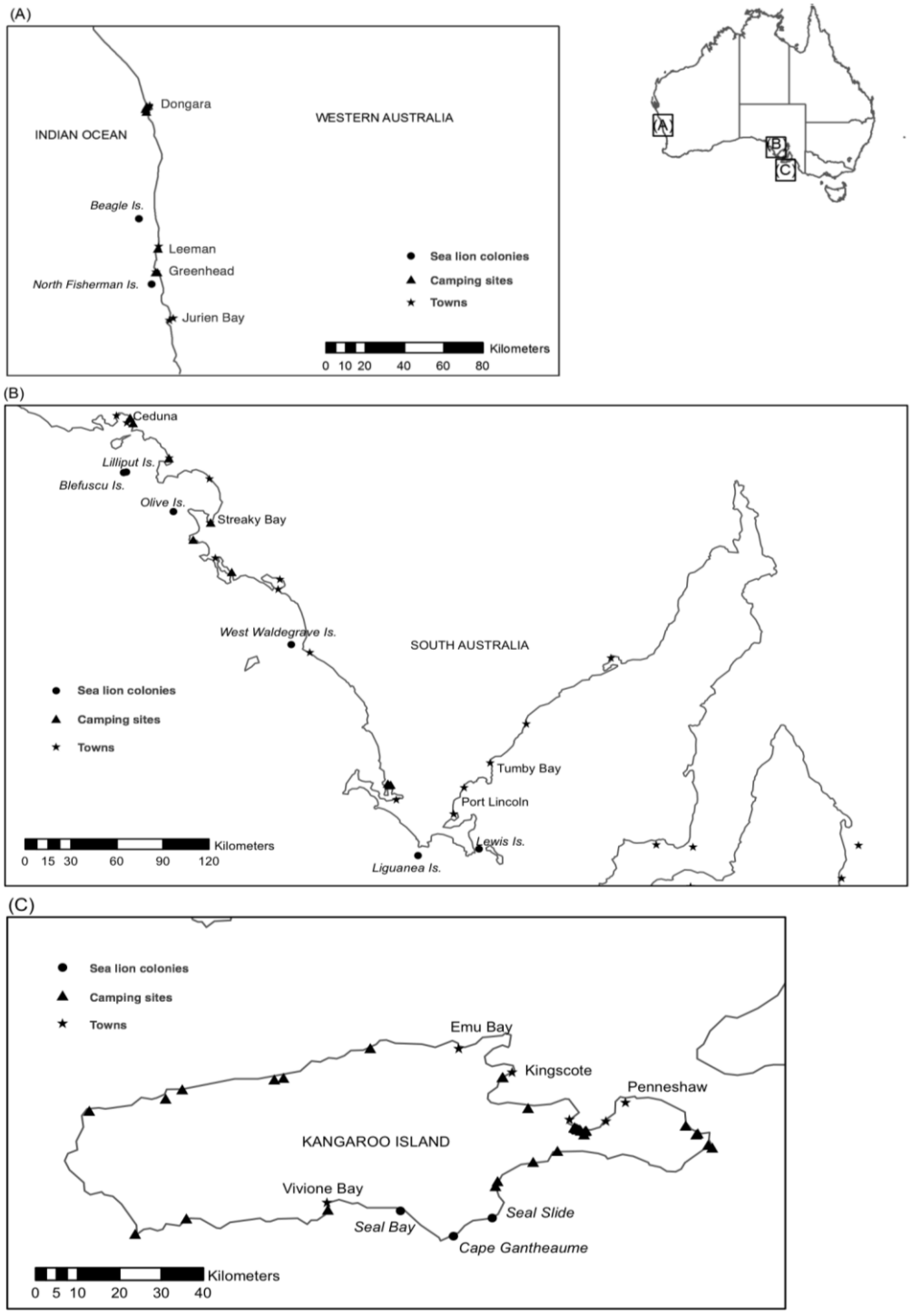

Figure 7.2.1 (A) Western Australia sampling locations. Faecal samples were collected from Beagle and North Fisherman Islands. Coastal settlements and human impacted camping locations within close proximity to sea lion colonies are indicated. **(B)** South Australia sampling locations. Faecal samples were collected from South Australia colonies; Blefuscu, Lewis, Liguanea, Lilliput, Olive and West Waldegrave Islands. **(C)** South Australia sampling locations: Kangaroo Island. Three colonies were sampled from Kangaroo Island including Cape Gantheaume, Seal Bay and Seal Slide.

Nested PCR using the primers RH11/RH4LM in the primary reaction and GiAR18SeR/GiAR18SiR in the secondary reaction were used to amplify a ~175 bp fragment of the 18S rRNA gene. Primary and secondary reactions (25 μl) were prepared using the GC-RICH PCR system (Roche Diagnostics, Indianapolis, IN) (Asher et al., 2012). Thermocycling for both the primary and secondary reactions were performed using the conditions described by Hopkins et al. (1997).

