The role of protozoan genetic diversity in human disease: Implications for the epidemiology of cryptosporidiosis and giardiasis in New Zealand

A thesis submitted in partial

fulfilment of the requirements for the

degree of Doctor of Philosophy

Massey University

Palmerston North

New Zealand

Paul Chiamaka Ogbuigwe

2021

Prelude

If I were a parasite, I'd curse the gods

for burdening me with such a terrible fate

To linger in the guts of naturist frauds

and serve my time as diarrhoea bait

It's not a fate I'd wish on my enemy

or stoically endure in feigned masculinity

But there is another side to the condition

a level of resilience that defies cognition

The ability to withstand the most potent of acids

and resist frigid winds without going flaccid

They leap from man to cow evading pursuit

and no vaccines can sour their fruit

Forgive them the children that perish from their kiss

or the immunocompromised they lead to the abyss

They have enough qualities to fill a book

I've written one now have a look!

Abstract

Cryptosporidium and *Giardia* are two common causes of diarrhoea in humans and livestock and responsible for multiple outbreaks of gastroenteritis every year in New Zealand and around the globe. Despite their prevalence, there are few effective therapies or vaccines, against either parasite. This is largely due to the difficulty of manipulating these parasites *in vitro*. The understanding of the epidemiology of this parasite in New Zealand is incomplete, due to the presence of multiple dominant subtypes of each parasite within samples from the same outbreak. In this thesis, new techniques are employed to investigate the genetic diversity of these parasites within hosts and develop an *in vitro* assay for comparing the infectivity of multiple subtypes of *Cryptosporidium*.

Current methods for the purification of *Giardia* cysts from faecal samples do not adequately remove debris from the sample and produce low numbers of purified cysts. This hampers molecular techniques that benefit from uncontaminated samples resulting in the use of expensive methods like immunomagnetic separation. Here, a novel method for the purification of cysts from faecal samples was developed, which produced purified oocysts with negligible debris and a 10-fold increase in yield over current techniques.

Epidemiological and molecular investigations of past giardiasis and cryptosporidiosis outbreaks in New Zealand have highlighted inconsistent results, where epidemiologically linked cases can have different dominant subtypes identified through Sanger sequencing. Here, amplicon-based metabarcoding was utilised to resolve *Giardia* and *Cryptosporidium* outbreak epidemiology in New Zealand. Human faecal samples from past outbreaks previously classified using Sanger sequencing were analysed using next-generation sequencing. This strategy uncovered significant within-host diversity and identified potential emerging subtypes of *Cryptosporidium* that could have public health significance in the future. Analysis of diversity within outbreaks provided previously unidentified genetic links between samples from the same outbreak.

Previous studies show that people experience different symptoms depending on the subtype of *Cryptosporidium* they are infected with. Also, the dominant subtypes of the parasite in a region, like the USA and Australia, have changed multiple times within the past 20 years. This suggests there are differences in infectivity between subtypes, but further analysis of this problem has been hampered by the lack of adequate cell culture systems that allow the complete development of the parsite *in vitro*. To better understand the differences in infectivity between subtypes of *Cryptosporidium,* an analysis of the expression of *Cryptosporidium* genes in the COLO-680N cell line at multiple timepoints during infection was carried out using the NanoString nCounter analysis system. This was done to investigate whether differences in gene expression could account for differences in infectivity. Furthermore, utilising flow cytometry a system was developed capable of identifying and quantifying infection in infected cells with and without the use of a fluorescent antibody. A novel signal was identified in the near-infra red range that was specific to *Cryptosporidium* infection and showed better signalling characteristics than the fluorophore.

Acknowledgments

It has been 12 years since I began my undergraduate and set my sights on reaching this point someday. I thank all the friends and my lovely, supportive family for standing by, and sometimes dragging me by the ear, as I strove to make my dream of being a scientist a reality.

Table of Contents

Table of Figures

Tables

Glossary

Confluency – a term used in cell or tissue culture to determine how much of the surface of the culture flask has been covered by cells.

Disability adjusted life years (DALYs) – a term-based measure that combines years of life lost due to premature mortality and years of life lost due to time lived in states of less than full health, or years of healthy life lost due to disability.

Encystation – the process by which the infectious stages of *Giardia* are enclosed in a protective shell forming a cyst before releasing into the environment.

Excystation – the process by which the wall of the environmental stage of *Cryptosporidium* and *Giardia* open along a suture to allow the release of the infectious stages held within.

Flow cytometry – a laser-based method used to analyse the physical and chemical characteristics of cells.

Gating – a term used in flow cytometry to define the process of the sequential identification and refinement of a cellular population of interest using a panel of markers.

Multiplicity of Infection (MOI) – the ratio of infectious agents to infection targets.

NanoString nCounter – A microarray that uses molecular barcodes and fluorescence microscopy to detect and count up to 800 RNA, DNA or protein targets within a sample.

Next-generation sequencing (NGS) – DNA sequencing techniques that allow millions of fragments to be sequenced in a single run allowing the researcher to separate the signal originating from each target molecule, thus allowing the efficient isolation, detection, and quantification of rare events.

Outbreak – an increase in the occurrence of a disease at a particular time and a specific place, or, two or more cases linked to a common source, in particular, where the common source is exposure at a common event, or to food or water dispersed in a community, an environmental source or a source in an institutional setting.

Pathognomonic – A symptom specifically characteristic of a disease or condition.

RNA interference (RNAi) – and RNA-dependent gene silencing process where short interfering RNAs bind to a target mRNA molecule and cleave it, thereby preventing translation.

Sanger sequencing – a DNA sequencing technique that combines the contribution of all DNA fragments present in the reaction mixture to produce one forward and reverse read.

Sequelae – long-term conditions resulting from a temporary disease.

1 General Introduction

1.1 Abstract

This chapter provides a general introduction for the topics covered in this thesis. The public health significance of the two parasites, *Cryptosporidium* and *Giardia,* upon which this work is based is laid out in full. A significant amount of research has been published covering both parasites. This section analyses the most current understanding of the taxonomy and life cycle, infection and transmission, epidemiology of outbreaks, and *in vitro* culture methods of both parasites, focusing on topics relevant to this thesis.

1.2 Introduction

Cryptosporidium and *Giardia* are two protozoan parasites that are among the most common causes of diarrhoea in humans and farm animals worldwide. The diseases caused by *Cryptosporidium* and *Giardia* are referred to as cryptosporidiosis and giardiasis, respectively. In immunocompetent human beings infections are usually acute and self-limiting, however, they can result in death in immunocompromised individuals, such as those infected with the human immunodeficiency virus (HIV) (Huang & White, 2006a; Wang et al., 2018), and children under five years old (Einarsson et al., 2016a; Khalil et al., 2018; Mmbaga & Houpt, 2017). According to the World Health Organisation (WHO), globally, *Cryptosporidium* and *Giardia* are, respectively, the second and third most common causes of diarrhoea in children under the age of five years old, preceded only by rotavirus (Lanata et al., 2013). Both parasites are capable of infecting all taxonomic classes of vertebrates excluding Agnatha. However, due to the difficulty of diagnosing the diseases in wildlife, it is unclear if the presence of *Cryptosporidium* and *Giardia* in most wild animals is a result of asymptomatic carriage or actual disease progression (Robertson et al., 2019; Thompson et al., 2011). The dominant mode of transmission of these parasites in humans is through the faecal-oral route. While contaminated water and food are the main vectors in this mode of transmission (Wuhib et al., 1994), it is possible for infection to occur directly through contact with contaminated surfaces (Rose, 1997).

Cryptosporidium and *Giardia* are endemic in all countries and considered particularly prevalent in low- and middle-income countries (LMICs) (Putignani & Menichella, 2010; Thompson et al., 2013). Globally, multiple outbreaks and as many as 300 million reported cases of giardiasis occur each year (Cernikova et al., 2018). The obvious acceptance of the public health significance of these diseases has led to the classification of cryptosporidiosis and giardiasis as notifiable diseases in most countries, as well as their inclusion in the Neglected Diseases Initiative by the WHO in 2004 (Savioli et al., 2006a). *Cryptosporidium* and *Giardia* are associated with both sporadic cases and outbreaks. Here, outbreaks are defined as an increase in the occurrence of a disease at a particular time and a specific place, the duration of an outbreak can range from days or weeks to seasons or years, and the location can range from a small localised group to an entire continent (Smith et al., 2014). In 1993, a particularly notable outbreak of cryptosporidiosis occurred in Milwaukee, United States of America (USA) involving a total of 403,000 cases, making it the largest waterborne disease outbreak documented in the USA since records began (Satcher et al., 1996). A study compiled by Efstratiou et al., (2017) covering reported cases from around the world between the years of 2011 and 2016 found 239 waterborne outbreaks of cryptosporidiosis and 142 of giardiasis occurred during that timeframe. Although outbreaks caused by *Cryptosporidium* species are more common, *Giardia* species are more prevalent, infecting approximately 10% of the world's population while *Cryptosporidium* infects around 3-5% (Huang & White, 2006b). However, the reporting of these diseases is imperfect in many countries, especially in LMICs, so the burden of these diseases and the number of outbreaks could be significantly underestimated (Checkley et al., 2015; Current & Garcia, 1991), particularly in outbreaks of gastroenteritis where multiple competing parasites are found in infected individuals (Thompson et al., 2011). Previous studies have found that some cases of the severe paediatric diarrhoeal disease initially attributed to rotavirus are now attributed to *Cryptosporidium* (Love et al., 2017).

Despite the global prevalence of cryptosporidiosis and giardiasis there are few effective treatments for humans and no vaccines. Currently, nitazoxanide is the only drug approved by the United States Food and Drug Administration (FDA) for the treatment of cryptosporidiosis and is sometimes employed in the treatment of giardiasis (Bones et al., 2019; Ordóñez-Mena et al., 2018). Metronidazole is usually the drug of choice for the

treatment of giardiasis (Lane & Lloyd, 2002), still, some other drugs from the 5-nitroimidazole class, such as tinidazole and secnidazole, can prove effective in the treatment of the disease to a somewhat lesser extent (Ordóñez-Mena et al., 2018). A major hindrance in the search for effective treatments against these parasites is their recalcitrance towards *in vitro* culture. However, the availability of *in vivo* culture methods and advances in molecular biology and sequencing technologies are providing new knowledge and tools that facilitate a better understanding of the biological processes of these organisms. The following review will provide information on the current understanding of the epidemiology of these organisms both in New Zealand and globally, and then a summary of the current trends and advances in the field of molecular biological techniques applicable to *Cryptosporidium* and *Giardia*.

1.3 Taxonomy and Life Cycle

1.3.1 *Cryptosporidium* Taxonomy

Cryptosporidium belongs to the phylum Apicomplexa and the family Cryptosporidiidae. The parasites of the family Cryptosporidiidae possess a multi-membranous attachment organelle, a defining feature used in the taxonomic classification of this family since 1961 (Levine, 1961) that distinguishes it from other coccidian parasites (U. Ryan et al., 2016). For many years after the discovery of the genus *Cryptosporidium* by Ernest Edward Tyzzer, who first described the parasites from the gastric glands of mice (Tyzzer, 1907), morphological characteristics served as the basis for the characterisation and classification of this parasite. This created some confusion due to the morphological similarity of the environmental stages (oocysts) of *Cryptosporidium*, making it difficult to differentiate between species when the specimens were examined microscopically. The advent of molecular techniques such as PCR and DNA sequencing overcame these difficulties, using genetic differences in concert with morphological measurements to greatly advance our understanding of the taxonomy of *Cryptosporidium*. As a result of these efforts, 38 species of *Cryptosporidium,* listed in table 1.1 have been described to date (Chalmers & Katzer, 2013; Feng et al., 2018; Holubová et al., 2019, 2020; Ježková, Limpouchová, et al., 2021; Ježková, Prediger, et al., 2021; Khan et al., 2018a; U. Ryan et al., 2014; Traversa, 2010). The species of *Cryptosporidium* most frequently associated

with human infections are *Cryptosporidium parvum* and *Cryptosporidium hominis*. *C. parvum* is the species with the widest host range, covering humans, domestic mammals and wildlife (Cacciò & Widmer, 2013). *C. hominis* was initially thought to be a genotype within *C. parvum* but was found to be its own species based on genetic data (Morgan-Ryan et al., 2002). This species infects humans but has been found in other animals such as pigs and cows. This is thought to be due to reverse zoonotic transmission in most cases (Abeywardena et al., 2012; Widmer et al., 2020).

Table 1.1. List of accepted species of *Cryptosporidium* (Feng et al., 2018; Holubová et al., 2019, 2020; Ježková, Limpouchová, et al., 2021; Ježková, Prediger, et al., 2021; Traversa, 2010).

Species	Primary Host	Reference
C. hominis	Humans	(Morgan-Ryan et al., 2002)
C. parvum	Livestock	(Tyzzer, 1912)
C. muris	Rodents	(Tyzzer, 1907)
C. tyzzeri	Domestic Mice	(Ren et al., 2012)
C. meleagridis	Turkeys	(Slavin, 1955)
C. ubiquitum	Varied Mammals	(Fayer et al., 2010)
C. suis	Pigs	(U. M. Ryan et al., 2004)
C. cuniculus	Rabbit	(Robinson et al., 2010)
C. bovis	Cows	(Fayer et al., 2005)
C. ryanae	Cattle	(Fayer et al., 2008)
C. canis	Dogs	(Fayer et al., 2001)
C. felis	Cats	(Iseki, 1979)
C. macropodum	Kangaroos	(Power & Ryan, 2008)
C. andersoni	Cattle	(Lindsay et al., 2000)
C. baileyi	Chickens	(Current et al., 1986)
C. galli	Finches	(U. M. Ryan et al., 2003)
C. serpentis	Snake	(Jirků et al., 2008)
C. molnari	Fish	(Barugahare et al., 2011)
C. wrairi	Guinea Pigs	(Vetterling et al., 1971)
C. homai	Guinea Pigs	(Zahedi, Durmic, et al., 2017)
C. ducismarci	Tortoises	(Traversa, 2010)
C. ornithophilus	Ostriches	(Holubová et al., 2020)

Species	Primary Host	Reference
C. xiaoi	Sheep	(Fayer & Santín, 2009)
C. fayeri	Marsupials	(U. M. Ryan et al., 2008)
C. fragile	Toads	(Jirků et al., 2008)
C. scrofarum	Pigs	(Kváč et al., 2013)
C. viatorum	Humans	(Elwin et al., 2012)
C. agni	Sheep	(Barker & Carbonell, 1974)
C. varanii	Emerald Monitors	(Pavlasek & Ryan, 2008)
C. erinacei	Hedgehogs	(Kváč et al., 2014)
C. scophthalmi	Turbot	(Alvarez-Pellitero et al., 2004)
C. proliferans	Rodents	(Kváč et al., 2016)
C. avium	Birds	(Holubová et al., 2016)
C. huwi	Guppies	(U. Ryan et al., 2015)
C. rubeyi	Squirrels	(X. Li et al., 2015)
C. apodemi	Mice	(Čondlová et al., 2018)
C. ditrichi	Mice	(Čondlová et al., 2018)
C. testudinis	Tortoises	(Ježková et al., 2016)
C. alticolis	Voles	(Horčičková et al., 2019)
C. microti	Voles	(Horčičková et al., 2019)
C. occultus	Rats	(Kváč et al., 2018)
C. ratti	Brown Rats	(Ježková, Prediger, et al., 2021)
C. myocastoris	Coypu	(Ježková, Limpouchová, et al., 2021)
C. proventriculi	Birds	(Ježková, Limpouchová, et al., 2021)

1.3.2 *Cryptosporidium* Life Cycle

The environmental stage of *Cryptosporidium* is called the oocyst. The oocyst possesses a particularly stalwart wall composed of a glycocalyx layer, carbohydrates, fatty acids, aliphatic hydrocarbons, hydrophobic proteins and glycoproteins in a complex lattice structure (Harris & Petry, 1999; Jenkins et al., 2010). This gives it resistance to common disinfectants such as chlorine and bleach and allows it to remain viable in the environment for up to 24 weeks, depending on temperature conditions and exposure to UV light (Bogan, 2019; Drummond et al., 2018; Fayer et al., 1998; Keegan et al., 2008). The life cycle of *Cryptosporidium* is shown in figure 1.1. Within each oocyst are 4 sporozoites, the life cycle stage that initiates infection in the host. Upon ingestion of oocysts by a viable host, the parasite travels through the gastrointestinal tract where the environmental conditions, particularly temperature and pH of the distal ileum, trigger excystation thereby releasing the 4 motile sporozoites contained within. The sporozoites attach to the plasma membrane of the epithelial cells of the small intestine. They are then internalised, forming a parasitophorous vacuole, developing into trophozoites and progressing through the rest of their life cycle stages (Carey et al., 2004). This vacuole is intracellular but extra-cytoplasmic (Elliott & Clark, 2000).

Cryptosporidium is capable of both sexual and asexual reproduction. In the asexual life cycle, the trophozoites develop into type I meronts containing 6-8 type I merozoites. The type I merozoites are released into the lumen of the small intestine where they invade new host cells, developing into new trophozoites, and eventually becoming type II meronts containing 4 type II merozoites. They then begin the sexually reproductive stage of the life cycle. Type II merozoites develop into macrogamonts (corresponding to an ovum) or microgamonts. Microgamonts produce microgametes (corresponding to sperm cells), which fertilise macrogamonts to form zygotes. The zygotes can develop into 2 types of oocysts: thin-walled oocysts that maintain the infection within the host, or, thick-walled oocysts that are passed out into the environment through the faeces of the host (Tandel et al., 2019). Most oocysts are infective immediately upon excretion.
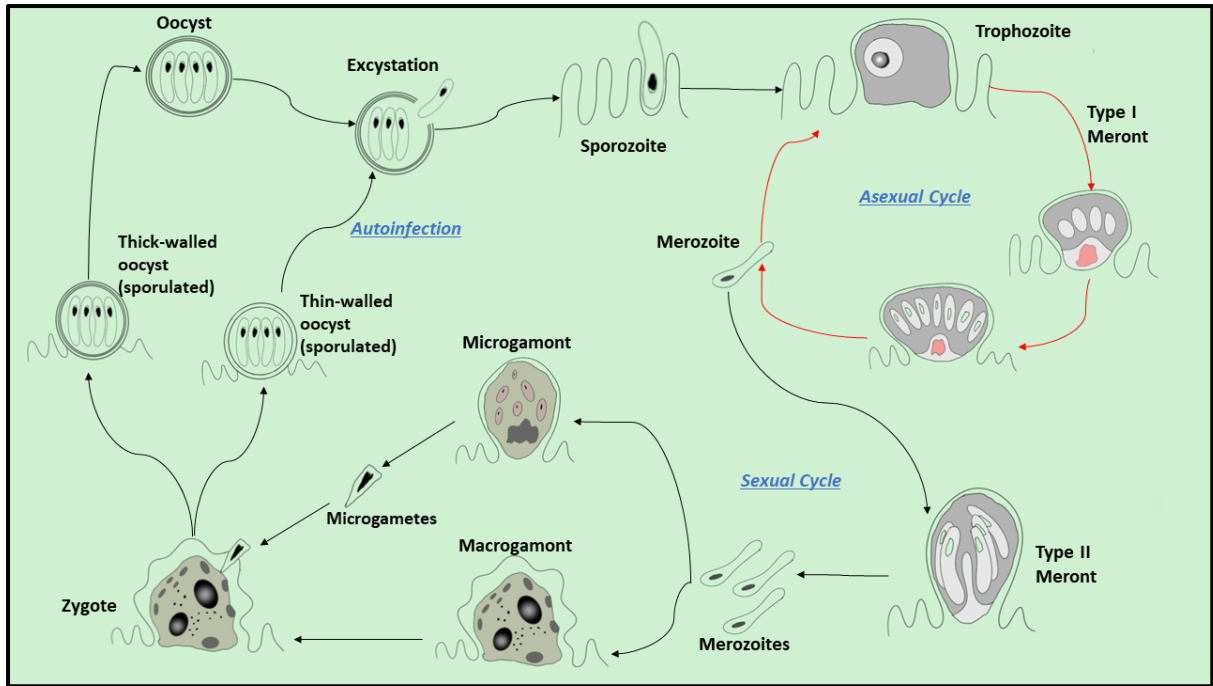
Figure 1.1. Life cycle of *Cryptosporidium*.