To characterize *Giardia* assemblages, 18S rRNA positive isolates (n= 35) were screened at the *gdh* and β-giardin loci. For *gdh* amplification the previously described semi-nested protocol of Read et al. (2004) was used. Primary and secondary reactions (25 μL) were prepared using the GC-RICH PCR system (Roche Diagnostics) and the primer set GdheF/GdhiR in the primary reaction and primers GdhiF/GdhiR in the secondary. Cycling was performed at denaturation for 2 min at 94° C, 1 min at 56° C, 2 min at 72° C; 35 cycles at 94° C for 30s, 56° C for 20s and 72° C for 45s; and a final extension step at 72° C for 7 min (Read et al., 2004). Where samples failed to produce an amplicon (n= 35), template DNA was increased from 1 to 2.5 μl in the primary reaction and secondary reactions prepared using identical PCR chemistry, with 2.5 μl of primary PCR product. Amplification of the β-giardin locus was achieved following the nested protocol described by Cacciò et al. (2002) and Lalle et al. (2005). Primary reactions (25 μl) were prepared using 1.5 mM MgCl2, 200 μM dNTPs, 200 nM of each primer G7/G759, 1U of DNA polymerase Tth Plus (Fisher Biotec, Wembley, Australia) and 2 μl template DNA and the cycling conditions used by Cacciò et al. (2002). Secondary reactions (25 μl) were prepared using PCR chemistry identical to the primary reactions, with 2 μl of primary PCR product and following the thermocycling conditions described by Lalle et al. (2005). PCR was performed with the internal primers described by Lalle et al. (2005) with a slight modification to the internal forward primer (GAA CGA GAT CGA GGT CCG) after β-giardin sequence comparisons available on the NCBI GenBank database

(http://www.ncbi.nlm.gov/genbank/index.html) showed a 4 bp difference between *Giardia* sequences and the internal forward primer sequence.

A spike analysis using DNA extracted from an existing *Giardia* laboratory trophozoite isolate was performed on 18S rRNA positive samples that failed to amplify at the *gdh* and β-giardin loci.

7.2.3.4 PCR screening for characterization of *Cryptosporidium* sp.

Screening for *Cryptosporidium* was conducted using a nested PCR protocol targeting the small subunit 18S rRNA (Xiao, 1999). RedHot Taq (Thermo Scientific, Scoresby, Australia) was used for reactions and all conditions were as described by Xiao et al. (2000). To confirm that the absence of *Cryptosporidium* was not the result of faecal inhibitors impairing DNA amplification, all samples (n = 290) were spiked with *Cryptosporidium parvum* DNA from an existing laboratory isolate (Waldron et al., 2011a) and PCR screening repeated as described above.

All PCRs were performed in an Eppendorf Mastercycler (Eppendorf, North Ryde, Australia). PCR products (8 μl) were resolved by agarose gel electrophoresis (2% w/v, 110 V for 30 min) in TBE (Tris, boric acid, EDTA pH8) with 2 μl SYBR safe (Invitrogen, Mulgrave, Australia) using a HyperLadder II DNA marker (Bioline) to estimate amplicon size.

7.2.3.5 DNA sequencing and phylogenetic analyses

To identify *Giardia* assemblages the 18S rRNA products (n = 35) from GiAR18SeR/GiaR18SiR secondary reaction were purified for sequencing using the QIAquick PCR Purification Kit (Qiagen, Melbourne, Australia) and sequenced in the forward direction using the internal primer GiAR18SeR and in the reverse direction using GiAR18SiR. To identify *Giardia* sub-assemblage the β-giardin product (n = 1) was

purified and sequenced in the forward and reverse direction using the secondary β-giardin PCR primers. All sequencing was performed by Macrogen Inc. (Seoul, Korea) on a 3130x1 genetic analyser (Applied Biosystems, Foster City, California) using the standard run protocol for a 50 cm, 16 capillary array using a Big Dye terminator kit (Applied Biosystems).

Forward and reverse sequences (18S rRNA and β-giardin) were checked manually and trimmed in GeneiousPRO version 5.0.3 (Biomatters Ltd, Auckland, New Zealand) and a single contiguous sequence (contig) was assembled for each sample. BlastN sequence searches were performed to assign contiguous 18S rRNA and β-giardin sequences to an assemblage. To allow for assemblage identification sequences were analyzed within a phylogenetic framework. Representative 18S rRNA sequences for *Giardia* assemblages A to G were obtained from the NCBI GenBank database using accession numbers AF199446, AF199447, AF199449, AF199443, AF199448, AF199444 and AF199450 for 18S rRNA analyses (Wielinga and Thompson, 2007). 18S rRNA and β-giardin sequences representing assemblage H were not available on GenBank and could not be included in analyses. Representative β-giardin sequences for *Giardia* assemblages AI-III to G were obtained from the NCBI GenBank database using accession numbers X85958, AY072724, FJ971410, AY072727, AY545646, AY545647, DQ116608, AY647264 and EU769221 (Kosuwin et al., 2010; Lalle et al., 2005; Lebbad et al., 2010; Wielinga and Thompson, 2007). Contiguous 18S rRNA and β-giardin sequences generated in this study were aligned to GenBank sequences using ClustalW (Thompson et al., 1994) in MEGA version 6.0 (Tamura et al., 2013). For phylogenetic analyses, nucleotide substitution models were tested for maximum likelihood in MEGA6 (Tamura et al., 2013). Akaike Information Criterion corrected (AICc) values were used to determine the optimal parameters. Phylogenetic trees were constructed for 18S rRNA and β-giardin sequences using maximum likelihood (Tamura 3-parameter distance model with the uniform distribution

parameter) and bootstrap analysis (1000 replicates) and compared to existing assemblages (Kosuwin et al., 2010; Lalle et al., 2005; Lebbad et al., 2010; Wielinga and Thompson, 2007). 18S rRNA sequences generated in this study have been submitted to The European Nucleotide Archive (ENA) under the accession numbers LN610171-LN610198. The β-giardin sequence generated in this study has been submitted to GenBank under accession number KM497498.

7.2.3.6 Mapping and statistical analyses

Maps illustrating the locations of wild sea lion populations sampled and proximity of towns and camping grounds were developed using ArcGIS version 10.0 (ESRI, 2010). A Pearson's χ^2 test was used to identify differences in the occurrence of *Giardia duodenalis* between wild and captive populations. For the wild populations only, a generalized linear model (GLM) with a binomial probability distribution was used to examine the effect of sea lion colony distance from human settlements and sampling season on presence/absence of *Giardia*. For this analysis, colonies were grouped into three distance-from-settlement categories: < 25 km, 26-69 km and > 70 km (Table 7.2.1). Differences in occurrence between distance categories were determined using a Tukey's post-hoc test.

7.2.4 Results

7.2.4.1 *Giardia* detection and species identification

Screening of genomic DNA using a *Giardia* specific 18S rRNA protocol resulted in the detection of *Giardia* in 28 samples from wild sea lions (10.3%) and in seven samples from captive sea lions (36.8%). There was a significant difference in *Giardia* presence between wild and captive individuals ($\chi^2 = 11.758$, df = 1, p= <0.001). In wild colonies, the distance from human settlement had a significant effect on the presence or absence of

Giardia (Wald χ^2 = 39.078, df = 2, p= <0.001). Colonies less than 25 km from human settlements had a higher occurrence of *Giardia* than colonies more than 26 km away (Table 7.2.1). There was no effect of sampling season on *Giardia* presence (Wald χ^2 = 6.112, df = 3, p= 0.106).

DNA sequences were obtained for 28 of the 35 18S rRNA positive samples. BlastN search identified 27 sequences as belonging to *Giardia duodenalis* assemblage B and one belonging to assemblage A. Analysis using the phylogenetic framework clustered all samples from wild sea lions (n= 24) and three samples from captive animals within a clade that also contained the assemblage B reference sequence from GenBank (Figure 7.2.2). One captive sample clustered within a clade containing the reference sequences from GenBank for assemblages A, E and F.

Table 7.2.1. Australian sea lion colony groupings and analysis of *G. duodenalis* presence. Wild sea lion colony distance-from-settlement categories. Differences in occurrence between distance categories were determined using a Tukey's post-hoc test.

Distance category (km)	Colonies in category	Mean occurrence of *G. duodenalis* (%)	Total number of samples
< 25	Beagle Island Cape Gantheaume North Fisherman Island Seal Bay Seal Slide	23.8[a]	80
26-69	Blefuscu Island Lewis Island Liguanea Island Lilliput Island	5.8[b]	120
>70	Olive Island West Waldegrave Island	2.8[b]	71

[a, b] Significant difference in *Giardia duodenalis* presence between groupings.