1.3.3 *Giardia* Taxonomy

Giardia belongs to the phylum Metamonada and the family Hexamtidae. Organisms belonging to this family notably possess paired organelles including a pair of transcriptionally active diploid nuclei; accordingly, they are referred to as diplozoic-flagellated protozoa (Luján & Svärd, 2011). *Giardia* was initially discovered by Antony van Leeuwenhoek in 1681, however, the name was established by Kunstler approximately 200 years later in 1882. Since then, there has been much debate about the taxonomy of this parasite, and, similarly to *Cryptosporidium*, much of the difficulty was due to the morphological similarities between the various species of this parasite. Nevertheless, eight species of *Giardia,* shown in table 1.2, have been described in vertebrate hosts (U. Ryan et al., 2019; Thompson & Monis, 2004).

Table 1.2. List of accepted species of *Giardia* (U. Ryan et al., 2019; Thompson & Monis, 2004)*.*

Species	Primary Host	References
G. intestinalis	Humans	(Filice, 1952)
G. agilis	Amphibians	(Künstler, 1882)
G. ardae	Great Blue Herons	(Erlandsen et al., 1990)
G. cricetidarum	Hamsters	(Lyu et al., 2018)
G. microti	Voles	(Feely, 1988)
G. muris	Rodents	(Kulda, 1978)
G. peramelis	Southern Brown Bandicoots	(Hillman et al., 2016)
G. psittaci	Budgerigars	(Erlandsen & Bemrick, 1987)

According to our current understanding, *Giardia intestinalis* (also known as *Giardia lamblia* or *Giardia duodenalis*) is the only species of this parasite known to be responsible for infections in humans. *Giardia intestinalis* has a wide range of hosts including livestock, pets, and wildlife. *G. intestinalis* has further been divided into 8 assemblages (A-H), with assemblages A and B being the ones most frequently identified in human infections. These assemblages correspond to genotypes or genetic clusters. However, due to the continuing uncertainty vis-à-vis the taxonomy of *Giardia,* some assemblages are sometimes referred to as distinct species. For example, a review suggested assemblages E, F and G should be alternatively called *G. bovis, G. cati* and *G. simondi*, respectively (Thompson et al., 2011), nevertheless, these names have not been formally described so they do not comply with the International Code of Zoological Nomenclature (ICZN). Recent advances in molecular techniques, such as PCR, sequencing technologies, and cell culture, have the potential to further aid the taxonomic classification of *Giardia*, especially as the understanding of the host specificity of the various subtypes of this ubiquitous protozoan develops.

1.3.4 *Giardia* Life Cycle

Giardia begins its life cycle as a cyst. The cyst is the environmental/transmissible stage of this parasite and it possesses a hardy wall made up of a mesh of cyst wall proteins complexed to a singular sugar polymer of (β1–3)-linked N-acetylgalactosamine (GalNAc) (Samuelson & Robbins, 2011), which makes it resistant to most solvents and able to survive in the environment for several months in cold conditions (~10 °C) (Bingham et

al., 1979; Jarroll & Hoff, 1988). Each cyst encloses 2 trophozoites, the release of which is triggered by environmental cues in the host duodenum in a process called excystation. The trophozoites float freely or attach to the epithelial cells in the lumen of the small intestine using a ventral adhesive disk, however, unlike their *Cryptosporidium* counterparts, they proliferate and cause disease while remaining in the extracellular space (Horlock-Roberts et al., 2017). Each trophozoite is covered in a thick coat of variable surface proteins (VSPs). Most species of *Giardia* have approximately 270 VSP genes, only one of which is expressed at a time but periodically interchanged by another in a process known as VSP switching (Cernikova et al., 2018). The exact process by which switching occurs remains up for debate, but some studies suggest it does so using a mechanism similar to RNA interference (RNAi) (Rivero et al., 2010). The VSP complex presents an attractive target for the development of drugs and vaccines for the treatment of giardiasis, and various studies are looking into deregulation of the expression of this complex (Rivero et al., 2010). In fact, vaccines targeting the VSP repertoire have been tested in gerbils, cats and dogs (Serradell et al., 2016). The trophozoites multiply through mitosis. This was thought to be their only form of reproduction, but several studies have found evidence of the ability to undergo sexual reproduction, however, the mechanism is not fully understood (Adam, 2021; Birky, 2010; Cooper, Adam, Worobey, & Sterling, 2007). The final stage of the life cycle of *Giardia* is triggered as the parasite transits towards the colon. This triggers encystation – the formation of new cysts – and the new cysts are excreted by the host spreading the infection out in the environment.
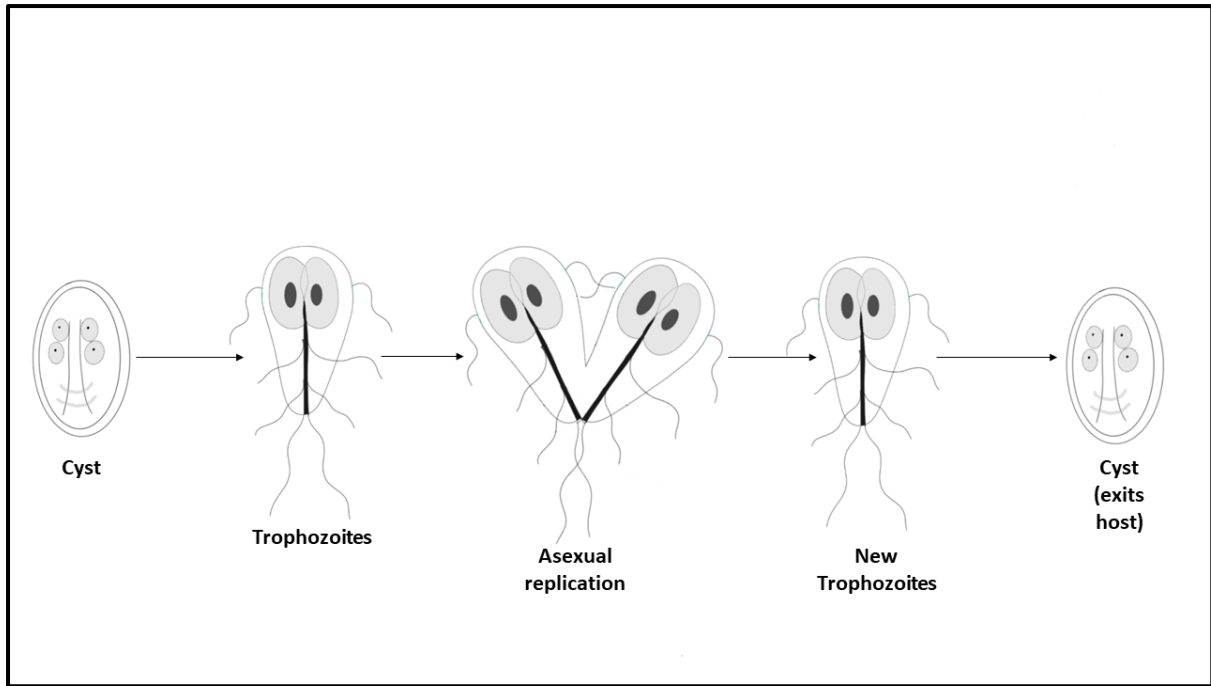
Figure 1.2. Life Cycle of *Giardia*.

1.4 Infection

1.4.1 *Cryptosporidium* Infection in Humans

The species of *Cryptosporidium* responsible for most human infections are *C. hominis* and *C. parvum*. *C. hominis*, subdivided into over ten subtypes, is mainly found in humans and its presence in other animals is usually attributed to anthroponotic transmission (Feng et al., 2018). In contrast, *C. parvum* is most common in livestock and is considered chiefly a zoonotic infection in humans, apart from the anthroponotic subtype IIc (Bouzid et al., 2013; Ng et al., 2012; Thomson et al., 2017). Due to the wide host range of *C. parvum* it can be subdivided into over twenty subtypes (Feng et al., 2018). The differences in the host range of the two main species of *Cryptosporidium* allow for insight into the epidemiology of the parasite. A study looking at the distribution of *Cryptosporidium* genotypes globally and locally found that in New Zealand, two spikes in the number of cases of cryptosporidiosis occur per annum (Garcia–R et al., 2017a), and the largest increase occurred during the spring calving season (August – October) which coincided with an increase in *C. parvum* genotypes, suggesting zoonotic transmission (Learmonth,

Ionas, Pita, & Cowie, 2002). A second peak (January – May) in cases of cryptosporidiosis in New Zealand has been attributed to anthroponotic transmission (Lake, Pearce, & Savill, 2008; Learmonth, Ionas, Ebbett, & Kwan, 2004). From this, it can be inferred that while humans serve as an important reservoir for the pathogen, zoonotic transmission also plays a significant role in the seasonal spread of the disease. Many other species of *Cryptosporidium* have been found in humans, including *C. erinacei, C. cuniculus, C. meleagridis* and *C. tyzzeri* (Cacciò & Widmer, 2013; Chalmers & Katzer, 2013; Swaffer et al., 2018). Their presence is not simply due to asymptomatic carriage, these studies confirmed they actually have the ability to cause disease in humans.

The signature symptom of cryptosporidiosis is severe, non-bloody diarrhoea but other potential symptoms include nausea, vomiting, abdominal pain, weight loss and fever (Bones et al., 2019; Cacciò & Widmer, 2013). The incubation period of the disease ranges from 4-28 days, with most infections lasting 6-7 days (Hunter et al., 2004a). Interestingly, while fatigue and joint pain have been found to develop as post-infectious sequelae in some people infected with *Cryptosporidium*, it has been found that eye pain and recurring headaches occur only from *C. hominis* infections, and these symptoms are not observed in *C. parvum* infections (Carter et al., 2020; Hunter et al., 2004a). Further investigations into the differences in clinical effects triggered by the various species of this parasite would have to be carried out to determine the reason for this difference in sequelae.

As the group most vulnerable to infection, the effects of cryptosporidiosis in children can be significantly more severe when compared to the effects observed in immunocompetent adults (Troeger et al., 2017). Chronic infection in children can result in severe malabsorption syndrome, growth retardation and even cognitive defects, particularly in resource-limited settings (Delahoy et al., 2018; Tumwine et al., 2003). The severity of the disease is affected by age, nutrition, immune status, and potentially by the species and subtype of *Cryptosporidium* that infects the host (Cama et al., 2008). The Global Burden of Disease (GBD) study carried out in 2016 found 80% of deaths from cryptosporidiosis occurred in children under the age of 5 (Naghavi et al., 2017), further highlighting the severe effects of the disease on this age group. It does not help that the only FDA-approved drug, nitazoxanide, is less efficacious in young children and immunocompromised individuals (Amadi et al., 2002; Chavez & Jr, 2018). Chronic

cryptosporidiosis can negatively affect the normal gut function of young or immunocompromised hosts by altering endothelial cells, flattening microvilli, and causing chronic inflammation (Khalil et al., 2018). Together, this causes a reduction in the area of absorptive tissue. Similar to adults, it has been found that *C. hominis* infection causes more severe illness in children relative to *C. parvum* and other species (Bushen et al., 2007). This shows that public health bodies need to come up with better strategies to tackle human-human transmission to relieve the burden of this disease on the most vulnerable members of our communities. A study in 2016 estimated that separate from the 4.2 million disability-adjusted life years (DALYs) due to diarrhoeal episodes and deaths caused by *Cryptosporidium,* an additional 7.85 million DALYs are due to its effects on growth (Khalil et al., 2018).

The differences in the severity and sequelae of infections from *Cryptosporidium* highlight the importance of capturing the global and local genetic diversity of the parasite to better inform the public health officials tasked with devising strategies for the management of the disease. In addition, better *in vitro* techniques for the manipulation of the parasite would speed up the development of drugs to combat the disease and could give an understanding of the biological mechanisms that account for the observed differences in clinical symptoms between the various species and subtypes of *Cryptosporidium*.

1.4.2 Transmission and Risk Factors of Cryptosporidiosis

There are multiple transmission routes through which a person can get infected with *Cryptosporidium*. The two most common routes are by contact with contaminated food and water. The waterborne route includes both the ingestion of water as part of diet and exposure to contaminated recreational bodies of water, such as community swimming pools, lakes, and rivers. The resistance of oocysts to common disinfectants, such as chlorine and bleach, means that if rigorous methods of water purification are not employed *Cryptosporidium* can persist in drinking and recreational water sources. In addition to this, studies of previous outbreaks show that it is possible to acquire infection through direct human-human contact, contact with infected animals, and travel to endemic countries (Cacciò & Widmer, 2013). Socioeconomic factors and geographic settings affect the dominant mode of transmission in a particular area or country. It has

been shown that children <5 years old in New Zealand have the greatest risk of cryptosporidiosis when they live in areas with high densities of dairy cattle (Lal et al., 2016), furthermore, when all age groups were considered the highest rates of the disease was observed in those living in low-deprivation areas (Snel, Baker, Kamalesh, French, & Learmonth, 2009). This is of particular importance because global studies show the highest incidence of cryptosporidiosis is found in high-deprivation communities. The difference in New Zealand could be due to the large and lucrative livestock industry, which results in many wealthy farmers living in rural areas surrounded by cattle, thereby increasing their risk of zoonotic transmission of cryptosporidiosis and offsetting socioeconomic patterns elsewhere. As previously stated, studies in New Zealand have shown that there are two peaks in cases of cryptosporidiosis each year, one at early autumn attributed to anthroponotic transmission (e.g. human-human contact in recreational swimming facilities), and a larger peak in spring that coincides with lambing/calving season suggesting largely zoonotic transmission (e.g. contact with infected animals) (Garcia–R et al., 2017a; Learmonth et al., 2002; Learmonth et al., 2004).

In addition to the faecal-oral route of transmission, recent studies have shown evidence of respiratory infections by *Cryptosporidium.* Sponseller et al., (2014) describe evidence for the presence of respiratory cryptosporidiosis in multiple avian and mammal species. With regards to humans, they describe studies showing the presence of *Cryptosporidium spp.* In samples from infected individuals over 3 months, suggesting that respiratory cryptosporidiosis may be common in immunocompetent children suffering unexplained coughs in addition to the usual diarrhoea expected during infection. This evidence indicates that it is not merely a case of the parasite asymptomatically showing up in the lungs, ostensibly, *Cryptosporidium* is a more versatile and adaptive organism than the simple enteric parasite it was thought to be. Sayed et al., (2016) showed that infection of mice with two strains of *C. parvum* led to hypercellularity and acute inflammatory cell infiltration in the brain parenchyma of mice infected with both strains among other pathologies. Interestingly, while both strains used were *C. parvum*, there were substantial differences in the oocysts shedding rate, clinical outcomes, and the histopathological pictures of the intestines, lungs, and brains of infected mice. Furthermore, Audebert et al., (2020), using carcinogenic isolates of *C. parvum* – virulent strains that caused cancer in severe immunodeficiency (SCID) mice - extracted from infected humans and animals to

infect mice, noticed phenotypic differences, including differences in clinical manifestations, mortality rate, rate of neoplastic lesion progression, and the development of extra-gastrointestinal lesions. However, considering that the study used immunocompromised mice and the phenomenon has not been noticed outside laboratory settings, it could not be definitively stated that the parasite caused the cancer. The results suggest potential newly identified sequelae that can occur in *Cryptosporidium* infections in severely immunocompromised individuals.

Studies looking at the genetic diversity of *Cryptosporidium* species across the globe have found that *C. hominis* is the dominant species found in high-income countries, while the anthroponotic *C. parvum* subtype IIc is the dominant subtype found in LMICs, with high-income countries showing a lower prevalence of this subtype (King, Tyler, & Hunter, 2019). This finding was based on data from children and HIV-positive individuals in multiple countries, including Nigeria, Kenya, Uganda, India, Sweden, and the UK. *C. parvum* IIc has not been observed in livestock or companion animals, verifying its status as a likely anthroponotic subtype of *C. parvum* (Nader et al., 2019)*.* Due to its wide host range, *C. parvum* is more dominant in rural environments where the likelihood of contact with infected animals and contact with untreated water sources (e.g. streams, lakes, etc.) is higher (Cacciò & Widmer, 2013; Thompson & Smith, 2011). To highlight the difference that socioeconomic status can play in transmission of *Cryptosporidium* a review found that, on a yearly basis, there were approximately 6 million cases of foodborne transmission in the Eastern Mediterranean Region compared with 27 million in the African region (Thompson & Smith, 2011).

At the community level, it is easier to isolate the specific factors that increase the risk of *Cryptosporidium* infection. Improper food handling can increase the risk of contracting cryptosporidiosis. Low hygiene standards when processing meat or fish can increase the risk of infection for those handling the food and the consumers downstream. The ability of cattle to serve as reservoirs of *Cryptosporidium* has previously been described in this thesis, in addition, previous studies have found evidence of zoonotic subtypes *C. parvum* in various species of fish and molluscs (Gómez-Couso et al., 2004; Robertson et al., 2019). Deforestation and loss of biodiversity can also increase the risk of infection by bringing human communities in closer contact with wild animal reservoirs of the parasite

(Thompson, 2013; Thompson & Polley, 2014). Improper handling of human waste, i.e. inefficient waste processing plants or liberal disposal of human faecal matter in the environment, greatly increases the ability of this parasite to spread within a community (Omarova et al., 2018).

Currently, the methods most frequently employed for the diagnosis of cryptosporidiosis are microscopy and immunoassays. However, these methods often result in false negatives and do not specify the genotype of the parasite detected (Chalmers et al., 2011), which may have important implications for human health. Wider adoption of advanced molecular techniques such as PCR and DNA sequencing would give us better insight into the importance of particular transmission routes in specific communities and aid the development of effective strategies to reduce the risk of contracting cryptosporidiosis in our communities.

1.4.3 *Giardia* Infections in Humans

Giardia is sometimes described as the "ubiquitous parasite", a moniker that highlights its prevalence across the globe. The species of *Giardia* involved with most human infections is *G. intestinalis*, specifically assemblages A and B within this species. There are 8 assemblages within *G. intestinalis*, C and D show specificity for predominantly canid hosts, E bovine and porcine, F is associated with felids, G rodents, and assemblage H is found predominantly in pinnipeds (Feng & Xiao, 2011). However, contrary to our previous understanding of host specificity which presumed that infections in humans were exclusively due to assemblages A and B, observation of other assemblages in humans has increased in recent years (Cacciò et al., 2017). A review by Ryan et al., (2013) found that assemblage B is the most common subtype of *G. intestinalis* in most human infections (58% of cases) when compared to assemblage A (37% of cases). This proportion is not linked to the socioeconomic status of the country observed. In New Zealand assemblage B is the dominant subtype of *Giardia* found in human infections, with this assemblage identified in 79% of cases in New Zealand between 2009 and 2015, nevertheless, assemblages A-F have also been observed in the country (Garcia–R et al., 2017a). Clinical outcomes differ when a host is infected with either assemblage A or assemblage B. Previous studies suggest that infections with assemblage A are generally short and acute

whereas infections with assemblage B show a high frequency of transmission often leading to chronic infections and growth stunting in children (Thompson & Monis, 2011). Acute infections normally build up over 3 weeks with a peak at 8 days post-infection (Cernikova et al., 2018). Assemblages A and B have a wide range of hosts besides humans, including cattle, dogs, cats and ducks, all of which have close contact with humans compared to other animals (Luján & Svärd, 2011). This highlights the zoonotic potential of this parasite. Nonetheless, it has been found that *Giardia* infections in dogs and cats are usually asymptomatic (Thompson, Palmer, & O'Handley, 2008), and various studies suggest that humans are the main source of infection in these situations. This was observed in African wild dogs in the African continent, marsupials in Australia, beavers and coyotes in North America, muskoxen in the Canadian arctic, house mice on remote islands and marine mammals in various parts of the world (Thompson & Smith, 2011). Despite the apparent anthroponotic nature of the majority of *Giardia* infections, livestock still serves as an important reservoir of the parasite.

Looking at the seasonality of *Giardia* infections in New Zealand between 1997 and 2014 the number of cases appears relatively uniform when compared to that of cryptosporidiosis, with significant peaks in cases occurring from January-May in most years (Garcia–R et al., 2017a; Snel, Baker, Kamalesh, et al., 2009). This suggests that the dominant mode of transmission is anthroponotic because these are the months where increased use of recreational water sources is observed in the population (January-May corresponds to the summer to early autumn in New Zealand).

The most common symptoms of giardiasis are diarrhoea, nausea, weight loss, and vomiting (Einarsson et al., 2016a). Previous studies have attributed the signature symptom of giardiasis, diarrhoea, to a disruption of the electrolyte balance and increased permeability of the intestines caused by infection with the parasite. In addition to the aforementioned common symptoms of giardiasis, chronic infections can lead to severe malabsorption syndrome, irritable bowel disease (IBD), arthritis, food allergies, and lethargy (Bartelt & Sartor, 2015), and sometimes the symptoms can last long after the parasite is cleared from the host. Accordingly, the field would benefit from in-depth studies linking assemblages and sub-assemblages to the varied symptoms observed during giardiasis.

1.4.4 Transmission and Risk Factors of Giardiasis

The main transmission route for *Giardia* infections is through contaminated water, either drinking or recreational sources. Other routes include person to person transmission, contact with contaminated food, contact with contaminated animals and livestock and travel to endemic countries (Luján & Svärd, 2011). As mentioned before, children under the age of 5 and immunocompromised individuals have the highest risk of infection with *Giardia*, either as a primary or an opportunistic infection. The vulnerability of this group poses a significant risk to the population. A study in Auckland, New Zealand found that housewives and nursing mothers stood at a higher risk of infection compared with other occupational groups, largely due to the risk of contracting the disease while changing nappies (Hoque et al., 2001). Similar to *Cryptosporidium* oocysts, *Giardia* cysts are resistant to ozonolysis and disinfecting solvents such as chlorine and bleach, so low water purification standards put a population at high risk of contracting the disease. Interestingly, giardiasis acquired the alias "beaver fever" due to an outbreak that was caused in Canada when hikers drank from a stream contaminated with *Giardia* from beavers in Banff National Park (Wiser, 2010), further highlighting the risk of contracting giardiasis from natural water sources.

Due to the lack of consensus with regards to the taxonomy of *Giardia*, and the fact that assemblages of *G. intestinalis* are alternatively referred to as distinct species, determining definitive links between subtypes of *Giardia* and specific geographic locations or socioeconomic status is problematic. The field would benefit from unanimity vis-à-vis the classification of the various species and subtypes of *Giardia*. Also, the application of advanced molecular techniques, such as PCR, next-generation sequencing and whole genome sequencing (WGS), when diagnosing infection would help us better understand the relative burden of zoonotic and anthroponotic transmission of this enteric parasite. This will give us a better understanding of the epidemiology of the disease and inform public health strategies to combat giardiasis. A recent study used WGS to show that surface water plays a pivotal role in the transmission of giardiasis from wildlife to humans (Tsui et al., 2018).

1.5 Outbreaks

Many countries report multiple outbreaks of cryptosporidiosis and giardiasis each year, and the data gathered from these events has greatly developed our understanding of the epidemiology of these diseases. A study looking at worldwide waterborne outbreaks attributed to parasitic protozoa between 2011 and 2016 found that 239 outbreaks of cryptosporidiosis and 142 of giardiasis were reported during that time frame (Efstratiou et al., 2017). It should be noted that due to insufficient surveillance reporting mechanisms in some countries and the fact that only waterborne outbreaks were recorded in this study, the total number of outbreaks due to both parasites was probably significantly higher than the reported figures. This presents a significant burden to public health institutions. Interestingly, the source of 14% of the outbreaks reported in the study was contaminated recreational water, a phenomenon that is replicated when analysing the local data from New Zealand (Lake et al., 2008). Though waterborne outbreaks of cryptosporidiosis and giardiasis appear to be the most common form of outbreaks, the sources of the contamination can be varied. This could be due to contamination by infected wild animals, like in the previously mentioned outbreak in Banff National Park (Wiser, 2010), contamination of waterways by infected livestock, or recreational water by infected humans. Apart from the usual diagnostics used to confirm cases during outbreaks, genotyping is important when trying to determine the initial source of contamination. Compared to waterborne outbreaks there is less data on foodborne outbreaks of cryptosporidiosis and giardiasis, but they do occur frequently and can have a high economic cost (Budu-Amoako et al., 2011; Dixon, 2021).

A study of foodborne outbreaks in Europe in 2016 found that 0.4% of the total (4786 outbreaks) were due to *Cryptosporidium, Giardia,* and *Trichinella* (Rousseau et al., 2018), with contaminated fruits and vegetables being the most relevant sources of infection. Due to the incredible consumption of water in the agricultural industry, contaminated water is a significant source of contamination of agricultural produce resulting in foodborne outbreaks (U. Ryan et al., 2019). Between 1971 and 2011, 38 foodborne outbreaks of giardiasis were reported in the USA (E. A. Adam et al., 2016). Notably, fresh produce was

implicated in most of the outbreaks. The low infective dose of cysts (10 cysts) required to initiate infection (Leggett et al., 2012), and the fact that *Giardia* survives in the temperatures and humidity recommended for the storage of most leafy produce (4 °C and a relative humidity of 98-100% for the storage of lettuce), means there is an increased risk of contracting *Giardia* through the handling and consumption of fresh produce relative to other types of food (Dixon, 2021; Lutz & Hardenburg, 1968). The increasing globalisation of food distribution, which in turn increases the risk and scale of outbreaks, suggests that public health bodies should come up with new and effective strategies for the prevention of foodborne transmission of giardiasis and cryptosporidiosis. One factor hampering efforts to capture foodborne outbreaks of these diseases is the difficulty and expense of detecting *Cryptosporidium* and *Giardia* on contaminated food. The current standard for diagnosing contaminated food and water is through immunomagnetic separation followed by PCR, which is prohibitively expensive in LMICs (McAuliffe et al., 2017; U. Ryan et al., 2019). Future research should look into developing diagnostic assays capable of detecting small numbers of *Cryptosporidium* and *Giardia* (oo)cysts while being economic and efficient.

In New Zealand outbreaks of cryptosporidiosis and giardiasis are of particular public health significance. Surveillance data from 2017 implicated enteric agents in 88% of the total outbreaks that occurred in New Zealand (Institute of Environmental Science and Research Ltd (ESR), 2018b). During that year there were 27 outbreaks of cryptosporidiosis and 24 of giardiasis, accounting for 4.2% and 3.7% of total outbreaks respectively, and the most commonly reported mode of transmission was human-human. Outbreaks can involve multiple infectious agents, as was observed in Auckland in 2017 where both *Cryptosporidium* and *Giardia* were detected. A study conducted in the Hopkirk Research Facility looking at samples from an outbreak of campylobacteriosis in Hawke's Bay in 2016 detected the presence of *Cryptosporidium* and *Giardia* in a significant proportion of the samples, microscopically and using PCR. This suggests that mixed infections during outbreaks are not being adequately captured and the burden of these diseases in New Zealand, and indeed globally, might be higher than previously estimated. To date, epidemiologically linked cases of *Cryptosporidium* and *Giardia* have not been attributable to particular genotypes. A patient with an infection may carry multiple subtypes of the same infectious agent, the outcome of the competitive

interactions between them has an effect on the disease phenotype, which, in turn, affects the efficacy of treatment (Thompson & Smith, 2011). For this reason, understanding the within-host genetic diversity of a pathogen is essential for effective disease management. In New Zealand, questions remain as to whether epidemiologically linked cases were all part of the same event if there is genetic variation, or if they represent within- and between-host diversity (Garcia–R et al., 2017a). A possible reason for this could be a lack of resolution due to the standard detection methods used in New Zealand i.e., epidemiological analysis is limited by what immunoassays, conventional PCR and Sanger sequencing can detect. These limitations affect the surveillance of these diseases and will be discussed further in chapters 2 & 4.

1.6 *In vitro* culture methods

Since the discovery of *Cryptosporidium* and *Giardia,* significant developments have been made in the culture of these parasites. Despite the success had with the *in vivo* culture of these parasites using systems such as mice, suckling pigs and calves, less success has been observed in *in vitro* culture. A lack of efficient methods for the *in vitro* culture of these parasites has prevented a better understanding of the factors influencing the infectivity of the various subtypes of *Cryptosporidium* and *Giardia* and hampered the development of drugs and vaccines against the diseases they cause. The development of cell-free culture methods for the culture of these protozoans, such as Keister's Modified TYI-S-33 medium used to culture *Giardia* (Keister, 1983), has shown some success. However, this review will focus on cell culture methods.

The application of human cell culture systems for the culture of *Cryptosporidium* and *Giardia* has the potential to give us a better understanding of host-pathogen interactions when compared to cell-free methods. Human intestinal cell lines such as Caco-2 and HCT-8 have allowed for the successful culture of *G. intestinalis* (Luján & Svärd, 2011). However, less success has been seen when culturing *Cryptosporidium* due to the inability of the parasite to progress through its full life cycle in these systems (only asexual intracellular life cycle stages have been observed in *in vitro* culture) (Cacciò & Widmer, 2013). This led

to the development of alternatives such as the hollow fibre system developed by Morada et al., (2016) which they say provides an environment that mimics the gut allowing long-term culture (>6 months) and an increased yield of oocysts (1×10^8 oocysts ml^{-1} day^{-1}). Recently, the human oesophageal carcinoma cell line, COLO-680N, was characterised and displayed successful long-term cultivation of *Cryptosporidium* (Bones et al., 2019; Miller et al., 2018)*.

The ability to purify clean (oo)cysts from faecal samples, while reducing the level of contamination due to bacteria and particulate matter, has a great effect on the success of *in vitro* culture of *Cryptosporidium* and *Giardia*. In this respect, the purification of *Cryptosporidium* oocysts from faecal samples has been more successful when compared to the purification of *Giardia* cysts. Various floatation methods have been developed for the purification of oocysts from faecal samples. A particularly successful technique is the Ficoll floatation method, which results in abundant and exceptionally clean samples of oocysts suspended in phosphate buffer saline (PBS) (Meloni & Thompson, 1996). Purification of large numbers of *Giardia* cysts from faecal samples is relatively more difficult. There are salt-based and sucrose-based floatation techniques, but the highest yield from the best of these is only 5.1×10^4 cysts from 2 grams of faeces (significantly more is needed for infectivity studies), and even then, there are still significant levels of debris present in the resulting solution (Afshin et al., 2011). This highlights the need for the development of better methods for the purification of cysts from faecal matter, which, if successful, could provide more insights into the biological processes, host-pathogen, and competitive interactions of this parasite.

1.7 Conclusion

Cryptosporidium and *Giardia* are protozoan parasites of global public health significance. They typically cause severe, self-limiting diarrhoea in humans but are particularly lethal to the immunocompromised and children under the age of five. *Cryptosporidium parvum* and *Giardia intestinalis* have a wide range of hosts, which means there is a high risk of zoonotic transmission of the parasites from animals to humans. In New Zealand, and in

most countries that report, multiple outbreaks of cryptosporidiosis and giardiasis occur each year, the largest proportion is attributed to waterborne outbreaks. Great advances have been made in the understanding of the genetic diversity of these parasites at the global and local scale, giving us a better understanding of the epidemiology of these diseases and allowing the researcher to track the dominant sources of transmission. Cryptosporidiosis and giardiasis place a significant burden on the public health system in New Zealand, but the adoption of advanced techniques for the diagnosis and genotyping of these parasites would lead to a better understanding of the spread of these parasites in the country. The lack of affordable and efficient methods for the *in vitro* culture of *Cryptosporidium* and *Giardia* acts as a roadblock to the development of therapies and the better understanding of the biological processes and host-pathogen interactions of these parasites. Further work needs to be carried out in this respect to aid in the development of strategies to combat this enteric duo.

1.8 Structure of Thesis

The purpose of this thesis was to use NGS to gain new insights into the epidemiology of cryptosporidiosis and giardiasis in New Zealand and to develop an *in vitro* system for the analysis of *Cryptosporidium* infection in human cell lines. This chapter has provided an introduction to the literature upon which this study is built.

Initially, one of the aims of this thesis was to investigate the infectivity of *Giardia* using *in vitro* techniques. However, limitations were encountered, particularly with regards to the purification of cysts from faecal samples. Current methods produce low yield samples with high levels of contaminating debris. For this reason, Chapter 2 focused on developing a new method for the purification of cysts from faecal samples.

In Chapter 3 human faecal samples from outbreaks of giardiasis that occurred in New Zealand between 2010 and 2018 are analysed using Sanger and next-generation sequencing. This was done to highlight the ability of NGS to discriminate between multiple sequences within one sample, which makes it a superior tool for improving the

understanding of the epidemiology of cryptosporidiosis, and to gain a better understanding of the genetic diversity of *Giardia* present in New Zealand.

The investigation in Chapter 4 is similar to that done in Chapter 2, except in this case the same techniques were carried out on samples from outbreaks of cryptosporidiosis.

Chapter 5 investigates the infectivity of *Cryptosporidium*. Previous studies have found that symptoms in humans vary depending on the species of *Cryptosporidium* they are infected with. The development of a versatile *in vitro* system for the culture and manipulation of the parasite would advance the understanding of the mechanisms underpinning the differences in infectivity observed between species. In addition, it provides a tool that can be used in the development of new therapies. Successful culture of the full life cycle of *Cryptosporidium* was accomplished using the COLO-680N cell line. In this study flow cytometry was utilised to distinguish between infected and uninfected populations of cells within the same samples, with and without the use of a fluorescent antibody.

Chapter 6 builds upon the results of Chapter 5 by using NanoString to examine the differences in gene expression between species and subtypes that might account for their differences in infectivity. A review of current literature identified 48 genes highly expressed intracellularly and 48 highly expressed extracellularly, also, twelve potential drug targets were identified and included in the panel of genes. Gene expression was assessed in infected cells and sporozoite samples.

Chapter 7 provides a general discussion that ties together the aims and findings of all investigations in this thesis.

1.9 Research Questions

Chapter 3: Uncovering the genetic diversity of *Giardia* isolates from outbreaks in New Zealand.

- Are there common genotypes linking outbreaks of giardiasis that cannot be identified by consensus sequencing?
- What subtypes of *Giardia* are common to human hosts in New Zealand?

Chapter 4: Capturing genotype diversity in outbreaks of Cryptosporidiosis occurring in New Zealand from 2010 to 2018.

- Are there common genotypes linking outbreaks of cryptosporidiosis that cannot be identified by consensus sequencing?
- What genotypes are common to human hosts in New Zealand?

Chapter 5: An *in vitro* platform for studying the inter-species infectivity of *Cryptosporidium* spp.

- Are the differences in symptoms observed between patients infected with *Cryptosporidium* due to differences in infectivity between species/subtypes of the parasite?
- Are host-pathogen interactions determined by genetic differences at the between-species level and ecological differences at the between genotype level?

Chapter 6: Comparative gene expression analysis of *Cryptosporidium* species

- Are there differences in gene expression between species/subtypes that account for their differences in infectivity?

1.10 Understanding *Cryptosporidium* glycoprotein 60 and *Giardia intestinalis* nomenclature

1.10.1 *Cryptosporidium* glycoprotein 60 nomenclature

For example, IIaA15G2R1:

- The first two letters of a subtype denote whether it is *C. hominis* (I) or *C. parvum* (II).
- Followed by a lower case letter which indicates which of the ten allele families (five *C. hominis* and five *C. parvum*) the subtype belongs to.
- An uppercase A, G, or T denotes the tri-nucleotide TCA, TCG or TCT which code for the amino acid serine, followed by a number indicating the number of respective tri-nucleotide repeats.
- An R1 or R2 indicates whether one or two copies of the sequence ACATCA immediately following the tri-nucleotide repeats

1.10.2 *Giardia intestinalis* nomenclature

Giardia intestinalis consists of eight distinct genotypes or assemblages designated A-H. Assemblages A and B can infect humans and other mammals but assemblages C-H appear to be host-specific. One of the most common genetic loci for *Giardia intestinalis* typing analysis is the *gdh* (glutamate dehydrogenase) gene. The differences between the gene sequences produce subgroups within the assemblages, e.g., AI, AII, AIII and BI, BII, BIII, BIV etc.

2 High-yield purification of *Giardia intestinalis* cysts from faecal samples

2.1 Abstract

Giardia is an enteric protozoan parasite that causes gastroenteritis in all classes of vertebrates. It is globally ranked among the leading causes of death in children under 5 years of age. Giardiasis affects approximately 280 million people worldwide annually, a situation exacerbated by the low availability of effective treatments and the lack of a

vaccine. In addition, the parasite is difficult to manipulate in *in vitro* environments, which hampers the development of effective disease management strategies. This chapter outlines the development of a method for the purification of viable *Giardia* cysts from faecal samples, verified by a trypan blue dye exclusion test. This protocol produces a 10-fold increase in yield over current methods. By combining sucrose flotation with gated filtration, the protocol significantly reduces the amount of debris in the purified cysts suspension. Cyst viability is verified by a trypan blue dye exclusion test. The ability to purify large quantities of *Giardia* from faecal samples could advance the development of effective treatments to target this worldwide prevalent parasite. The published version of this paper is attached in Appendix A.

2.2 Introduction

Giardia intestinalis is a protozoan parasite that has the ability to infect the epithelial cells of the gastrointestinal tract in all classes of vertebrates. *Giardia* is one of the leading causes of diarrhoea worldwide and can be lethal to immunocompromised individuals and children under 5 years of age (Dib et al., 2008; Saaed & Ongerth, 2019). The main mode of transmission of this parasite in humans and other animals is through contaminated food and water (Efstratiou et al., 2017; U. Ryan et al., 2019). Transmission among people is aided by the ability of the environmental stage of *Giardia*, the cyst, to resist chlorine and bleach, two of the most common chemicals used in water purification (Jarroll & Hoff, 1988; Kim et al., 2001; Winiecka-Krusnell & Linder, 1998).

Giardia begins its life cycle as a cyst. The cyst is the environmental/transmissible stage of this parasite and it possesses a hardy wall made up of a mesh of cyst wall proteins complexed to a singular sugar polymer of (β1–3)-linked N-acetylglucosamine (GalNAc) (Samuelson & Robbins, 2011), which makes it tremendously stable in cool and moist conditions, resistant to most solvents and able to survive in the environment for several months in cold conditions ($\sim$10 °C) (Bingham et al., 1979; Jarroll & Hoff, 1988; Rovid Spickler, 2005). The infective dose required to initiate infection in humans is approximately 10 – 100 cysts (Leggett et al., 2012). Acute infections normally

build up over 3 weeks, with a peak at 8 days post-infection (Cernikova et al., 2018). According to current understanding, *Giardia* cysts come in two varieties: type 1 cysts which, when viewed with light microscopy, appear as bright ovals with the cell body uniformly distributed within a clearly visible cell wall; type 2 cysts which are ovoid in shape but appear darker when using a light microscope, and the cell body appears fully or partially detached from the clearly visible cell wall. Type 1 cysts are considered to be more viable and better for *in-vitro* excystation than type 2 cysts (Gillin et al., 1989; Luján & Svärd, 2011).

Current methods of detection of *Giardia* include microscopy, immunoassays, and molecular assays (Adeyemo et al., 2018). However, the inability to purify significant numbers of cysts from faeces has hindered the progress of *in vitro* studies looking at the molecular mechanisms of this parasite. This study details a new method for the purification of *Giardia* cysts from faeces, which builds on past methods (Afshin et al., 2011; Walderich et al., 1997) and gives a 10-fold increase while reducing the amount of debris in the sample. This protocol for purifying *Giardia* cysts from faecal samples involves the use of sucrose flotation to separate cysts from faecal matter, followed by gated filtration to purify the cysts from any remaining debris and a dye-exclusion test to verify the viability of the purified cysts. The efficacy of the protocol outlined herein is tested against the methods out lined by Afshin et al., (2011) and Walderich et al., (1997).

2.3 Methods

Faecal samples from symptomatic patients across New Zealand are delivered to accredited national diagnostic laboratories. Those found to be positive for *Giardia* were sent to the Hopkirk Research Institute, Massey University, New Zealand. Here the samples were genotyped at the glutamate dehydrogenase (*GDH*) locus by PCR and further characterised by Sanger sequencing. Samples chosen for cyst purification were picked based on positive PCR results.