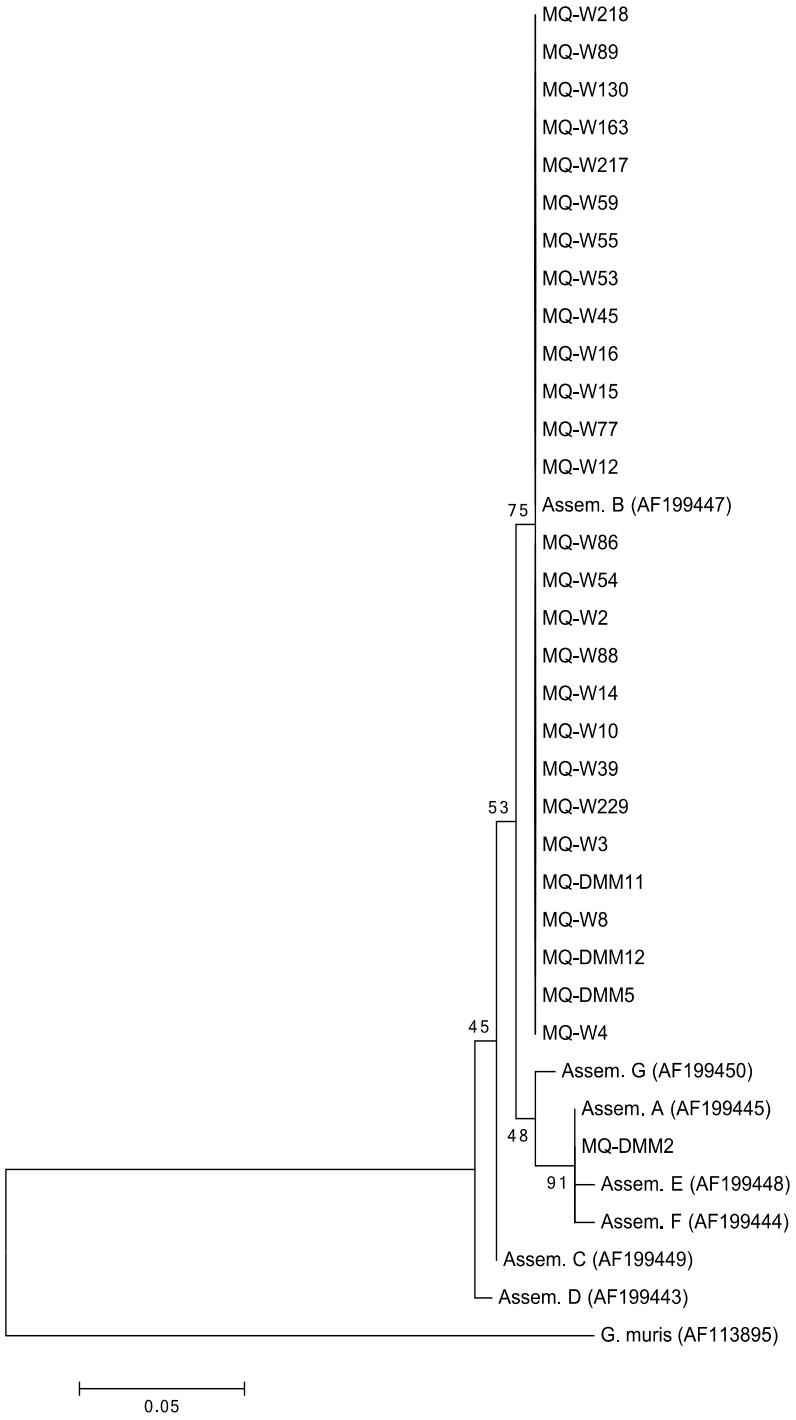

Figure 7.2.2 *Giardia duodenalis* 18S rRNA phylogenetic tree.
Phylogenetic analysis of *Giardia duodenalis* positive samples was performed using a fragment of the 18S rRNA gene. Analysis within the phylogenetic framework placed sea lion samples within the assemblage B (n= 27) and assemblage A clades (n= 1). Branch values indicate percent bootstrapping using 1,000 replicates.

Analysis of clustalW alignment showed a 2 bp polymorphism between the assemblages with the captive sample most closely aligned to assemblage A. Alignment of the sample to representative sequences showed that the sample was 100% identical to assemblage A but not E or F.

18S rRNA positive samples (n= 35) failed to amplify at the *gdh* locus. Representative samples all produced a *gdh* amplicon when spiked with *Giardia* isolate DNA. The β-giardin locus amplified in one of 35 18S rRNA positive samples identified as assemblage B. A DNA sequence was obtained for the β-giardin positive sample and a BlastN search identified the sequence as belonging to *Giardia duodenalis* assemblage AI. The inferred phylogeny placed the sample within a clade that also contained the assemblage AI reference sequence from GenBank. All samples spiked with *Giardia* positive DNA produced an amplicon when screened using β-giardin PCR.

7.2.4.2 *Cryptosporidium* screening

Cryptosporidium was not detected in any of the faecal samples (n= 290). The purified genomic DNA were deemed PCR competent using DNA spike analysis with all 290 samples generating an amplicon when screened using 18S rRNA PCR.

7.2.5 Discussion

In this study we examined the occurrence of *Giardia* and *Cryptosporidium* in the endangered Australian sea lion. *Giardia duodenalis* assemblage B, commonly found in humans and terrestrial mammals, was detected in wild and captive sea lion populations and *G. duodenalis* assemblage A was detected in a captive animal. Screening for *Cryptosporidium* failed to identify this parasite in any of the samples. Infections of *Cryptosporidium* are commonly reported at lower frequencies than *Giardia* in marine mammal populations (eg. Appelbee et al., 2005; Gaydos and Miller, 2008; Hueffer et al.,

2011; Lasek-Nesselquist et al., 2008). Of the eight *Giardia duodenalis* assemblages A and B are the most commonly identified in wild seal and sea lion populations (eg. Appelbee et al., 2010; Lasek-Nesselquist et al., 2010). Our findings indicate that assemblage B is the most common assemblage detected in wild and captive sea lions while assemblage A occurs at low frequency. The host range of assemblages A and B are broad including domestic animals, livestock and humans (c.f. Monis et al., 1999; Monis and Thompson, 2003). Infections with assemblages A and B are very common in human cases but based on the absence of subtype characterization, it is difficult to assess the association between parasite transmission and humans.