In a 50 ml centrifuge tube, 1 g of faeces was dissolved in 20 ml of refrigerated 0.1 M phosphate-buffered saline, pH 7.3 (PBS). The solution was mixed thoroughly using a vortex mixer. The suspension was then passed through a 40 μm cell strainer and the filtrate collected in a new 50 ml centrifuge tube to get rid of larger particles. The filtrate was washed once by centrifuging at $500 \times$ g, 4 ° C, for 5 min with the brakes off to reduce disturbance of the pellet. Using a 5 ml pipette the supernatant was discarded leaving 5 ml, then pellet was resuspended in 0.5% Tween 80 up to 20 ml. The suspension was centrifuged at $500 \times$ g, 4 ° C, for 5 min. The supernatant was discarded leaving 5 ml then the pellet was washed 2 more times in 0.5% Tween 80 at the aforementioned settings. After the third wash, the pellet was resuspended in dH_2O up to 20 ml and washed 2 more times to remove any remnants of Tween 80.

After the second wash, the pellet was resuspended in dH_2O up to 20 ml, then using a 5 ml pipette 20 ml of refrigerated 1.5 M sucrose was layered under the suspension taking care to minimise disturbance at the interface. The floatation was centrifuged at $1,300 \times$ g, 4 ° C, for 10 min with the brakes off. Following centrifugation, 15 ml of the mediated phase was carefully collected and transferred to a fresh 50 ml centrifuge tube, an equal volume of dH_2O added and the tube vortexed thoroughly to mix the sample. Then the solution was washed 3 times in dH_2O by centrifugation at $500 \times$ g, 4 ° C, for 5 min. After the last wash, the pellet was resuspended in dH_2O up to 20 ml and an equal volume of 0.85 M sucrose was carefully layered underneath, then the floatation was centrifuged at $1,600 \times$ g, 4 ° C, for 5 min with the brakes off. Following centrifugation, the cysts settle down at the bottom of the tube and the lighter debris is trapped at the mediated phase, therefore, most of the contents of the floatation were discarded leaving 5 ml at the bottom. The sample was topped up to 20 ml with dH_2O then washed 3 times at $500 \times$ g, 4 ° C, for 5 min.

After the last wash, the sample was topped up to 10 ml with refrigerated PBS, 0.2 g of 0.1 mm zirconia/silica beads was added to the solution then vortexed thoroughly for 1 min. Immediately after being vortexed, the sample was passed through a 20 μm pluriSelect cell strainer and the connector ring (*PluriSelect - The Cell Separation Company*, n.d.) inserted into a 50 ml centrifuge tube. By attaching a syringe to the connector ring and pulling the piston, low pressure was created to encourage the straining of the sample through the filter (Figure 2.1.). The filtrate was collected and the filter, now containing

the zirconia/silica beads and other debris, was discarded. The filtrate was vortexed briefly then passed through a 5 μm pluriSelect cell strainer and connector ring inserted into a fresh 50 ml centrifuge tube. A syringe was attached to the connector ring and utilised as previously described. After this step, the cysts collect on the 5 μm cell strainer and smaller matter are discarded in the filtrate. By reversing the cell strainer onto a new 50 ml centrifuge tube the cysts were washed off using 15 – 20 ml of cold dH$_2$O.

To assist the washing off and increase efficiency of recovery, a connector ring was attached to a clean empty 50 ml centrifuge tube. This construct was attached on the top of the reversed cell strainer. By attaching a fully extended syringe to the connector ring and pushing on the piston, high pressure can be created to support the washing off of the cysts by the dH$_2$O, greatly increasing yield and efficiency (Figure 2.2.). The filtrate with the cysts was then spun at 2000 × g for 8 min, reduced to 1 ml, resuspended, and transferred to a clean 1.5 ml microcentrifuge tube. 10 μl of an antibiotic solution consisting of 100 U/ml penicillin, 100 μg/ml streptomycin and 250 ng/ml amphotericin B was added to the sample. Cysts were counted using a haemocytometer, then stored at 4 ° C until further use. An aliquot of the cysts suspension was diluted 1:1 in trypan blue stain and mounted onto a haemocytometer to count type I and type II cysts to estimate the viability of the purified cysts (Gillin et al., 1989; Luján & Svärd, 2011). The cysts purification method outlined here tested against the methods outlined by Afshin et al., (2011) and Walderich et al., (1997). All three methods were tested on each of three samples for the sake of comparison.
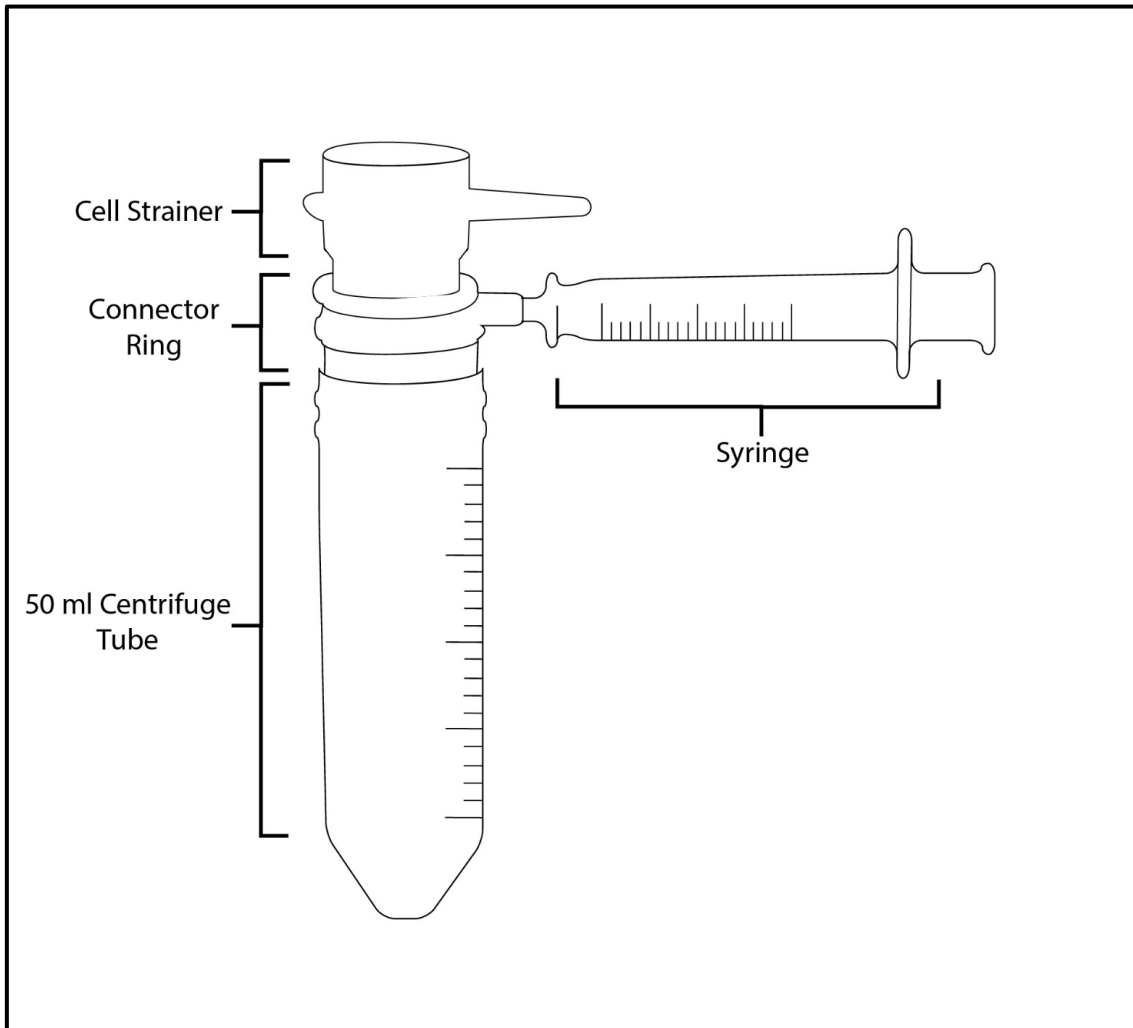
Figure 2.1. Schematic of filtration setup. The cell strainer is attached to a connector ring, which is in turn attached to the top of a 50 ml centrifuge tube. A syringe attached to the connector ring is used to create region of low pressure that assists and speeds up the filtration through the cell strainer.
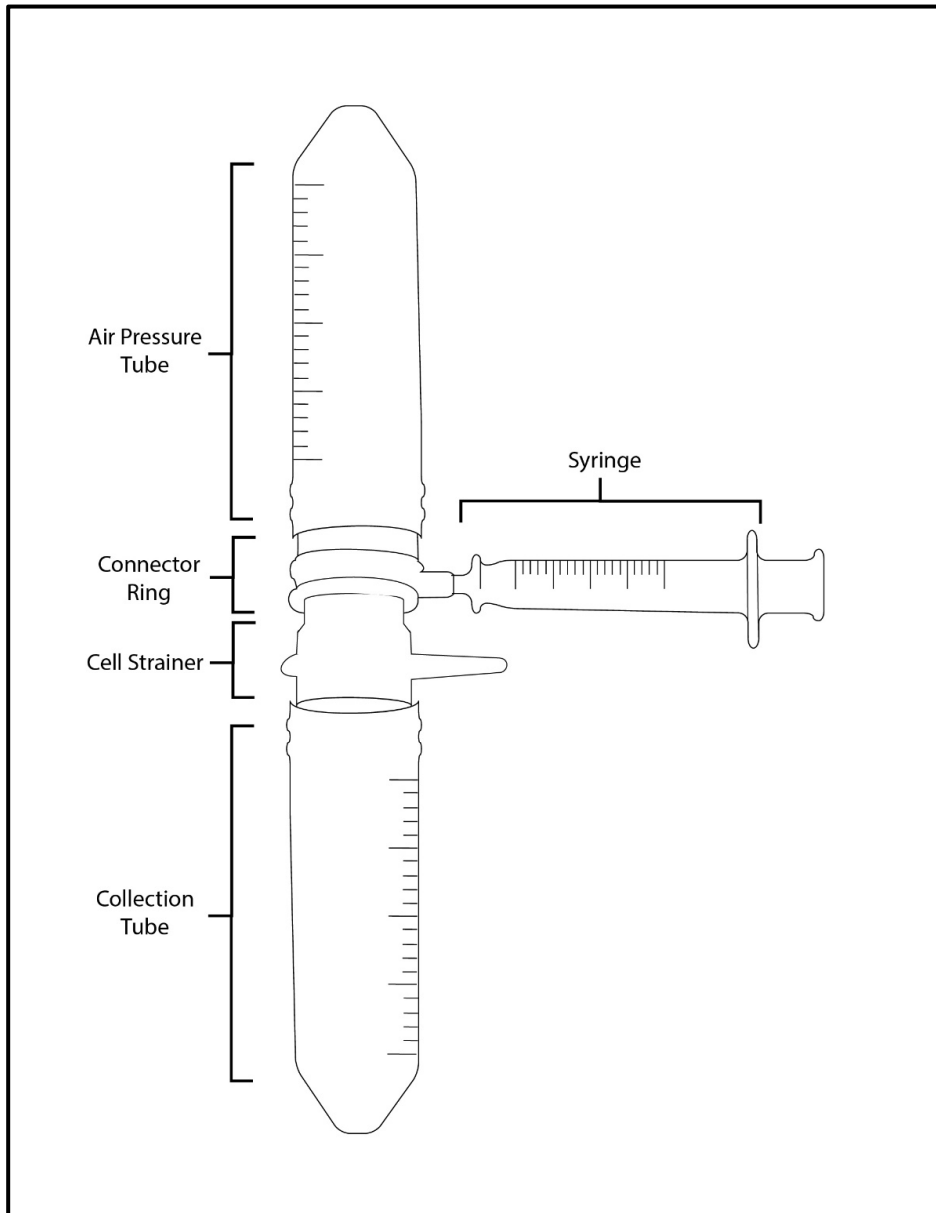
Figure 2.2. High pressure-assisted washing. After filtration the 5 μm cell strainer is flipped over onto a fresh 50 ml centrifuge tube to wash off the collected cysts. An empty centrifuge tube with a connector ring and syringe attached to it is used to create an area of high pressure to support the washing off of the cysts by the H_2O.

The method employed in this study was based off a method developed by Afshin et al., (2011). Briefly, 1g of faecal matter was dissolved in 0.5% v/v Tween 80 solution. The solution was centrifuged at $500 \times g$ for 5 min at RTP. This washing process was repeated until the surface liquids became clear. The supernatant was discarded, the pellet resuspended in distilled water, then carefully poured onto an equal amount of 1.5M

sucrose solution. This gradient was centrifuged for 10 min at 1,300 × g, 4°C. Contents of the mediated phase was collected and wash three times in distilled water by centrifugation at 500 × g for 5 mins at RTP. Then the supernatant was discarded, pellet resuspended in distilled water, and solution poured over an equal amount of 0.85M sucrose solution. This gradient was centrifuged at for 10 min at 1,600 × g, 4°C. The cysts settle at the bottom of the gradient, were collected, and washed three times in distilled water.

Another method for cysts purification employed in this study was that developed by Walderich et al., (1997). It is like the method outlined by Afshin et al., (2011), except the faecal matter was washed in 0.2M PBS instead of Tween 80, the 1.5M sucrose floatation was centrifuged at 1,700 × g for 10 min, the final floatation used a 0.75M sucrose solution and was centrifuged at 1,700 × g for 10 min.

2.4 Results and Discussion

Table 2.1. A comparison of *Giardia* cyst purification methods. The novel method outlined in this study (Ogbuigwe) was compared to the cyst purification methods by Afshin et al., (2011) and Walderich et al., (1997). The test was carried out on three separate human faecal samples from patients diagnosed with giardiasis.

	Method	Afshin	Walderich	Ogbuigwe
Sample	**Age of Sample (Days)**	**Cyst Yield ($\times 10^4$)**	**Cyst Yield ($\times 10^4$)**	**Cyst Yield ($\times 10^4$)**
1	58	0.3	1	11
2	58	0.8	10	15
3	58	0.5	1.7	13
Average		**0.5**	**4.2**	**13**
Standard Deviation		**0.25**	**5.01**	**2.00**

Current methods for the purification of *Giardia* cysts from faeces include sucrose and Percoll-based flotation (Afshin et al., 2011; Alvarado & Wasserman, 2006; Sauch, 1984; Walderich et al., 1997). Here, the two-phase sucrose flotation method described by Afshin

et al., (2011) is modified, using cell strainers and pressure-assisted filtration to obtain a 10-fold increase in the yield of cysts based on comparison of the aforementioned method. A comparison of the cyst purification methods outlined by Afshin et al., (2011), Walderich et al., (1997), and that outlined in this study, tested on three samples, is shown in table 2.1. Furthermore, the application of zirconia/silica beads functioned to reduce the amount of contaminants in the sample by separating any debris that coagulated in the suspension, thereby making the subsequent filtration steps more efficient (Figure 2.3). Table 2.2. shows the result of repeated testing of the method outlined in this study on 11 separate samples.

Table 2.2. Results of testing of new cyst purification method on multiple replicates. The novel method outlined in this study (Ogbuigwe) was tested in multiple samples to test efficacy.

Sample	Age of Sample (Days)	Cyst Yield ($\times 10^5$)
1	62	1.7
2	62	1.1
3	34	1.2
4	8	0
5	96	1.4
6	97	1.5
7	97	0
8	197	0.7
9	197	1.3
10	148	0.4
11	148	0
Average	**104**	**0.85**
Standard Deviation		**0.65**
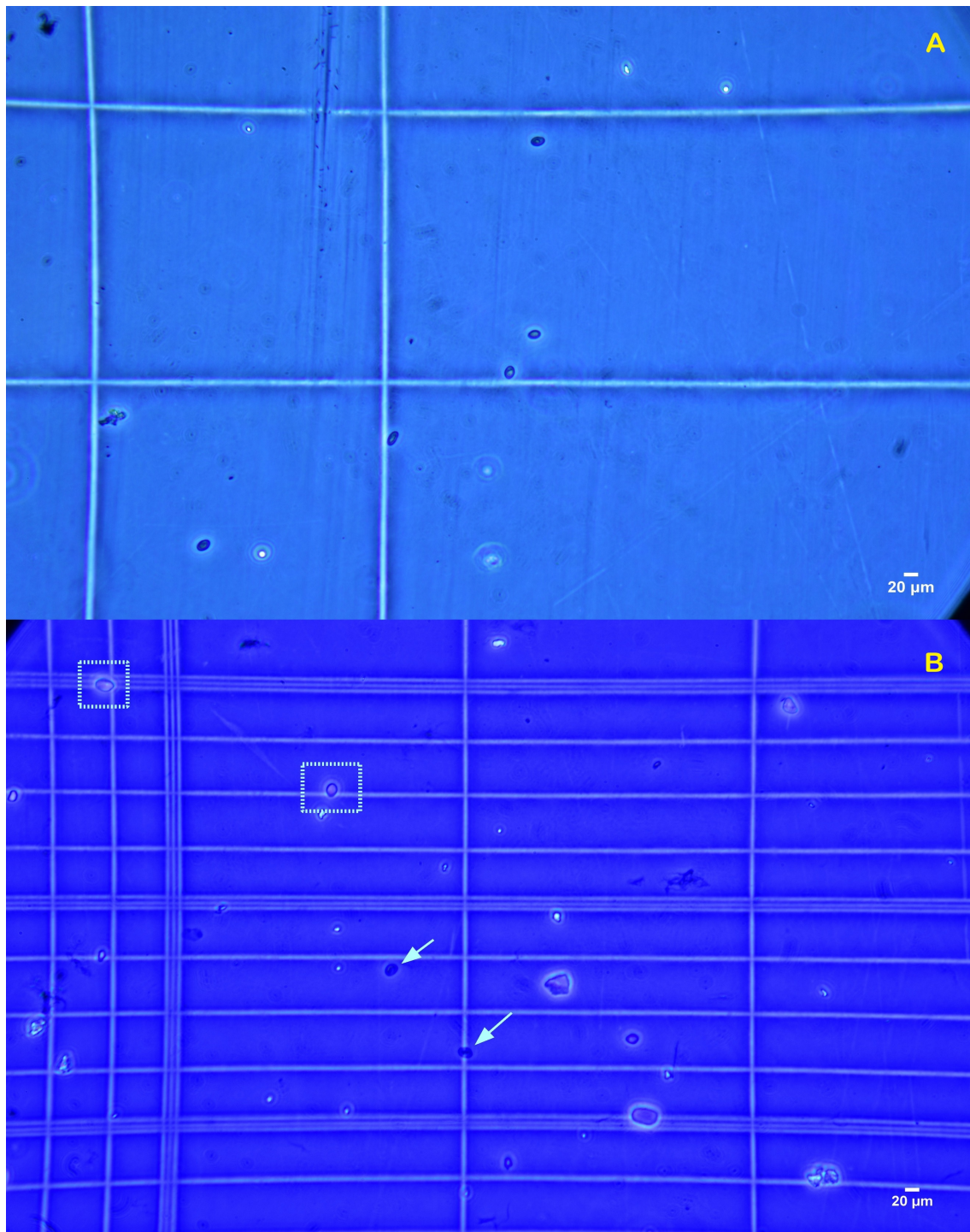
Figure 2.3. Brightfield microscopy images of purified *Giardia intestinalis* cysts. [A] 20x magnification of purified cysts showing little debris. [B] trypan blue staining of purified cysts to show presence of viable cysts. Type 1 cysts (highlighted by dashed boxes) appear as bright ovals and the cell body is uniformly distributed within the cyst wall (the cyst wall is clearly visible), these are trypan-blue negative (white). In type 2 cysts (highlighted by arrows) the main body of the cell appears to be fully or partially detached from the cyst wall, these are trypan-blue positive (blue).

On average, *Giardia* cysts have a length of 11 – 14 μm and a width of 7 – 10 μm (Luján & Svärd, 2011). By using the cell strainers to filter matter larger and smaller than the dimensions of the cysts, it was possible to significantly reduce the amount of debris and contaminants in the sample. The standard method for the identification of *Giardia* cysts in faeces has been the ether sedimentation technique, which relied on the ability of the diagnostician to differentiate cysts from faecal matter and other microbes present in the sample (Hooshyar et al., 2019). This method requires two or more examinations of faeces from the same individual for reliable diagnosis. As a result of this, and the fact that cyst shedding varies during the course of an infection, it has been difficult to estimate the average load of cysts in infected individuals. The procedure outlined in this protocol could provide a method for acquiring accurate estimates of the cyst load in infected individuals. Furthermore, our method does not require any stains or fluorescent microscopes to identify *Giardia* cysts. The lack of debris in the sample makes it reliable, easy and efficient to identify cysts based on their morphological characteristics with little training or prior experience needed. The purification of large numbers of cysts from faecal samples could give us a better understanding of host-parasite interactions and advance the manipulation and understanding of the molecular biology of this parasite.

In conclusion, the original method achieves a recovery rate of 1.5×10^4 cysts from 2 grams of faeces (Afshin et al., 2011), and a previous one, by Walderich et al., (1997), achieves a return rate of approximately 5×10^4 cysts from 2 grams of faeces. With the method outlined herein, it is possible to reliably recover between $1 – 1.5 \times 10^5$ cysts from 1 gram of faeces, and the dye exclusion test shows that approximately 50% of the cysts in the suspension were viable.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL
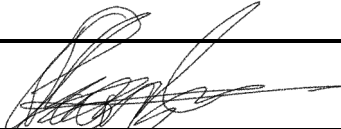
STATEMENT OF CONTRIBUTION

DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Paul Chiamaka Ogbuigwe
Name/title of Primary Supervisor:	Professor David Hayman

Name of Research Output and full reference:	
Ogbuigwe, P., Pita, A. B., Knox, M. A., Velathanthiri, N., & Hayman, D. T. S. (2020). High-yield purification of *Giardia intestinalis* cysts from faecal samples, in press at Current Protocols in Microbiology, 59, e117. doi: 10.1002/cpmc.177	

In which Chapter is the Manuscript /Published work:	2

Please indicate:	
The percentage of the manuscript/Published Work that was contributed by the candidate:	
and	
Describe the contribution that the candidate has made to the Manuscript/Published Work:	

P Ogbuigwe did 80% of the lab work, compiled the draft of the manuscript and designed the figures.

For manuscripts intended for publication please indicate target journal:

37

Candidate's Signature:	
Date:	08/07/2021
Primary Supervisor's Signature:	
Date:	08/07/2021

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

3 Uncovering the genetic diversity of *Giardia* isolates from outbreaks in New Zealand

3.1 Abstract

Giardia is one of the most common causes of diarrhoea in the world and is a notifiable disease in New Zealand. Recent advances in molecular techniques, such as PCR and Sanger sequencing, have greatly improved our understanding of the taxonomic classification and epidemiology of this parasite. However, there has been an inability to identify shared subtypes between samples from the same epidemiologically linked cases, due to samples showing multiple dominant subtypes within the same outbreak when characterised using Sanger sequencing. Here, NGS was employed to uncover the genetic diversity within samples from sporadic and outbreak cases of giardiasis that occurred in New Zealand between 2010 and 2018. This strategy exposed the significant diversity of subtypes of *Giardia* present in each sample. The utilisation of NGS and metabarcoding at the glutamate dehydrogenase (*gdh*) locus enabled the identification of shared subtypes between samples from shared outbreaks, providing a better understanding of the epidemiology of outbreaks of giardiasis in New Zealand.

3.2 Introduction

Giardia is an enteric protozoan parasite with the distinction of being among the most common causes of diarrhoea in humans and farm animals worldwide (Cacciò & Sprong, 2011). *Giardia* infects the epithelial cells of the gastrointestinal tract causing self-limiting diarrhoea in all classes of vertebrates. This parasite transmits via the faecal-oral route, and in humans particularly, contact with contaminated water sources is the dominant mode of infection and cause of outbreaks.

Approximately 280 million people are infected with this parasite every year, and the prevalence of infections in humans ranges between 0.4% to 7.5% in high-income countries, and 8% to 30% in low-/middle-income countries (Einarsson et al., 2016b; Feng

& Xiao, 2011). However, the disease can be fatal in immunocompromised individuals and ranks among the leading causes of death in children under the age of 5 (Luján & Svärd, 2011). This is why giardiasis, the disease for which *Giardia* is the causative agent, was recognised by the Word Health Organisation (WHO) in its neglected diseases initiative, highlighting the public health significance of this parasite (Savioli et al., 2006a). Since then, reporting of this parasite has improved in many countries. Further exacerbating the burden of this disease is the lack of any effective vaccines against the pathogens.

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

The clinical effects of giardiasis vary among individuals, ranging from asymptomatic carriage to severe malabsorption syndrome in some acute cases (Luján & Svärd, 2011). However, the mechanisms underlying the differences in phenotypes within these diseases are poorly understood. Previous studies suggest that differences in infectivity exist between assemblages. Experimental observations found that human volunteers inoculated with assemblage B were more likely to succumb to infection and develop

symptoms than those inoculated with assemblage A (Cacciò et al., 2017). Nevertheless, studies looking at the correlation between symptoms and assemblages have produced contradictory results. The ability to link phenotypic features with assemblages would greatly increase our understanding of transmission patterns. Failing that, an investigation of the genetic structure at the population level is essential for the proper inference of the transmission patterns of *Giardia* and its epidemiology.

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

Questions remain as to whether epidemiologically linked cases in New Zealand were all part of the same events or if they represent within- and between-host diversity (Garcia–R et al., 2017a). A possible reason for this could be a lack of resolution due to the standard detection methods used. Over the years the methods for the detection and classification of *Giardia* have progressed from microscopic analysis of physical characteristics to molecular tools such as PCR and Sanger sequencing of notable genes like *bg, gdh, tpi* and

SSU rRNA genes. However, because Sanger sequencing combines the contribution of all DNA fragments present in the reaction mixture, even this may lack sufficient resolution where mixed assemblages are present. PCR amplification of the *gdh* gene will amplify sequences from any *Giardia* assemblage that is present in the extracted DNA, which can lead to a mixed signal in the resulting Sanger sequence or failure to detect rare assemblage types. These limitations affect disease surveillance and make it difficult to capture within-host diversity. In contrast to Sanger sequencing, next-generation sequencing (NGS) techniques like amplicon-based sequencing allow millions of fragments to be sequenced in a single run allowing the researcher to separate the signal originating from each target molecule, thus allowing the efficient isolation, detection and quantification of rare types. In recent years, researchers have applied NGS techniques to study the epidemiology of cryptosporidiosis and giardiasis, which has led to great advances in the understanding of these infectious diseases (Ortega-Pierres et al., 2018).

In this study, NGS techniques are used to gain a better understanding of the genetic diversity of giardiasis outbreaks in New Zealand. Taking human faecal samples from three outbreaks of giardiasis that occurred between 2010 and 2017 in various regions across the country and some samples from routine surveillance, and utilising amplicon-based metabarcoding at the glutamate dehydrogenase (*gdh*) locus, the hypothesis that epidemiologically linked cases share subtypes undetectable with consensus sequencing technologies was tested. In addition, NGS was used to detect the degree of genetic diversity present in samples from patients diagnosed with giardiasis. Comparing these results to the results of Sanger sequencing at the same locus it was possible to detect the presence of mixed infections and gained a better understanding of the assemblages of *Giardia* present in New Zealand. This study shows that amplicon-based sequencing provides better tools for painting a clearer picture of the role of protozoan genetic diversity in giardiasis outbreaks in New Zealand, which could lead to a better perception of protozoan outbreak epidemiology.

3.3 Methods

3.3.1 <u>Sampling</u>

The Protozoa Research Unit (PRU) at the Hopkirk Institute, Palmerston North, New Zealand, receives human faecal samples diagnosed as positive by accredited diagnostic laboratories from routine surveillance and outbreaks of cryptosporidiosis in New Zealand. All the samples included in this study, both from outbreaks and routine surveillance, were from patients diagnosed with giardiasis. A list of the samples from routine surveillance and outbreaks of giardiasis that occurred in New Zealand between 2010 and 2018 can be found in Table 3.1.

Table 3.1. List of samples from outbreaks and routine surveillance along with the regions in which they occurred. Outbreaks where 'Organism' is annotated with (*) highlight situations in which *Cryptosporidium* and *Giardia* were identified in the same sample. A full list of the samples used in this study can be found in Table B.1.

YEAR	REGION	ORGANISM	SAMPLE ORIGIN	NUMBER OF CASES
2010	Hawke's Bay	*Giardia*	Giardiasis Outbreak	3
2014	Gisborne	*Giardia**	Giardiasis Outbreak	5
2015	Hawke's Bay	*Giardia*	Giardiasis Outbreak	5
2016	Christchurch	*Giardia*	Routine Surveillance	1
2017	Auckland	*Giardia**	Cryptosporidiosis Outbreak	1
2017	Palmerston North	*Giardia*	Routine Surveillance	1
2017	Otago	*Giardia*	Routine Surveillance	1
TOTAL				**17**

3.3.2 DNA purification and Sanger sequencing

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

3.3.3 PCR

A partial fragment of the glutamate dehydrogenase (*gdh*) gene was amplified by nested PCR using a previously established PCR programme and set of primers (Read et al., 2004). The external primers were modified to contain MiSeq™ adapter sequences on the 5' end according to standard protocols (Illumina Inc., 2013). Agarose gel electrophoresis was used to verify the presence of fragments of the correct size (432 bp) from all the PCR reactions. A blank containing deionised H_2O was used as a negative control, and DNA from a sample that had already been verified by PCR and Sanger sequencing as containing *Giardia* DNA was used as a positive control.

3.3.4 Next-Generation Sequencing (NGS)

The PCR products for all 17 samples were cleaned according to Illumina recommended protocols (Illumina Inc., 2013). The DNA concentration in each sample was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, United States), the samples diluted to a 5 ng/µl concentration according to the Illumina protocol referenced above then delivered to the Massey Genome Service (Massey University, Palmerston North, New Zealand) for library preparation and amplicon-based sequencing. Sequencing was carried out on an Illumina MiSeq™ using 500-cycle V2 chemistry according to the manufacturer's recommendations, producing 2 × 250 base paired-end reads. Due to the potential uneven representation of bases at each

cycle with amplicon sequencing, an Illumina PhiX control library was loaded onto the Illumina MiSeq™ run at 20% volume, to even out the base composition and prevent biases in the initial few cycles that otherwise would result in base calling errors.

3.3.5 Construction of a *gdh* database

Through our collaboration with the New Zealand Ministry of Health the PRU receives anonymised faecal samples from patients diagnosed with giardiasis. The samples are analysed through PCR and Sanger sequencing at the *gdh* locus. The assembly of sequences and compilation of databases was done using Geneious v.10.2.6 (Kearse et al., 2012). Using the *Giardia intestinalis* sequences from our in-house database a separate database was compiled consisting of 858 unique *gdh* sequences, most had previously been submitted to GenBank by Garcia–R et al. (2017) and can be found in GenBank with accession numbers MT265681 – MT265802. To capture the greatest possible extent of known diversity of *Giardia gdh* sequences, a dataset of all available *gdh* sequences for *G. intestinalis* from GenBank (Benson et al., 2013) was extracted and imported into Geneious. The search strategy employed one search string (*Giardia*) and included the keywords glutamate dehydrogenase, and *gdh*. The sequences were trimmed to the length of the primers employed in this study and all sequences less than 393 bp were discarded. This left 337 unique sequences from GenBank. The 337 GenBank sequences were combined with the 858 sequences extracted from our in-house database, then duplicate sequences were extracted to create a collection of 1109 unique sequences covering most of the assemblages of *G. intestinalis* that have been characterised at the *gdh* locus.

3.3.6 Sequence processing

The Illumina sequence reads for the 17 samples involved in this study were analysed inside the Quantitative Insights Into Microbial Ecology 2 (QIIME 2) environment (Bolyen et al., 2019). The dada2 methodology (Callahan et al., 2016) was used to filter and trim the forward and reverse sequence reads, dereplicate them, calculate and plot error rates, merge paired reads and construct a sequence table, and remove chimeras. Then our

database of 1109 known unique sequences was used as a reference to assign taxonomy to the merged sequences. To remove the impact of index hopping or PCR error, from the processed and merged sequences only the top 1971 sequences were imported from dada2 into the phyloseq R package (McMurdie & Holmes, 2013) for plotting, ranking of the most expressed sequences and creation of a heatmap. The resulting table of sequences was run against the reference database to exclude any sequences that did not match known sequences of *G. intestinalis* then put through phyloseq again for further analysis. Only the top 50 sequences present across all the samples were used for the creation of bar plots and heatmaps to reduce the possibility of sequencing errors being included in the analysis. The source code used for sequence processing is shown in Appendix B.

3.4 Results

3.4.1 Overview of sample data

Of the 17 historical faecal samples from cases of giardiasis that had occurred in New Zealand between 2010 and 2018 fragments of the *gdh* gene were successfully amplified for all of them using nested PCR. All negative controls returned negative, and all positive controls returned positive. The number of reads generated from each sample after filtering, trimming and dereplication are shown in Table 3.2. The samples for which the assemblage according to Sanger sequencing were known and the most dominant assemblage according to NGS are also shown in Table 3.2. There were no disagreements in assigned dominant assemblage between the two sequencing methods. According to the NGS data, and focusing on the dominant assemblage in each sample, 11/17 samples were found to belong to sub-assemblage BIV, 1/17 to BIII, 2/17 to AII, 1/17 to AIII, 2/17 to E. Figure 3.1. provides a comparison of the assemblage assigned by Sanger sequencing and the diversity captured by NGS. It shows that even in genetically diverse samples, like the one from the outbreak in Hawke's Bay in 2015, there are agreements between the Sanger sequence data and the NGS data. Analysis of the NGS data was conducted to probe the intra-sample diversity of these samples.

Table 3.2. Sample Data. Sample assemblages according to results of Sanger sequencing compared with most abundant assemblages according to NGS. The number of reads generated by NGS from each sample after filtering, trimming and dereplication are shown for reference.

SAMPLE NO.	ID	SANGER	NGS	NGS READS
1	1997	BIV	BIV	55235 reads in 15404 unique sequences
2	1998	BIV	BIV	113042 reads in 21061 unique sequences
3	1999	BIV	BIV	136387 reads in 25493 unique sequences
4	10015	AII	AII	118718 reads in 23734 unique sequences
5	10046	BIV	BIV	141257 reads in 24507 unique sequences
6	10047	BIV	BIV	95812 reads in 19814 unique sequences
7	10048	BIV	BIV	95836 reads in 28269 unique sequences
8	10049	BIV	BIV	106343 reads in 24744 unique sequences
9	10936	BIV	BIV	8184 reads in 3121 unique sequences
10	10937	BIV	BIV	144483 reads in 39869 unique sequences
11	10938	AII	AII	116354 reads in 33518 unique sequences
12	10939	BIV	BIV	121820 reads in 21446 unique sequences
13	10940	BIII	BIII	20678 reads in 6331 unique sequences
14	13273	BIV	BIV	103784 reads in 22832 unique sequences
15	14201	Unspecified	E	75624 reads in 15918 unique sequences
16	11359	Unspecified	AIII	112267 reads in 19785 unique sequences
23	13805	Unspecified	E	87460 reads in 11150 unique sequences
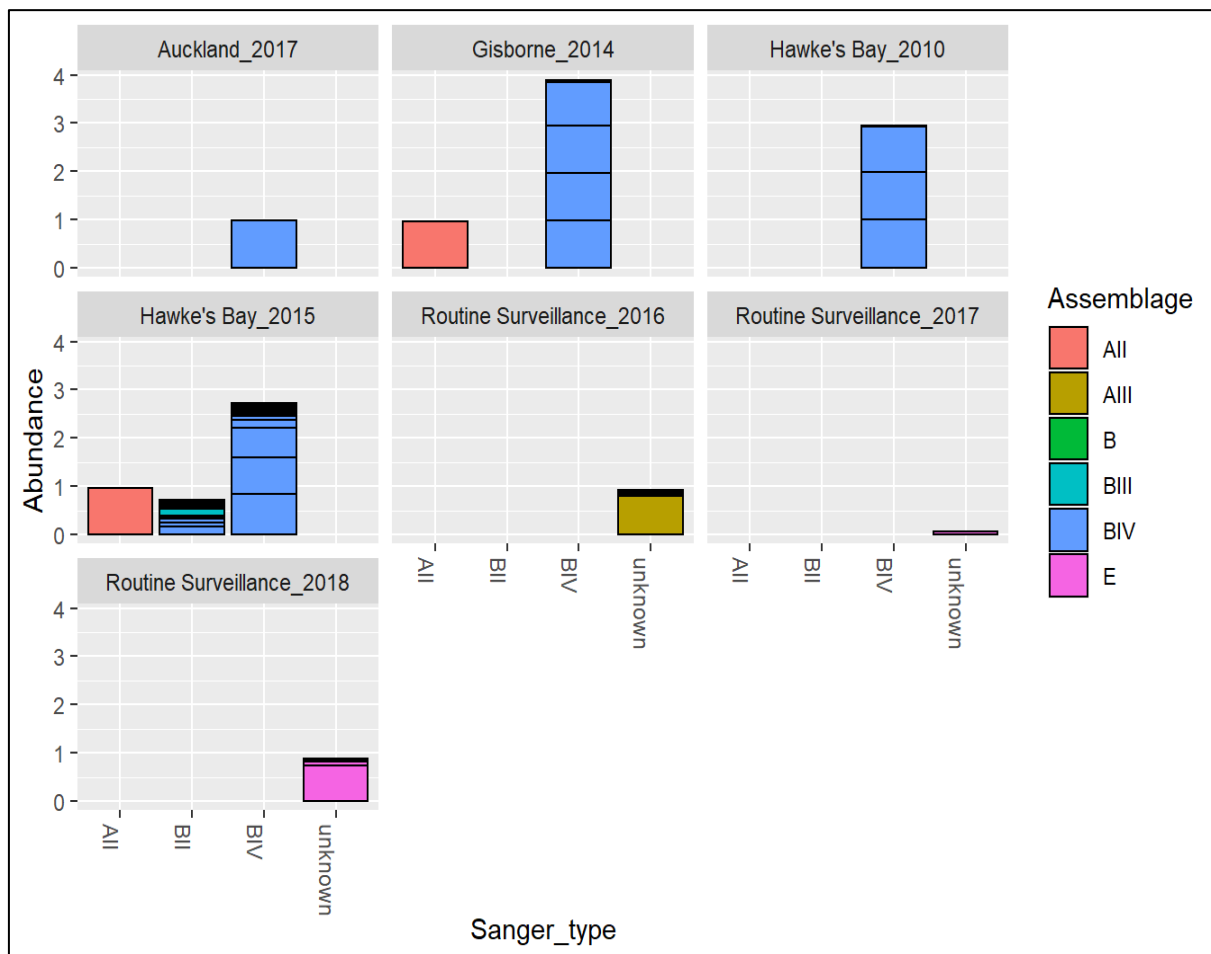
Figure 3.1. bar plot showing the taxonomic distribution of (sub) assemblages in samples from the routine surveillance and the multiple outbreaks included in this study. The X

47

axis shows the (sub) assemblage of each sample according to Sanger sequence data and the Y axis displays the number of samples corresponding to each assemblage; the colour codes in each bar represent the genetic diversity within each sample according to NGS.

3.4.2 <u>Metabarcoding analysis</u>

The diversity of assemblages found in each sample after processing and analysis of the NGS reads are shown in Figure 3.2. Similar to the results from the Sanger sequencing, the most abundant assemblage in most samples was assemblage B, specifically sub-assemblage BIV. This assemblage was present at some level in 16/17 samples, only one sample (13805_S23) did not have any variants of sub-assemblage BIV present in it. There was evidence of mixed infections in 13/17 samples. The majority of the genetic diversity within those 13 samples was due to the presence of multiple variants within assemblage B, for example, samples 10937_S10 and 10940_S13 showed evidence of multiple variants corresponding to assemblage B. The second most common assemblage present in this study was assemblage A, with 7/17 samples showing the presence of at least one variant of that assemblage.