The presence of *G. duodenalis* assemblages AI and B in this sea lion species is a strong indicator of the spread of parasites from terrestrial mammals to the marine environment. *Giardia duodenalis* is reported in higher frequencies in seal and sea lion species distributed within close proximity to human settlements and wastewater run-off localities (Appelbee et al., 2010; Gaydos and Miller, 2008). The presence of *G. duodenalis* in seals visiting haul-out sites distributed near coastal settlements can be up to 5 times greater than individuals at more sparsely populated sites with limited human exposure (Dixon et al., 2008; Hughes-Hanks et al., 2005; Lasek-Nesselquist et al., 2010). Some Australian sea lion colonies are within close proximity to coastal settlements and experience high levels of human interaction as a major tourist icon (Gales et al., 1994; Rodger et al., 2011). Sea lion behaviour such as hauling-out on human impacted beaches increases the potential for exposure to parasites from terrestrial sources. Compared with the more isolated Australian sea lion colonies (> 70 km from human settlement), *Giardia* presence is significantly higher in colonies nearer (< 25 km) to human coastal settlements and those colonies that experience high human visitation. Australian sea lions have limited dispersal and a high degree of philopatry so impacts are likely to be localized (Lowther et al., 2012). Future observation of protozoan prevalence in South Australian (Seal Bay and

Seal Slide on Kangaroo Island) and Western Australian colonies (Shoalwater Marine Park, Perth; North Fisherman Island, Jurien Bay and Recherche Archipelago, Esperance) is therefore essential for monitoring the spread of parasites and associated potential disease risks, and will assist in the development of conservation management strategies.

Giardia duodenalis presence is significantly greater in captive Australian sea lions (36.8%) than wild animals (10.3%) indicating that occurrence may be the result of atypical habitat interactions. Exposure to humans and interactions atypical to those within the natural habitat of sea lions may increase the risk of *Giardia* transmission within captive environments (Beck et al., 2011). Captive mammals may be exposed to *Giardia* through human contact during hand feeding and touching by zoo visitors (Thompson et al., 2008). The captive facilities observed in this study have varying levels of visitor interaction programs with sea lions, some even include activities such as swimming with and touching the animals. While *Giardia* presence in captive marine mammal populations is rarely observed or indeed investigated, screening of *Giardia* in other captive mammal species with similar levels of human interactions would provide an indication of the extent of transmission in the captive environment. This in turn may reveal potential avenues of dispersal of *Giardia* in the captive environment, and by deduction, potential mitigation strategies for improved husbandry.

While the use of molecular tools has facilitated a greater understanding of protozoan origins and host specificity, we had limited success in accurately confirming *Giardia* species sub-assemblage across multiple loci. This poses significant biological implications for inferring host specificity and transmission of *Giardia*. We failed to amplify 18S rRNA positive isolates at the *gdh* and β-giardin loci. Difficulty in confirming positive 18S rRNA detection at the *gdh* locus has been observed in other marine and captive mammal studies. Failure to amplify at the *gdh* locus in samples from Pacific harbor seals and captive mammals was attributed to variation in sequences and failure of primers to anneal (Beck et

al., 2011; Lasek-Nesselquist et al., 2010). While analyses at the 18S rRNA locus alone can enable assemblage identification, multilocus gene screening is required to determine *G. duodenalis* sub-assemblage and specific host origin.

Further, we were unable to consistently assign *G. duodenalis* assemblage across multiple loci for the one sample that amplified at β-giardin. Inconsistent assemblage identification across multiple loci has been observed in other marine mammal studies (Lasek-Nesselquist et al., 2008; Lasek-Nesselquist et al., 2010). Failure to confirm genotype across multiple loci in samples from grey and Pacific harbor seals was attributed to target gene amplification biases, where assemblages A and B preferentially amplified at different loci, and the presence of mixed assemblage infection (Lasek-Nesselquist et al., 2010). Mixed infection of *G. duodenalis* assemblages A and B are commonly reported in human and marine mammal studies, although there is much debate about whether this is the result of infection by multiple isolates or the haplotype of a single isolate (Cacciò and Ryan 2008; Cacciò et al., 2008; Lasek-Nesselquist et al., 2008; Lasek-Nesselquist et al., 2010). While the occurrence of recombination between *G. duodenalis* assemblages A and B has been supported by several studies, the mechanisms involved remain unclear (Cooper et al., 2007; Cacciò and Sprong, 2010; Teodorvic et al., 2007). Consequently, due to limited amplification across multiple loci, we were unable to draw inferences on the potential for target gene amplification biases or the presence of mixed infection. These findings emphasize the need for multilocus molecular characterization to definitively assign *G. duodenalis* assemblages present in wild sea lion populations and determine the origin of parasite dispersal into the marine environment.