Three samples from routine surveillance were included in this study (see Table 3.1.) to compare the genetic diversity between samples from outbreaks and samples from sporadic cases. No significant differences were observed. Two samples (11359_S16 & 14201_S15) from routine surveillance represented the first report of sub-assemblage AIII and assemblage E in human samples from the South Island in New Zealand. These samples were analysed further in another study (Garcia-R et al., 2021).
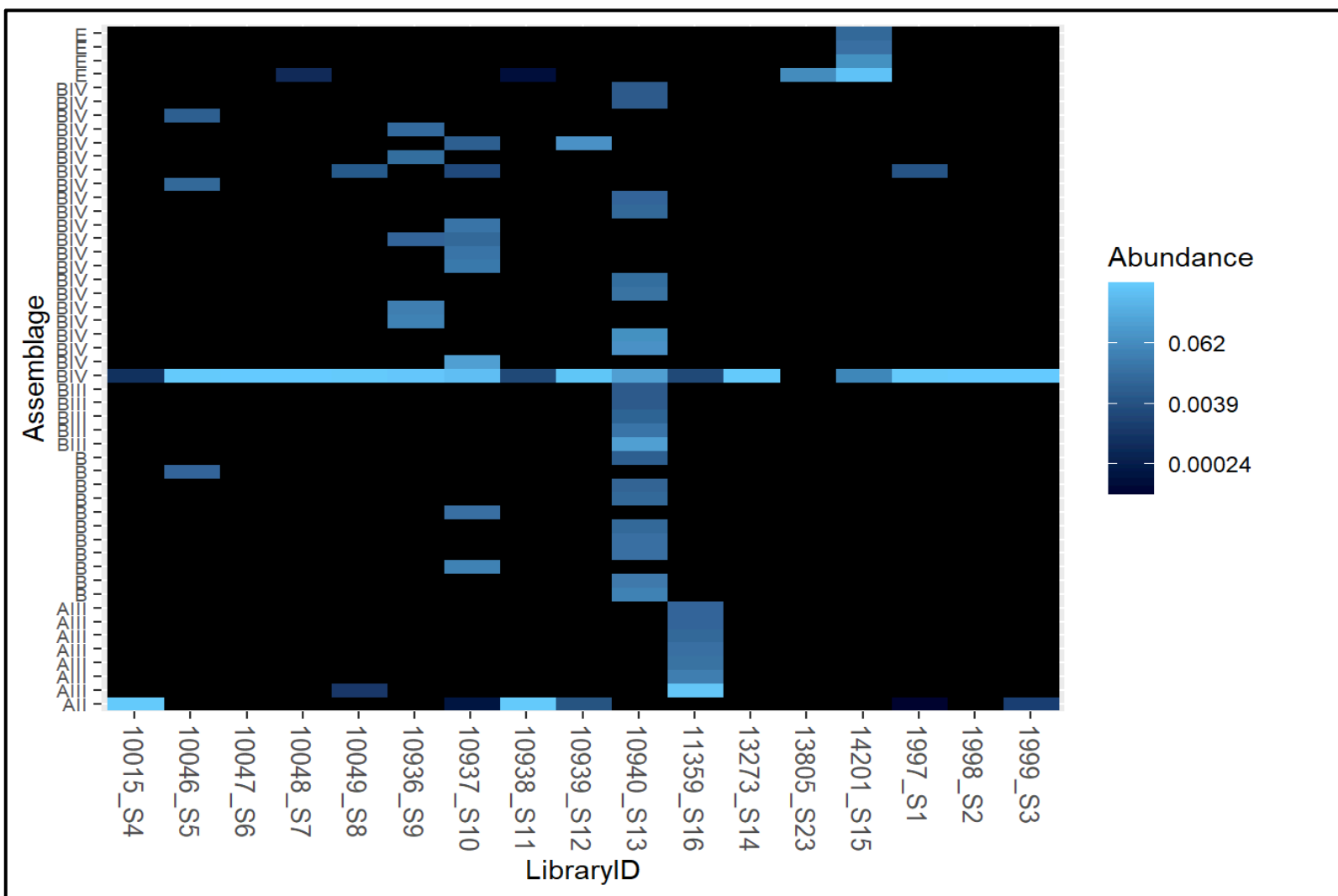
Figure 3.2. Heatmap showing the relative abundance of the top 50 sequences in each sample. The multiple variants of each assemblage present in each sample are displayed on the Y axis.

3.4.3 Identifying links between outbreak cases

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

Of the 5 samples from the outbreak that occurred in Gisborne in 2014, 4/5 were characterised as sub-assemblage BIV and 1/5 as AII according to Sanger and NGS data. Figure 3.3. is a heatmap showing the genetic diversity, captured by NGS, within the samples involved in this outbreak. From this, it is evident that a single variant of sub-assemblage BIV is shared by all the samples in this outbreak. Also, a copy of sub-assemblage AIII is present in one of the samples from this outbreak (10049_S8).

Of the 5 samples received from the outbreak in Hawke's Bay in 2015, 3/5 were identified as sub-assemblage BIV, 1/5 as BIII, and 1/5 as AII according to Sanger and NGS data. From the heatmap shown in Figure 3.4. it is evident that, despite the differences in dominant assemblages, sub-assemblage BIV is shared between all the samples from this outbreak. Sample 13273_S14 represented the only sample from an outbreak of cryptosporidiosis in Auckland. According to the NGS data this sample was also positive for *G. intestinalis* sub-assemblage BIV (Figure 3.2.). This represents an example of a mixed-species infection. The NGS abundance data for the rest of the outbreaks is available in Figure B.1.
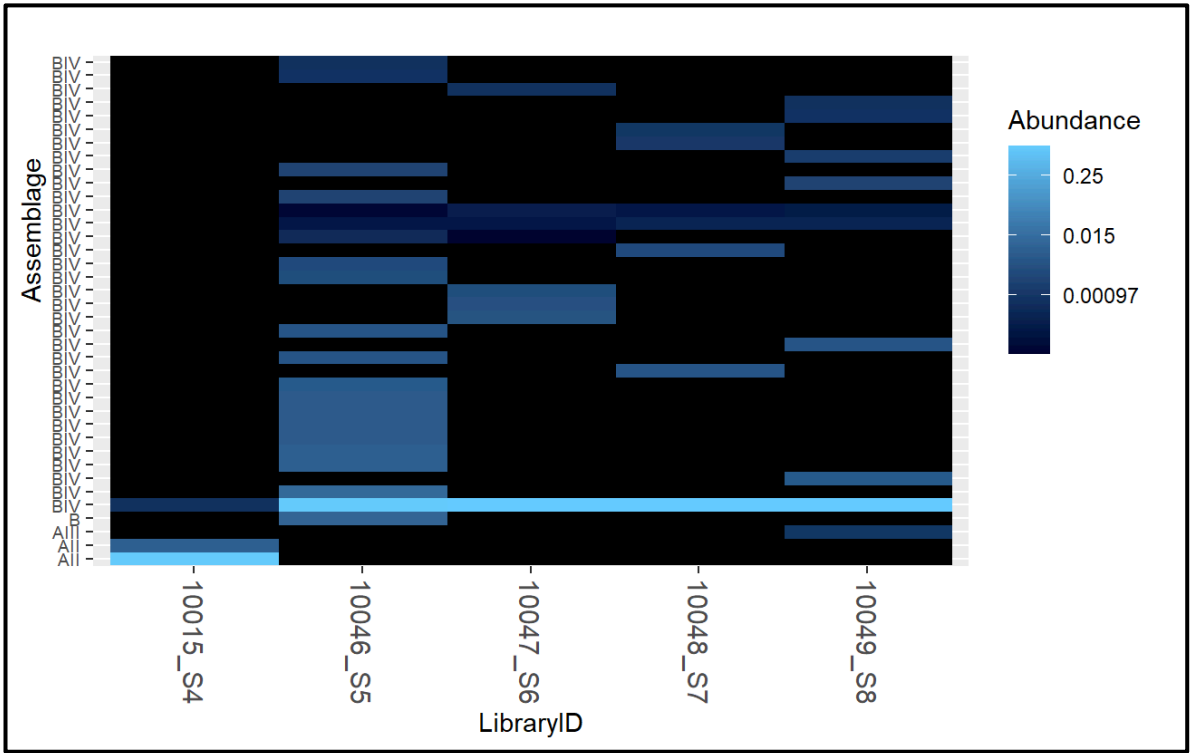
Figure 3.3. Heatmap showing the relative abundance of the top 50 sequences in each sample from the outbreak of giardiasis that occurred in Gisborne in 2014. The multiple variants of each assemblage present in each sample are displayed on the Y axis.
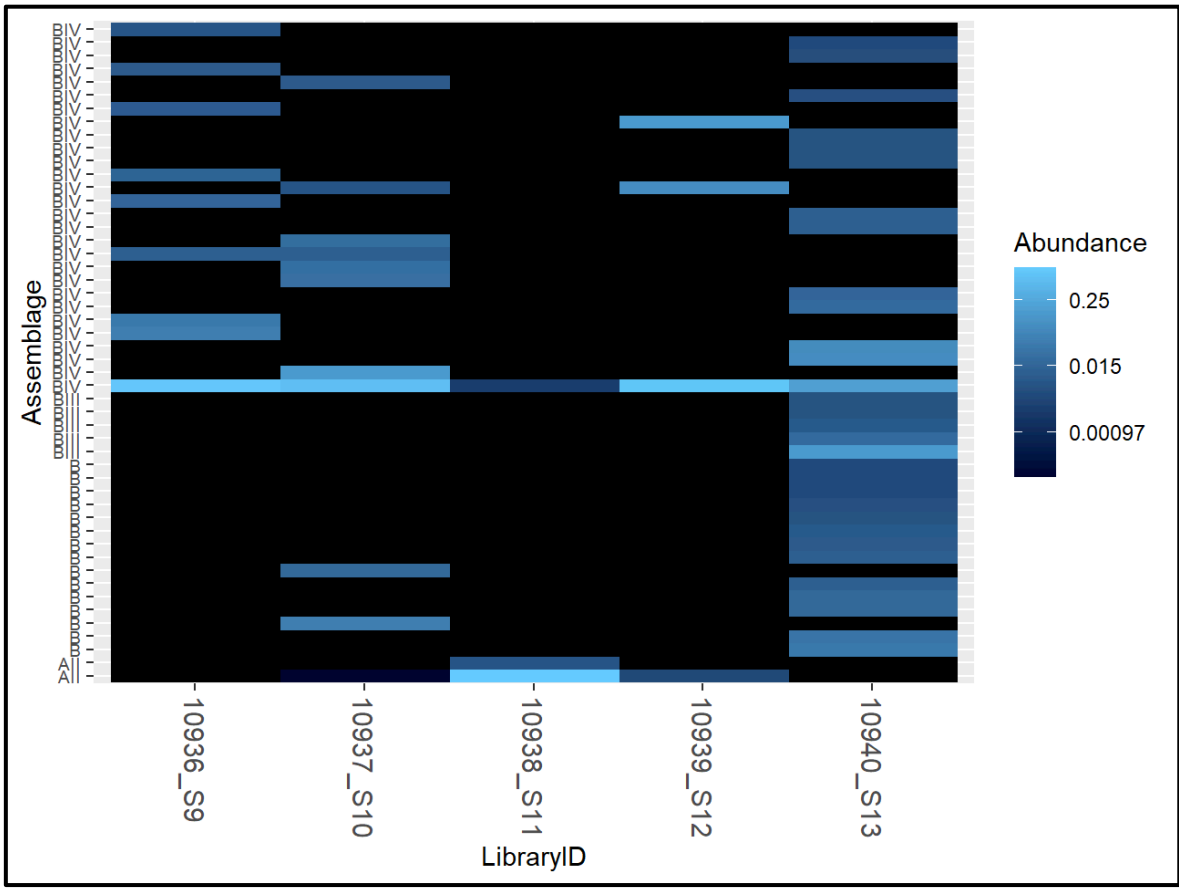
Figure 3.4. Heatmap showing the relative abundance of the top 50 sequences in each sample from the outbreak of giardiasis that occurred in Hawke's Bay in 2015. The multiple variants of each assemblage present in each sample are displayed on the Y axis.

3.5 Discussion

This investigation into the intra-sample diversity of *G. intestinalis* in patients from historical outbreaks of giardiasis in New Zealand compares the capabilities of NGS and Sanger sequencing technologies. The strength of Sanger sequencing lies in its ability to detect the dominant sequence within a sample. The results outlined here show that NGS is also capable of the same level of discernment with regards to the identification of dominant sequences, shown by the agreements between the data from Sanger sequence and NGS of samples from cases of giardiasis that occurred in New Zealand between 2010 and 2018. The aim of this study was to use NGS to capture the diversity within samples, and this is where the benefit of NGS over Sanger shows itself. NGS is capable of sequencing multiple reads in each sample, compared to the one consensus read per sample achieved with consensus sequencing technologies. Here, NGS was employed to uncover the genetic diversity present within cases of giardiasis in this country. As previously stated in the introduction to this study, the clinical manifestation of giardiasis can differ between individuals. Further work needs to be done to ascertain if there is a link between (sub)assemblage and clinical presentation, however recent advances in the *in vitro* culture of *Giardia* (Liu et al., 2020) have the potential to assist in the resolution of this question. So, the ability to capture the genetic diversity within samples from cases of giardiasis and link them to the symptoms displayed by the patient could greatly advance our understanding of the disease mechanisms of this parasite.

The data presented here suggest that assemblage B is still the most common assemblage of *Giardia* in New Zealand, present in 16/17 samples. However, the ability to capture the diversity of assemblages within samples showed that, although they might not be dominant, assemblages A and E reported with increased frequency in New Zealand, as evidenced by their presence in 7/17 and 4/17 samples respectively. This is particularly

significant since assemblage E was thought to be exclusively infectious to livestock. However, recent studies have shown that it is increasingly present in humans as well (Abdel-Moein & Saeed, 2016). The significance of this finding is discussed in a study published by our research group (Garcia-R et al., 2021).

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

Another aim of this study was to use NGS to uncover genetic links between epidemiologically linked samples. It was hypothesised that epidemiologically linked cases share assemblages undetectable with consensus sequencing technologies. The outbreaks that occurred in Gisborne in 2014 and Hawke's Bay in 2015 provided a perfect case study for this. In those outbreaks there were multiple dominant assemblages

present in the samples within each outbreak. By applying NGS and metabarcoding at the *gdh* locus it was shown that sub-assemblage BIV was shared between all samples from the Gisborne and Hawke's Bay outbreaks, thereby verifying the hypothesis. This improves our understanding of the epidemiology of these outbreaks.

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

4 Capturing genetic diversity in outbreaks of Cryptosporidiosis occurring in New Zealand from 2010 to 2018

4.1 Abstract

Cryptosporidiosis is a disease caused by the parasite *Cryptosporidium*. Globally, it is one of the leading causes of diarrhoea and is prevalent in all continents apart from Antarctica. Cryptosporidiosis is a notifiable disease in New Zealand due to an appreciation of its public health significance. To further the understanding of the outbreak epidemiology of this disease in New Zealand, this study analyses samples from outbreaks and sporadic cases occurring between 2010 and 2018. By comparing Next-Generation Sequencing (NGS) and Sanger sequencing of the glycoprotein 60 (*gp60*) locus, the benefits of NGS and limitations of Sanger sequencing in capturing the genetic diversity within and between samples from outbreaks and sporadic cases is highlighted. Implementation of NGS metabarcoding at the *gp60* locus uncovered significant intra- and inter-sample genotypic diversity in outbreaks and allowed for the identification of the subtypes shared by epidemiologically linked cases. The high resolution of this sequencing technology led to the identification of the formerly rare subtypes IfA12G1R5 and IgA20 as emerging variants in New Zealand.

4.2 Introduction

Cryptosporidiosis is a disease most commonly characterised by acute, watery diarrhoea affecting approximately 7.6% of the world's population across all regions excluding Antarctica (Cacciò & Widmer, 2013; Dong et al., 2020). The disease has a wide host range including humans, domestic animals, cattle, and a wide array of wildlife (Pumipuntu & Piratae, 2018). In healthy humans the disease is usually self-limiting with an incubation period of 4 – 28 days, and acute infection lasting 6 – 7 days (Hunter et al., 2004b). Nevertheless, the disease can be fatal in immunocompromised humans, and infants of human, bovine or ovine species. The Global Burden of Disease study found that in 2015

cryptosporidiosis was responsible for approximately 12.1% of deaths in children under 5 globally (Troeger et al., 2017). The disease is caused by the protozoan parasite *Cryptosporidium*, which infects the epithelial cells of the gastrointestinal tract to cause the disease. The omnipresence of this parasite and the severity of the disease has caused many countries to include it in their notifiable disease databases. In New Zealand cryptosporidiosis has been a notifiable disease since 1996 (Learmonth et al., 2004).

There are currently 44 recognised species of *Cryptosporidium*, and the 2 responsible for the majority of infections in humans are *C. hominis* and *C. parvum*, however, a total of 20 species have been identified in human infections to date (Feng et al., 2018). *C. hominis* causes the majority of anthroponotic cases of cryptosporidiosis in humans and is thought to mainly prefer humans as a host. However, it has been found in a variety of animal hosts including equine and non-human primate species (Inácio et al., 2017; Parsons et al., 2015; Widmer et al., 2020). Over 10 subtype families have been identified in *C. hominis*, with the virulent subtype IbA10G2 being the variant most commonly found in infected individuals across all socioeconomic settings (Xiao & Feng, 2017). *C. parvum* has a wider host range covering humans, companion animals, livestock and wildlife (Cacciò & Widmer, 2013). Close to 20 subtype families of *C. parvum* have been identified so far (Xiao & Feng, 2017).

According to public health surveillance data compiled on behalf of the New Zealand Ministry of Health, *Cryptosporidium* has been implicated as the causative agent in a significant percentage of outbreaks each year since 2001 (https://surv.esr.cri.nz/surveillance/annual_outbreak.php), making the disease of particular public health significance in this country. Between 2010 and 2019 the rate of cryptosporidiosis per 100,000 population in New Zealand has averaged 20.9/100,000, with a high of 33/100,000 in 2018 and a low of 13/100,000 in 2014 (Figure 4.1.). Initially, cryptosporidiosis was diagnosed microscopically, but this proved inefficient for characterisation due to the morphological similarity of the environmental stage of the parasite, the oocyst, between species. The advent of molecular typing technologies such as PCR and consensus sequencing techniques allowed for better characterisation of the parasite and understanding of the epidemiology and population genetics of the parasite (Feng et al., 2018; Xiao & Feng, 2017). A previous study by Massey University's Molecular

Epidemiology and Public Health Laboratory (mEpiLab), analysing 2931 human faecal samples over 11 years (2009 – 2019) identified 6 species of *Cryptosporidium* in humans in New Zealand, namely, *C. parvum*, *C. hominis*, *C. cuniculus*, *C. erinacei*, *C. meleagridis*, and *C. tyzzeri* (Garcia-R et al., 2020). Analysis of the 60 kDa glycoprotein (*gp60*) gene has been a pivotal subtyping tool utilised in characterisation of *Cryptosporidium spp.* and is still commonly used to this day. In the study referred to above *C. parvum* was the most common species (59%) identified with 46 subtypes present in those samples. Unlike most high-income countries (HICs), like those in Europe, where the hyper-transmissible IIaA15G2R1 subtype is most common (Khan et al., 2018b), IIaA18G3R1 was the most frequently reported subtype of *C. parvum* in New Zealand. *C. hominis* was the second most common species identified, covering 36 subtypes with the most common being IbA10G2, similar to observations in most HICs (Garcia-R et al., 2020).