Increasing exposure to agricultural run-off and untreated wastewater represents new challenges for managing the dispersal of protozoan parasites into the marine ecosystem. The high occurrence of *Giardia* and similarity to *Giardia* species found in humans in both wild and captive sea lions warrants the need for further molecular investigation to identify

the dispersal routes of parasites from terrestrial ecosystems into marine vertebrate populations.

7.2.6 Acknowledgements

This research was conducted under permission of the Department for Environment and Heritage, South Australia (permit Z25675), DEH SA Wildlife Ethics Committee (44/2008) and the Department of Environment and Conservation, Western Australia (DEC WA License SF007255 and SF008193). The research was supported by grants to Robert Harcourt from the Australian Marine Mammal Centre, Department of Environment, Heritage South Australia and West Australia, Water and the Arts through the Commonwealth Environment Research Fund and funds from Macquarie University (0809/27, 09/38, 0809/12). Funding was also received from Macquarie University Department of Biological Sciences postgraduate project support to Tiffany Delport and Michelle Power. We would like to thank Dolphin Marine Magic, Sea World Gold Coast and Taronga Zoo for kindly providing Australian sea lion faecal samples and also Ben Pitcher, Andy Lowther, Isabelle Charrier and Heidi Ahonen for wild Australian sea lion faecal collection. We would also like to thank Prof. John Ellis and Dr. Sonja Frolich from the University of Technology, Sydney, for kindly providing *Giardia* trophozoite cultures used in this study.

7.2.7 References

Appelbee, A.J., Thompson, R.C.A., Olson, M.E., 2005. *Giardia* and *Cryptosporidium* in mammalian wildlife – current status and future needs. Trends Parasitol 21, 370-376.

Appelbee, A.J., Thompson, R.C.A., Measures, L.M., Olson, M.W., 2010. *Giardia* and *Cryptosporidium* in harp and hooded seals from the Gulf of St Lawrence, Canada. Vet Parasitol 173, 19-23.

Asher, A.J., Waldron, L.S., Power, M.L., 2012. Evaluation of a PCR protocol for sensitive detection of *Giardia intestinalis* in human faeces. Parasitol Res 110, 853-858.

Bass, A.L., Wallace, C.C., Ynd, P.O., Ford, T.E., 2012. Detection of *Cryptosporidium* sp. in two new seal species, *Phoca vitulina* and *Cystophora cristata*, and a novel *Cryptosporidium* genotype in a third seal speces, *Pagophilus groenlandicus*, from the Gulf of Maine. J Parasitol 98, 316-322.

Beck, R., Sprong, H., Bata, I., Lucinger, S., Pozio, E., Caccio, S.M., 2011. Prevalence and molecular typing of *Giardia spp.* in captive mammals at the zoo of Zagreb, Croatia.Vet. Parasitol. 175, 40–46.

Bogomolni, A.L., Gast, R.J., Ellis, J.C., Dennett, M.R., Pugliares, K.R., Lentell, B.J., et al., 2008. Victims or vectors: a survey of marine vertebrate zoonoses from coastal waters of the Northwest Atlantic Dis Aquat Organ 81, 13–38.

Caccio, S.M., Ryan, U., 2008. Molecular epidemiology of giardiasis. Mol. Biochem Parasitol 160, 75–80.

Caccio, S.M., Sprong, H., 2010. Giardia duodenalis: genetic recombination and its implications for taxonomy and molecular epidemiology. Exp Parasitol 124, 107–112.

Caccio, S.M., De Giacomo, M., Pozio, E., 2002. Sequence analysis of the β-giardin gene and development of a polymerase chain reaction–restriction fragment length polymorphism assay to genotype *Giardia duodenalis* cysts from human faecal samples. Int J Parasitol 32, 1023–1030.

Caccio, S.M., Thompson, R.C., McLauchlin, J., Smith, H.V., 2005. Unravelling *Cryptosporidium* and *Giardia* epidemiology. Trends Parasitol 21, 430–437.

Caccio, S.M., Beck, R., Lalle, M., Marinculic, A., Pozio, E., 2008. Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B. Int J Parasitol. 38, 1523–1531.

Cooper, M.A., Adam, R.D., Worobey, M., Sterling, C.R., 2007. Population genetics provides evidence for recombination in Giardia. Curr. Biol. 17, 1984–1988.

Deng, M., Peterson, R.P., Cliver, D.O., 2000. First findings of *Cryptosporidium* and *Giardia* in California sea lions (*Zalophus californianus*). J. Parasitol. 86, 490–494.