Previous studies investigating the cases of cryptosporidiosis in New Zealand have utilised Sanger sequencing. This presents limitations because such technologies do not capture the full genetic diversity of *Cryptosporidium* within a host. Next-generation sequencing (NGS) technologies can capture significantly more diversity within each sample, capable of sequencing millions of reads per sample compared to the one consensus read achieved with Sanger sequencing. This helps paint a clearer picture of the genetic diversity and population structure of *Cryptosporidium* in humans (Feng et al., 2018), giving researchers a better understanding of the epidemiology of the disease, informing public health and enabling the attribution of specific subtypes to outbreaks. A study by Grinberg et al., (2013), applying NGS to two *C. parvum* human isolates showed extensive intra-host diversity in samples for which Sanger sequencing had identified a single subtype only. Application of these techniques in population-level studies could serve as a forecasting mechanism that will aid the identification of emerging subtypes of *Cryptosporidium* in each country and across the globe. For instance, in the USA IbA10G2 was the dominant *C. hominis* subtype implicated in outbreaks since records began. Then, in 2007 IaA28R4 took that position, and from 2013 IfA12G1R5 became the most dominant *C. hominis* subtype in sporadic and outbreak cases in the USA (Hlavsa et al., 2017; Putignani & Menichella, 2010). Also, a recent study in Australia found that in 2017 IfA12G1R5 was the most dominant subtype in cases of cryptosporidiosis in Western Australia (Braima et al., 2019). IfA12G1R5 was previously thought to be a rare subtype of *C. hominis* and now

it is dominant in two HICs (Australia and USA). Application of NGS technologies could aid the identification of such emerging subtypes, which will advance our understanding of the factors that lead to their introduction, spread and domination in a population. Morris et al., (2019) aptly stated the aim of genotyping in a context of public health is to understand transmission, improve detection resolution, investigation and interpretation of outbreaks. In addition, they explained that a potential impact of such a strategy lies in its ability to monitor emerging species and subtypes, identify links and risk factors and pinpoint the source of outbreaks and contamination.

In this study, 105 human samples from historic cases of cryptosporidiosis occurring between 2010 and 2018 in New Zealand were analysed using NGS. Eighty-six samples were from outbreaks and 19 from sporadic cases. The aim was to utilise NGS to gain a better understanding of the genetic diversity within and between samples to help in the identification of *Cryptosporidium* subtypes shared between samples from epidemiologically linked cases. Through this, potential candidates for emerging subtypes were identified, and the results provide a better understanding of the epidemiology of cryptosporidiosis in New Zealand.
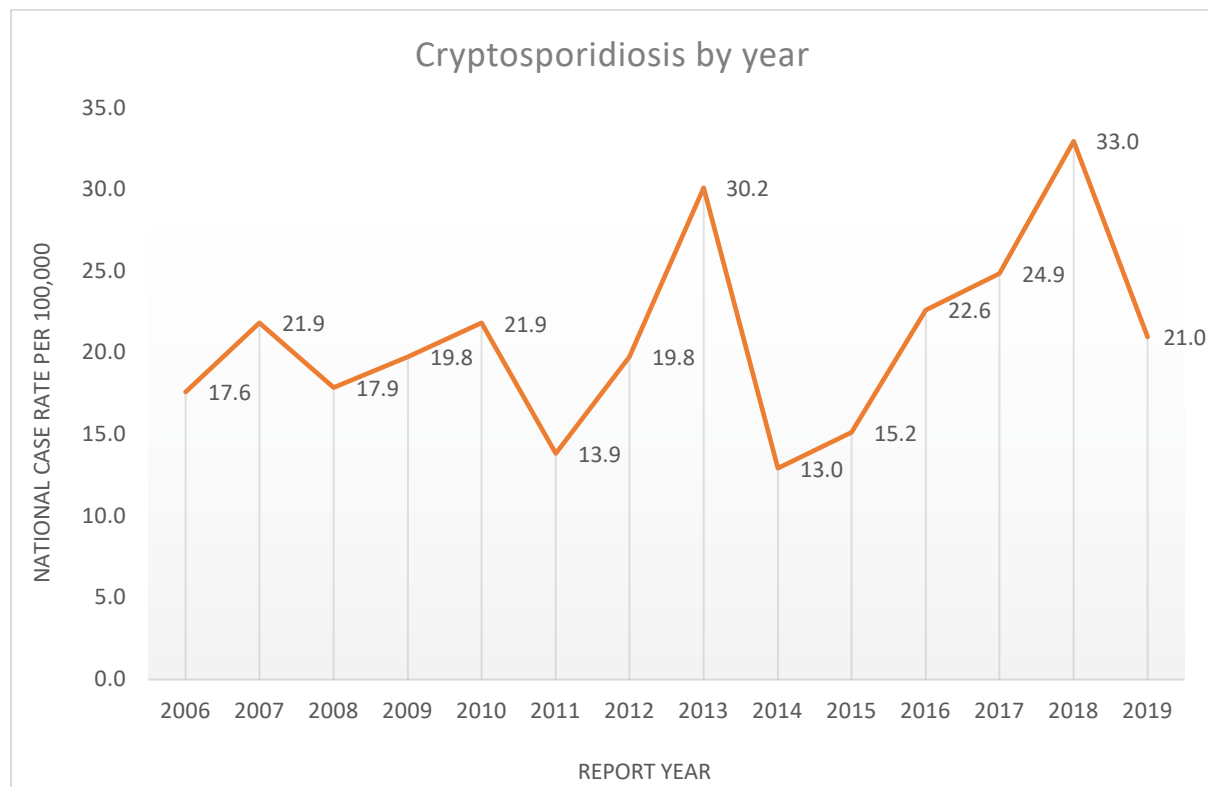
Figure 4.1. National case rate of cryptosporidiosis per 100,000 population in New Zealand from 2006 to 2019. Data collected from the national notifiable disease surveillance system (https://surv.esr.cri.nz).

4.3 Methods

4.3.1 Sampling

Anonymised human faecal samples from routine surveillance and outbreaks of cryptosporidiosis in New Zealand were collected as part of the same process outlined in Chapter 3. Only samples from patients diagnosed with cryptosporidiosis were included in this study. A list of the outbreaks that occurred in New Zealand between 2010 and 2018 that have been reported to the unit can be found in Table 4.1.

Table 4.1. List of outbreaks along with the regions in which they occurred. Outbreaks where 'Organism' is annotated with (*) highlight situations in which *Cryptosporidium* and *Giardia* were identified in the same sample in one of more cases. A full list of the samples used in this study can be found in Table 4.2.

YEAR	REGION	ORGANISM	NUMBER	OF
			SAMPLES	
2010	Auckland	*Cryptosporidium*		7
	Christchurch	*Cryptosporidium*		17
2013	Hawke's Bay	*Cryptosporidium*		22
	Waikato	*Cryptosporidium*		5
	Wellington	*Cryptosporidium*		5
	Taranaki	*Cryptosporidium*		3
2015	Auckland	*Cryptosporidium*		6
2017	Auckland	*Cryptosporidium**		9
	Blenheim	*Cryptosporidium*		3
2018	Wellington	*Cryptosporidium*		9
TOTAL				86

4.3.2 DNA purification and Sanger sequencing

Total DNA was isolated, nested PCR and Sanger sequencing was carried out as described in Chapter 3 with the only modification being the targeting of the glycoprotein 60 (*gp60*) locus in this study using the primers outlined below.

4.3.3 PCR

A partial fragment of the glycoprotein 60 (*gp60*) gene was amplified by nested PCR using a previously established PCR programme and set of primers (Zahedi, Gofton, et al., 2017). The external primers were modified to contain MiSeq™ adapter sequences on the 5' end according to standard protocols (Illumina Inc., 2013). Agarose gel electrophoresis was used to verify the presence of fragments of the correct size from all the PCR reactions. A blank containing deionised H_2O was used as a negative control, and DNA from a sample that had already been verified by PCR and Sanger sequencing as containing *Cryptosporidium* DNA was used as a positive control. The PCR products were approximately 400 bp for the *Cryptosporidium* samples.

4.3.4 Next-Generation Sequencing (NGS)

The PCR products for all 105 samples were cleaned according to Illumina recommended protocols (Illumina Inc., 2013). The DNA concentration in each sample was measured using NanoDrop (Thermo Fisher Scientific, Waltham, Massachusetts, United States), the samples diluted to a concentration of 5 ng/μl according to the Illumina protocol referenced above, then delivered to the Massey Genome Service (Massey University, Palmerston North, New Zealand) for library preparation and amplicon-based sequencing. Sequencing was carried out on an Illumina MiSeq™ using 500-cycle V2 chemistry according to the manufacturer's recommendations, producing 2 × 250 base paired-end reads. Due to the potential uneven representation of bases at each cycle with amplicon sequencing, an Illumina PhiX control library was loaded onto the Illumina MiSeq™ run at 20% volume, to even out the base composition and prevent biases in the initial few cycles that otherwise would result in base calling errors .

4.3.5 Construction of a *gp60* database

Through our collaboration with the New Zealand Ministry of Health, PRU receives anonymised faecal samples from patients diagnosed with cryptosporidiosis. The samples are analysed through PCR and Sanger sequencing at the *gp60* locus. The assembly of sequences and compilation of databases was done using Geneious v.10.2.6 (Kearse et al.,

2012). Using the *Cryptosporidium* sequences from our in-house database a separate database was compiled consisting of 139 unique *gp60* sequences from *C. hominis* and *C. parvum*, most had previously been submitted to GenBank by Garcia–R et al. (2017) and can be found in GenBank with accession numbers KY123918–KY124121 and MT265681-MT265802 (Garcia-R et al., 2020). To make the database more robust all available *gp60* sequences from GenBank were extracted. The sequences were subdivided by species and used to develop a barcode containing all known combinations of each nucleotide residue at each position within the range covered by the primers described above for each species. This barcode was used to develop a synthetic database of all possible combinations of residues for all known species and subtypes.

4.3.6 <u>Sequence processing</u>

Sequence processing was conducted by A/Prof Patrick Biggs. The Illumina reads for the 105 samples involved in this study were analysed inside the Quantitative Insights Into Microbial Ecology 2 (QIIME 2) environment (Bolyen et al., 2019). The dada2 methodology (Callahan et al., 2016) was used to filter and trim the forward and reverse sequence reads, dereplicate them, calculate and plot error rates, merge paired reads and construct a sequence table, and remove chimeras. Our synthetic database was used to assign taxonomy to the merged sequences. The taxonomic assignations were verified by BLAST. After sequence processing, only the top 3545 were carried over for further analysis to remove the impact of index hopping or PCR error.

4.3.7 <u>Sequence Analysis</u>

Analysis of the processed sequences, as shown in the results below, was conducted using the Phyloseq R library (McMurdie & Holmes, 2013). The source code is shown in Appendix C. Only the top 50 sequences present across all samples and within each outbreak were used for the creating of bar plots and heatmaps. This was done to reduce the chance of sequencing errors being including in the analysis.

4.4 Results

4.4.1 Intra-sample diversity and dominant subtypes of *Cryptosporidium* detected in human samples

All 105 samples were previously analysed by PCR at the *gp60* locus and found to be positive for the presence of *Cryptosporidium*. All negative controls came back negative and all positive controls showed strong bands of approximately 400 bp. 89 of these samples were also subtyped using Sanger sequencing at the same locus. NGS at the *gp60* locus of all 105 samples identified 5 species of *Cryptosporidium: C. hominis, C. parvum, C. cuniculus, C. tyzzeri* and *C. erinacei*. The most common dominant species identified in the samples was *C. hominis* (76.2%) followed by *C. parvum* (19.0%). Figure 4.2. shows a heatmap displaying the abundance of the various subtypes identified in this study across all samples. There is evidence of intra-sample diversity in most faecal samples analysed in this study. There is also significant diversity within subtype families, with Ib showing the largest number of variants, followed by IIa then Ig.

The most abundant subtype of *C. hominis*, as determined by the percentage of samples in which it was present, was IbA10G2 (77.1%), IgA17 was the second most common (48.6%), and IgA16 was the third most common (27.6%). It is worth noting that IgA20 was present in 21.9% of the samples, and IfA12G1R5 in 15.2% (Table C.1.). For *C. parvum* IIaA18G3R1 (38.1%) was the most abundant subtype, followed by IIdA17G1 (10.5%), and IIaA19G4R1 was the third most abundant (9.5%).
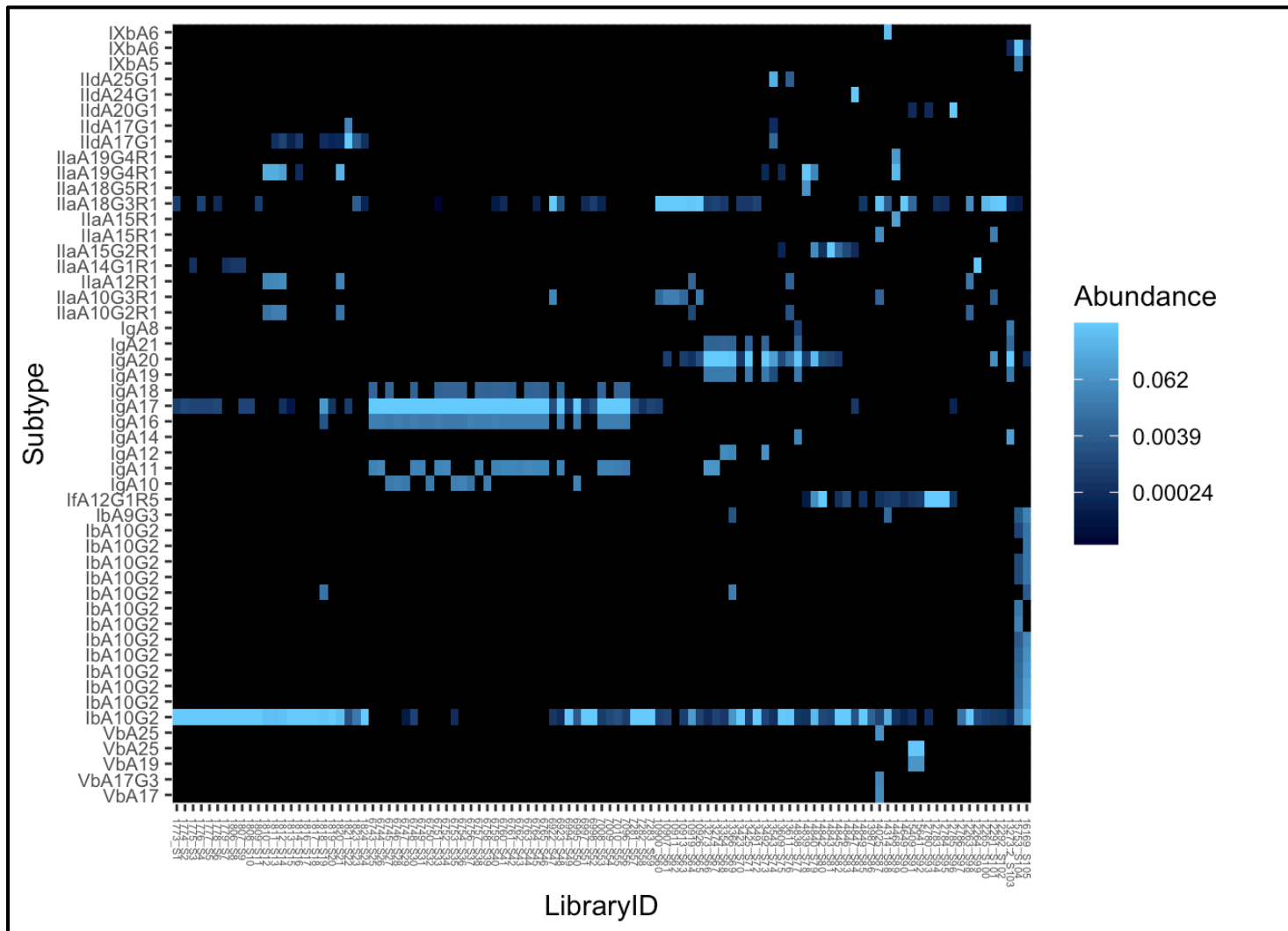
Figure 4.2. Heatmap showing the relative abundance of the top 50 sequences across all samples (LibaryID) involved in this study. The multiple subtypes present in each sample are displayed on the Y axis.

4.4.2 Sanger vs NGS

Of the 89 samples for which the subtype according to Sanger sequencing was available, after NGS and analysis of the same locus (*gp60*) 74 samples had the same dominant subtype family as Sanger, while 14 had different dominant subtype families from that identified by Sanger sequencing (Table 4.2.). This highlights the benefits of NGS over Sanger sequencing for taxonomic identification of dominant subtypes of *Cryptosporidium* from mixed infections.

Table 4.2. A comparison of the most abundant subtype families according to Sanger sequencing and NGS of all samples included in this study. Samples with disagreements between the two methods are highlighted in grey, and "Unspecified" denotes samples that had low-quality sequences that did not enable a subtype family classification.

Sample	MasseyID	Outbreak	Sanger	NGS
S1	1773	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S2	1774	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S3	1775	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S4	1776	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S5	1777	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S6	1778	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S7	1779	Auckland_2010	*C. hominis Ib*	*C. hominis Ib*
S8	1806	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S9	1807	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S10	1808	Christchurch_2010	Unspecified	*C. hominis Ib*
S11	1809	Christchurch_2010	Unspecified	*C. hominis Ib*
S12	1810	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S13	1811	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S14	1812	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S15	1813	Christchurch_2010	Unspecified	*C. hominis Ib*
S16	1814	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S17	1816	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S18	1817	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S19	1818	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S20	1819	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S21	1820	Christchurch_2010	*C. parvum IIa*	*C. hominis Ib*

Sample	MasseyID	Outbreak	Sanger	NGS
S22	1821	Christchurch_2010	*C. parvum IId*	*C. parvum IId*
S23	1823	Christchurch_2010	*C. erinacei*	*C. erinacei*
S24	1824	Christchurch_2010	*C. hominis Ib*	*C. hominis Ib*
S25	6743	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S26	6744	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S27	6745	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S28	6746	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S29	6747	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S30	6748	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S31	6749	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S32	6750	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S33	6751	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S34	6752	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S35	6753	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S36	6754	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S37	6756	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S38	6757	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S39	6758	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S40	6759	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S41	6760	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S42	6761	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S43	6762	Hawke's Bay_2013	Unspecified	*C. hominis Ig*
S44	6763	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S45	6764	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S46	6765	Hawke's Bay_2013	*C. hominis Ig*	*C. hominis Ig*
S47	6922	Waikato_2013	*C. parvum IIa*	*C. parvum IIa*
S48	6923	Waikato_2013	*C. hominis Ig*	*C. hominis Ig*
S49	6994	Wellington_2013	Unspecified	*C. hominis Ib*
S50	6995	Wellington_2013	*C. hominis Ig*	*C. hominis Ig*
S51	6997	Wellington_2013	*C. hominis Ib*	*C. hominis Ib*
S52	6998	Wellington_2013	*C. hominis Ib*	*C. hominis Ib*
S53	7008	Waikato_2013	*C. hominis Ig*	*C. hominis Ig*
S54	7009	Waikato_2013	*C. hominis Ig*	*C. hominis Ig*
S55	7010	Waikato_2013	*C. hominis Ig*	*C. hominis Ig*
S56	7096	Wellington_2013	*C. hominis Ig*	*C. hominis Ig*
S57	7281	Taranaki_2013	*C. hominis Ib*	*C. hominis Ib*
S58	7282	Taranaki_2013	*C. hominis Ib*	*C. hominis Ib*
S59	7283	Taranaki_2013	*C. hominis Ib*	*C. hominis Ib*
S60	10900	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*
S61	10907	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*
S62	10911	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*