Dixon, B.R., Parrington, L.J., Parenteau, M., Leclair, D., Santin, M., Fayer, R., 2008. *Giardia duodenalis* and *Cryptosporidium Spp.* in the intestinal contents of ringed seals (*Phoca hispida*) and bearded seals (*Erignathus barbatus*) in Nunavik, Quebec, Canada. J. Parasitol. 94, 1161–1163. ESRI Inc, 2010. ArcGIS. Version 10.0. Redlands, CA.

Feng, Y., Xiao, L., 2011. Zoonotic potential and molecular epidemiology of *Giardia* species and *Giardiasis*. Clin. Microbiol. Rev. 24, 110–140.

Gales, N.J., Shaughnessy, P.D., Dennis, T., 1994. Distribution, abundance and breeding cycle of the Australian sea-lion *Neophoca cinerea* (Mammalia: Pinnipedia). J. Zool. 234, 353–370.

Gaydos, J.K., Miller, W.A., 2008. Novel and canine genotypes of *Giardia duodenalis* in harbor seals (*Phoca vitulina richardsi*). J. Parasitol. 94, 1264–1268.

Goldsworthy, S., Gales, N., 2008. Neophoca cinerea. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.1. <http://www.iucnredlist.org> (accessed 08.06.30.).

Goldsworthy, S.D., Shaughnessy, P.D., Page, B., Dennis, T.E., MacIntosh, R.R., Hamer, D., et al., 2007. Developing population monitoring protocols of Australia sea-lions. Final report to the Department of the Environment andWater Resources. SARDI Res. Rep. Ser. 219, 75.

Hopkins, R.M., Meloni, B.P., Groth, D.M.,Wetherall, J.A., Reynoldson, R.C., Thompson, A., 1997. Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. J. Parasitol. 83, 44–51.

Hueffer, K., Holcomb, D., Ballweber, L.R., Gende, S.M., Blundell, G., O'Hara, T.M., 2011.

Serologic surveillance of pathogens in a declining harbor seal (*Phoca vitulina*) population in Glacier Bay National Park, Alaska, USA and a reference site. J. Wildl. Dis. 47, 984–988.

Hughes-Hanks, J.M., Rickard, L.G., Panuska, C., Sauciert, J.R., O'Harat, T.M., Dehn, L., et al., 2005. Prevalence of *Cryptosporidium spp*. and *Giardia spp*. in five marine mammal species. J. Parasitol. 91, 1225–1228.

Kosuwin, R., Putaporntip, C., Pattanawong, U., Jongwutiwes, S., 2010. Clonal diversity in Giardia duodenalis isolates from Thailand: evidences for intragenic recombination and purifying selection at the beta giardin locus. Gene 449, 1–8.

Lalle, M., Pozio, E., Capelli, G., Bruschi, F., Crotti, D., Caccio, S.M., 2005. Genetic heterogeneity at the *β-giardin* locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes. Int. J. Parasitol. 35, 207–213.

Lasek-Nesselquist, E., Bogomolni, A.L., Gast, R.J., Welch, D.M., Ellis, J.C., Sogin, M.L., et al., 2008. Molecular characterization of *Giardia intestinalis* haplotypes in marine animals: variation and zoonotic potential. Dis. Aquat. Organ. 81, 39–51.

Lasek-Nesselquist, E., Welch, D.M., Sogin, M.L., 2010. The identification of a new *Giardia duodenalis* assemblage in marine vertebrates and a preliminary analysis of *G. duodenalis* population biology in marine systems. J. Parasitol. 40, 1063–1074.

Lebbad, M., Mattsson, J.G., Christensson, B., Ljungstrom, B., Backhans, A., Andersson, J.O., et al., 2010. From mouse to moose: multilocus genotyping of *Giardia* isolates from various animal species. Vet. Parasitol. 168, 231–239.

Lowther, A.D., Harcourt, R.G., Goldsworthy, S.D., Stow, A., 2012. Population structure of adult female Australian sea lions is driven by fine-scale foraging site fidelity. Anim. Behav. 83, 691–701.

Monis, P.T., Thompson, R.C., 2003. *Cryptosporidium* and *Giardia* zoonoses: fact or fiction? Infect. Genet. Evol. 3, 233–244.

Monis, P.T., Andrews, R.H., Mayrhorfer, G., Ey, P.L., 1999. Molecular systematics of the parasitic protozoan *Giardia intestinalis*. Mol. Biol. Evol. 16, 1135–1144.

Monis, P.T., Andrews, R.H., Mayrhorfer, G., Ey, P.L., 2003. Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. Infect. Genet. Evol. 3, 29–38.

Olson, M.E., Roach, P.D., Stabler, M., Chan, W., 1997. *Giardiasis* in ringed seals from theWestern Arctic. J. Wildl. Dis. 33, 646–648.

Read, C.M., Monis, P.T., Andrew Thompson, R.C., 2004. Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP. Infect. Genet. Evol. 4, 125–130.

Rengifo-Herrera, C., Ortega-Mora, L.M., Gomez-Bautista, M., Garcia-Moreno, F.T., Garcia-Parraga, D., Castro-Urda, J., et al., 2011. Detection and characterization of a *Cryptosporidium* isolate from a Southern elephant seal (*Mirounga leonina*) from the Antarctic Peninsula. Appl. Environ. Microbiol. 77, 1524–1527.

Rengifo-Herrera, C., Ortega-Mora, L.M., Gomez-Bautista, M., Garcia-Pena, F.J., Garcia-Parraga, D., Pedraza-Diaz, S., 2013. Detection of a novel genotype of *Cryptosporidium* in Antarctic pinnipeds. Vet. Parasitol. 191, 112–118.

7.3 Human Research Ethics Approvals

7.3.1 Macquarie University Human Research Ethics Approval

Amy Asher <amy.asher@mq.edu.au>

Final Approval- Ethics application reference-5201100403

Ethics Secretariat <ethics.secretariat@mq.edu.au> Tue, Jul 12, 2011 at 3:27 PM
To: Dr Michelle Power <michelle.power@mq.edu.au>
Cc: Miss Amy Jeannette Asher <amy.asher@students.mq.edu.au>

Dear Dr Power

Re: "Molecular epidemiology of Giardia in Australia" (Ref: 5201100403)

The above application was reviewed by the Human Research Ethics Committee. Final Approval of the above application is granted, effective 12 July 2011, and you may now commence your research.

The following personnel are authorised to conduct this research:

Dr Michelle Power- Chief Investigator/Supervisor
Miss Amy Asher- Co-Investigator

NB. STUDENTS: IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF THIS APPROVAL EMAIL TO SUBMIT WITH YOUR THESIS.

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports. Your first progress report is due on 12 July 2012

If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the

Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy/

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of final ethics approval.

Yours sincerely
Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee

7.3.2 NSW Population and Health Services Human Research Ethics Approval

Dr Michelle Power
Department of Biological Sciences
Macquarie University
North Ryde
NSW 2109

13 February 2013

Dear Dr Power,

NSW Population & Health Services Research Ethics Committee

AU RED Reference: HREC/12/CIPHS/87

Cancer Institute NSW reference number: 2012/12/435

Project Title: Giardiasis transmission in NSW communities: an epidemiological assessment of Giardia strains contributing to human disease

Thank you for your correspondence of 21 January 2013 responding to a request for further information/clarification of the above referenced study, submitted to the NSW Population & Health Services Research Ethics Committee for single ethical and scientific review. The Executive of the Committee reviewed your response on 11 February 2013 and I am pleased to inform you that full ethical approval has been granted.

The Ethics Committee granted a waiver of the usual requirement for the consent of the individual to the use of their health information in a research project, in line with the State Privacy Commissioner's Guidelines for Research and the Health Records and Information Privacy Act 2002 (NSW).

The documents reviewed and approved include:

National Ethics Application Form, submission code AU/1/7120112, dated 21/11/2012
NSW Privacy Addition to NEAF
Research Protocol Version 1.0 dated November 2012
Data Custodian sign off, NDD, dated 3 September 2012
Macquarie University HREC final approval dated 12 July 2011
Request for Release of Data NSW Health Communicable Diseases
Curriculum Vitae: Michelle Power
Curriculum Vitae: Amy Asher

The NSW Population & Health Services Research Ethics Committee has been accredited by the NSW Department of Health to provide single ethical and scientific review of research proposals conducted within the NSW public health system.

The Committee is a joint initiative of the Cancer Institute NSW and NSW Department of Health. The Committee has been constituted and operates in accordance with the National Health and Medical Research Council's *National Statement on Ethical Conduct in Human Research (2007)* and relevant legislation and guidelines.

Please note that ethical approval is valid for **5 years**, conditional on the following:

Principal investigators will immediately report anything which might warrant a review of ethical approval of the research, including unforeseen events that might affect continued ethical acceptability.
Proposed amendments to the research proposal or conduct of the research which may affect the ethical acceptability of the research are to be provided to the NSW Population & Health Services Research Ethics Committee for review.
The NSW Population & Health Services Research Ethics Committee will be notified giving reasons, if the research is discontinued before the expected date of completion.
The Principal Investigator will provide an annual progress report to the NSW Population & Health Services Research Ethics Committee and at the completion of the study.

You are reminded that this letter constitutes *'ethical approval'* only. This research project must not commence at a site until separate authorisation from the Chief Executive or delegate of that site has been obtained. It is your responsibility to forward a copy of this letter together with any approved documents as enumerated above, to all site investigators for submission to the site's Research Governance Officer. Where relevant, copies will also need to be provided to the CHeReL and the data custodian.

For further information about the NSW Population & Health Services Research Ethics Committee, please refer to our website www.cancerinstitute.org.au/research.

Should you have any queries about the ethical review of your research proposal, please contact me on 02 8374 3562 or email ethics@cancerinstitute.org.au.

The NSW Population & Health Services Research Ethics Committee wishes you well in your research endeavours.

Yours sincerely,

V Turner

Virginia Turner
Ethics Coordinator
Cancer Institute NSW