Sample	MasseyID	Outbreak	Sanger	NGS
S63	10913	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*
S64	10919	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*
S65	10926	Auckland_2015	*C. parvum IIa*	*C. parvum IIa*
S66	13273	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S67	13274	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S68	13354	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S69	13366	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S70	13423	Auckland_2017	*C. hominis Ib*	*C. hominis Ib*
S71	13425	Auckland_2017	Unspecified	*C. hominis Ig*
S72	13491	Auckland_2017	*C. hominis Ib*	*C. hominis Ib*
S73	13492	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S74	13543	Blenheim_2017	*C. parvum IId*	*C. parvum IId*
S75	13609	Blenheim_2017	*C. hominis Ib*	*C. hominis Ib*
S76	13611	Blenheim_2017	*C. hominis Ib*	*C. hominis Ib*
S77	14838	Wellington_2018	*C. hominis Ig*	*C. hominis Ig*
S78	14839	Wellington_2018	*C. parvum IIa*	*C. parvum IIa*
S79	14840	Wellington_2018	*C. parvum IId*	*C. hominis Ig*
S80	14842	Wellington_2018	*C. hominis If*	*C. hominis If*
S81	14843	Wellington_2018	*C. parvum IIa*	*C. parvum IIa*
S82	14845	Wellington_2018	*C. hominis Ib*	*C. hominis Ib*
S83	14846	Wellington_2018	*C. hominis Ib*	*C. hominis Ib*
S84	14847	Wellington_2018	*C. parvum IId*	*C. parvum IId*
S85	14849	Wellington_2018	*C. hominis Ib*	*C. hominis Ib*
S86	13987	Routine Surveillance_2017	*C. erinacei*	*C. erinacei*
S87	14025	Routine Surveillance_2017	*C. cuniculus*	*C. parvum IIa*
S88	14314	Routine Surveillance_2017	*C. tyzzeri*	*C. tyzzeri*
S89	14468	Routine Surveillance_2018	*C. erinacei*	*C. parvum IIa*
S90	14649	Routine Surveillance_2018	Unspecified	*C. parvum IIa*
S91	15409	Routine Surveillance_2018	*C. cuniculus*	*C. cuniculus*
S92	15641	Routine Surveillance_2018	*C. cuniculus*	*C. cuniculus*
S93	12780	Routine Surveillance_2015	*C. hominis Ib*	*C. hominis If*
S94	12783	Routine Surveillance_2015	*C. parvum IIa*	*C. hominis If*
S95	12784	Routine Surveillance_2015	*C. parvum IIa*	*C. hominis If*

Sample	MasseyID	Outbreak	Sanger	NGS
S96	12785	Routine Surveillance_2015	*C. parvum IIa*	*C. parvum IId*
S97	12786	Routine Surveillance_2015	*C. parvum IIa*	*C. parvum IId*
S98	12263	Routine Surveillance_2015	*C. hominis If*	*C. hominis Ib*
S99	12264	Routine Surveillance_2015	*C. hominis If*	*C. parvum IIa*
S100	12265	Routine Surveillance_2015	*C. hominis If*	*C. parvum IIa*
S101	12291	Routine Surveillance_2015	*C. parvum IId*	*C. parvum IIa*
S102	12292	Routine Surveillance_2015	*C. parvum IId*	*C. parvum IIa*
S103	13273.2	Auckland_2017	*C. hominis Ig*	*C. hominis Ig*
S104	15753	Routine Surveillance_2018	Unspecified	*C. tyzzeri*
S105	16169	Routine Surveillance_2018	Unspecified	*C. hominis Ib*
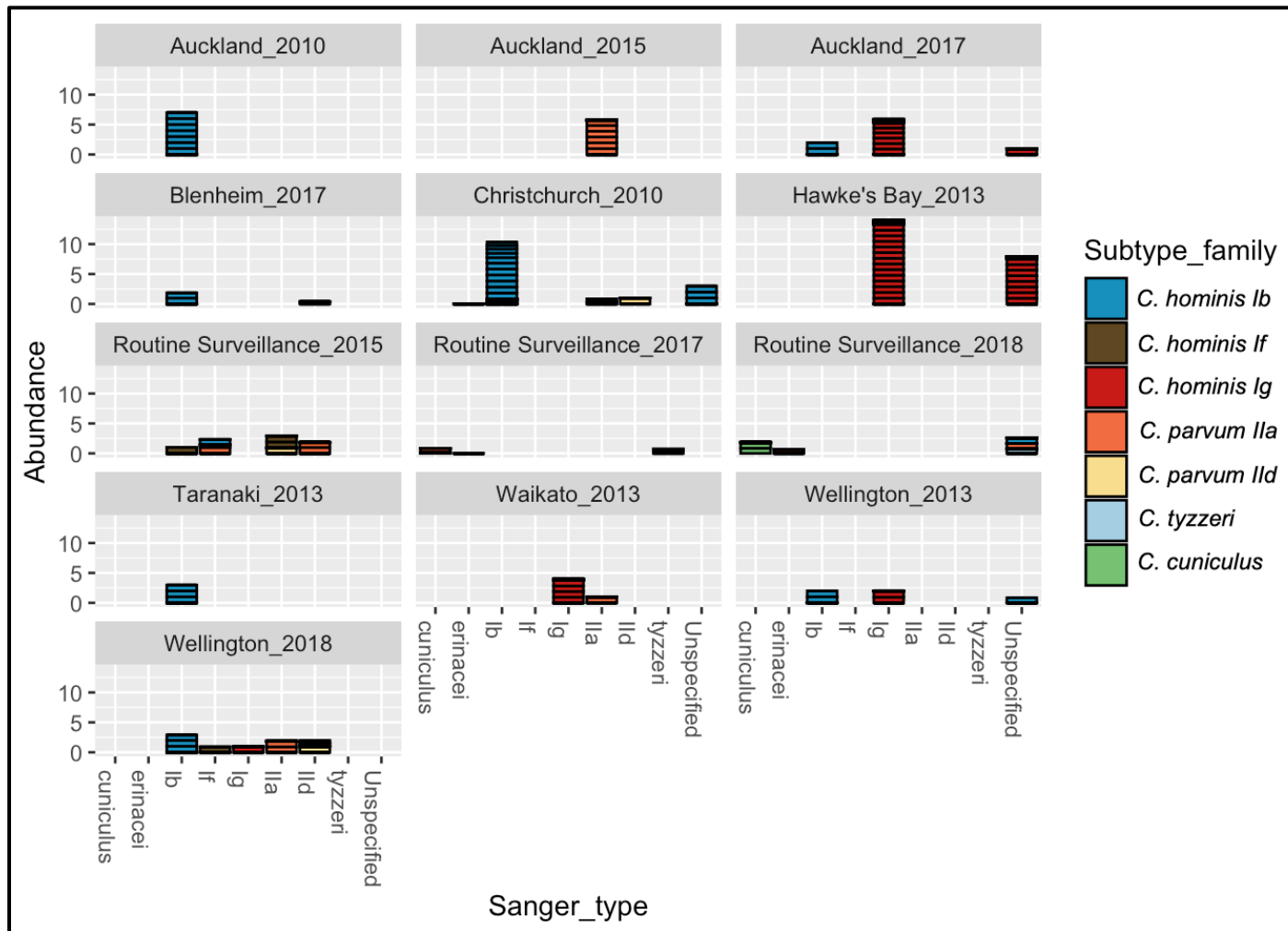
Figure 4.3. Exploratory bar plot showing the taxonomic distribution of subtype families in samples from each outbreak and from the samples collected as part of routine surveillance each year. Within each bar each line represents one sample, and the colour codes represent the most abundant subtype families.

4.4.3 Analysis of outbreaks

Figure 4.3. presents the diversity of variants of each of the subtype families found by NGS compared to the Sanger data from the same samples. For most of the outbreaks, the dominant NGS subtype families correspond to the Sanger data although extensive subtype diversity is present, as evidenced by the presence of multiple variants. In addition, the Sanger data for most of the outbreaks showed one dominant subtype that could be used to link the cases epidemiologically. Conversely, there are two outbreaks in which multiple dominant subtypes were present according to Sanger and NGS data, these will be discussed in the following paragraphs.

In the outbreak that occurred in Christchurch in 2010, of the 17 samples investigated 14 were classified by Sanger (Table 4.2). Of those 14, 11/14 of shared *C. hominis* subtype family Ib, 1/14 *C. parvum* IIa, 1/14 *C. parvum* IId, and 1/14 *C. erinacei*. Figure 4.4. displays a heatmap showing the relative abundance of the top 50 reads in each sample from the 2010 Christchurch outbreak according to the NGS data. Sample 1820, which initially was classified as IIa by Sanger sequencing, showed IbA10G2 as its most abundant sequence read (closely followed by IIaA19G4R1). The two samples (1821 and 1823) that had previously been classified as IId and *C. erinacei* respectively, displayed the same dominant subtype according to NGS. However, analysis of the heatmap (Figure 4.4.) shows that *C. hominis* IbA10G2 was present in all the samples from that outbreak, providing a shared subtype for these epidemiologically linked cases. In addition, 10/17 of the samples involved in that outbreak shared the same variant of *C. parvum* IId, IIdA17G1, providing two subtypes by which most of those cases could be linked.

The outbreak that occurred in Wellington in 2018 presents another interesting case. All 9 samples from this outbreak were classified at the *gp60* locus by Sanger sequence analysis (Table 4.2.). 3/9 were classified as *C. hominis* Ib, 1/9 *C. hominis* Ig, 2/9 *C. parvum* IIa, 2/9 *C. parvum* IId, and 1/9 *C. hominis* If. The NGS classification corresponded with all bar one of the samples (14840) in which it showed *C. hominis* Ig as the dominant subtype instead of *C. parvum* IId. This provides an example of an outbreak with no common subtype shared among all the samples according to consensus sequence analysis. From Figure 4.5. it is evident that although there are multiple dominant subtypes in each

sample, *C. hominis* IbA10G2 is abundant in all samples and could be used as evidence for a genetic link between these epidemiologically linked cases. In addition, *C. hominis* IgA20 is present in 6/9 samples, and *C. hominis* IfA12G1R5 is also present in 6/9 samples. The NGS abundance data for the rest of the outbreaks is available in Figures C.1. – C.8.
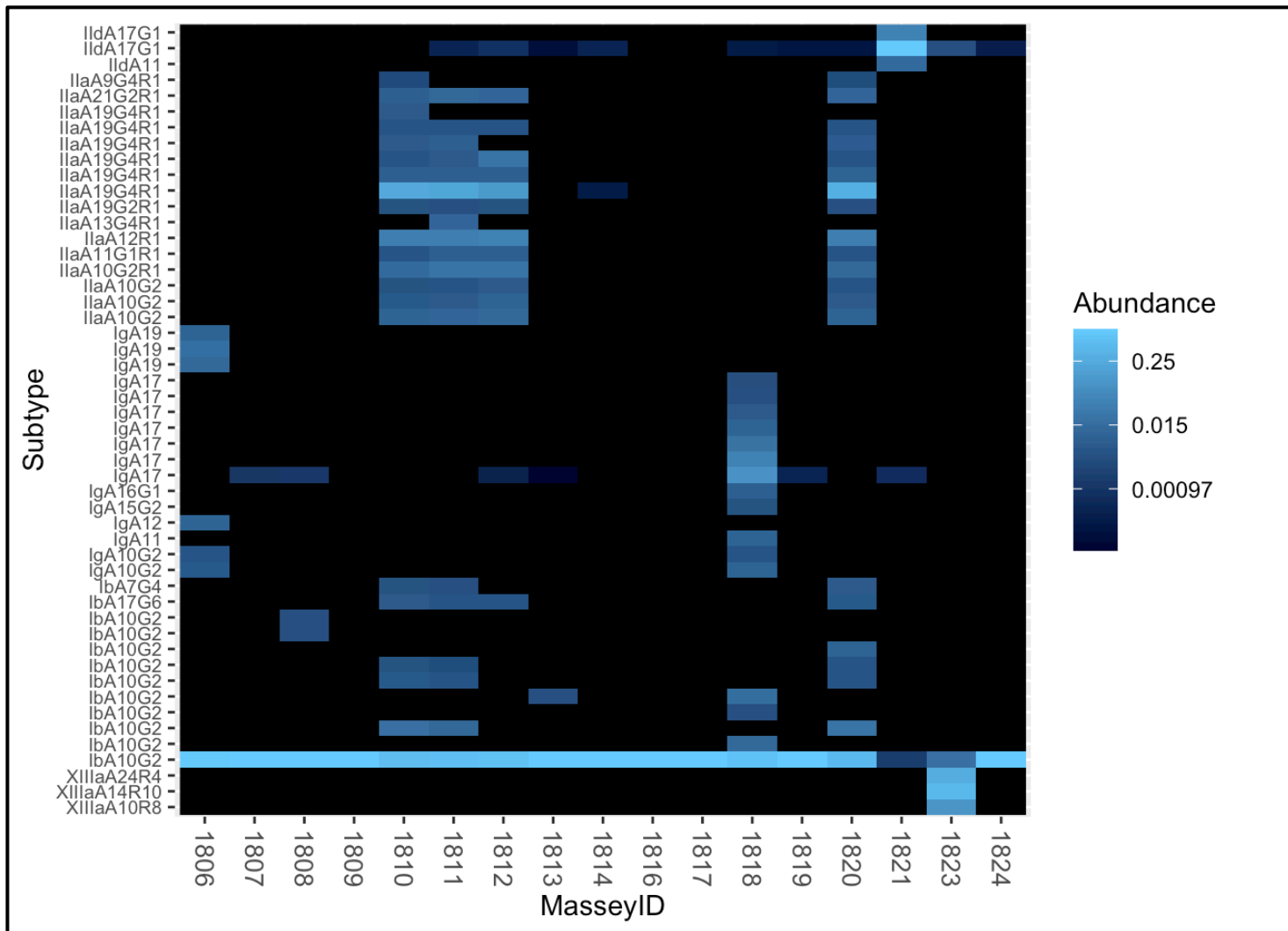
Figure 4.4. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Christchurch in 2010. The multiple subtypes present in each sample are displayed on the Y axis.
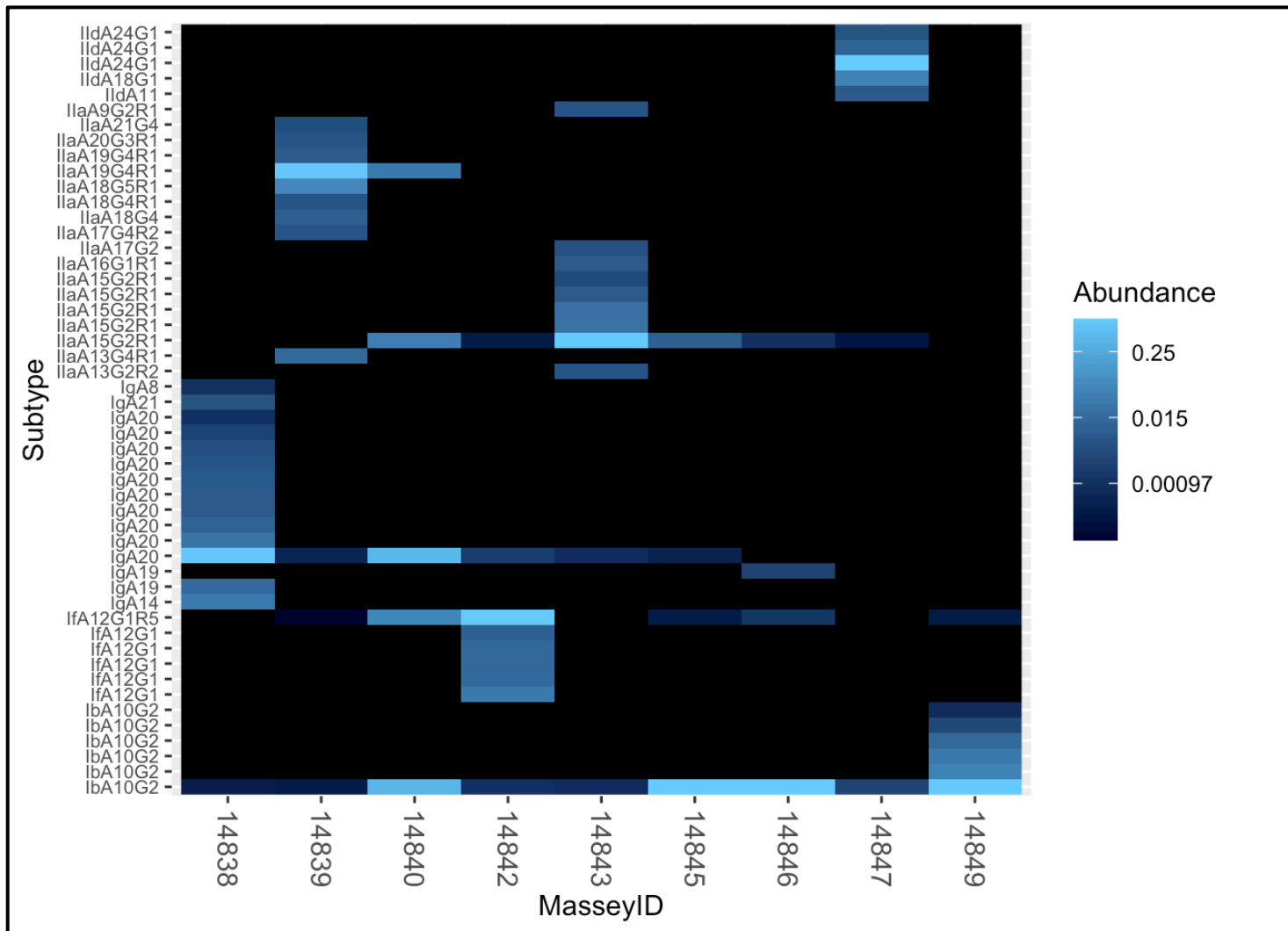
Figure 4.5. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Wellington in 2018. The multiple subtypes present in each sample are displayed on the Y axis.

4.5 Discussion

One of the aims of this study was to conduct a comparative analysis of samples from past sporadic cases and outbreaks of cryptosporidiosis in New Zealand to examine and contrast the efficiency of classification by Sanger sequencing and NGS at the *gp60* locus. An additional aim was to identify emerging or previously unidentified variants of *Cryptosporidium* that could have public health significance in the future. NGS data identified the presence of five dominant species (*C. hominis*, *C. parvum*, *C. cuniculus*, *C. tyzzeri* and *C. erinacei*) that have previously been found in New Zealand. The two most abundant subtypes identified in this study were *C. hominis* IbA10G2 (77.1%) and IgA17 (48.6%), which further suggests that although New Zealand fosters a substantial livestock industry, the dominant mode of transmission of *Cryptosporidium* in the country is anthroponotic. Similar to previous studies in New Zealand (Garcia-R et al., 2020; Garcia–R et al., 2017b), IIaA18G3R1was identified as the dominant subtype of *C. parvum* in this country (38.1%).

Cryptosporidium hominis IfA12G1R5 (15.2%) and IgA20 (21.9%) were identified as emerging subtypes in New Zealand, based on the number of samples in which they were present. The ability of NGS to sequence multiple reads in each sample allows for the uncovering of the hidden genetic diversity contained therein, which can serve as a method for the early identification of emerging variants of a pathogen before they become dominant in a country if it is adopted for the real-time surveillance of cryptosporidiosis. For this reason, and because of their prevalence in the samples analysed in this study, IfA12G1R5 and IgA20 are assigned the label of emerging subtypes in New Zealand. This is of particular importance because previous studies have documented the rise of IfA12G1R5 in the USA and Western Australia (Braima et al., 2019; Hlavsa et al., 2017), and the same could be gradually occurring in New Zealand. 14 out of 89 samples that had been classified by Sanger sequence analysis showed different dominant subtypes according to the NGS data. In all 14 samples, the subtype according to Sanger was still present in the samples but at lower levels than initially suggested. Further analysis showed evidence of significant intra-sample diversity in most of the samples involved in this study, with significant intra-species and subtype diversity observed from the NGS results. These results suggest, as has been hypothesised before

(Thompson & Ash, 2016), that *Cryptosporidium* infections in humans are usually mixed; either because of ingestion of genetically diverse oocysts, genetically diverse sporozoites within oocysts (potential results of sexual reproduction), or both. Advances are being made in the *in vitro* manipulation of *Cryptosporidium* and it is possible they could provide a clearer picture of the source of this diversity in individuals in the future. Taken together, the results outlined above *Cryptosporidium* infections in humans are frequently genetically diverse, and NGS is superior to consensus sequencing at capturing this diversity.

This study was limited by the omission of no-template and DNA extraction reagent controls which are usually used to assess the level of cross- and environmental contamination in a sequence run. The strategies taken to mitigate this limitation were the same as for the *Giardia* metabarcoding study and are outlined in full in chapter 3. Furthermore, nested PCR was used to specifically amplify *Cryptosporidium* DNA at the *gp60* locus, and the DADA2 algorithm used for sequence processing reports fewer false positives than other methods (Callahan et al., 2016). Only the top 50 sequences with high abundance across all samples were used to make inferences. From all the reads generated the bottom 1% were removed, and each outbreak showed different patterns and abundance of subtypes present, which suggest negligible levels of cross-contamination.

One of the problems hampering a full understanding of the epidemiology of cryptosporidiosis in New Zealand has been the inability to identify shared subtypes in epidemiologically linked cases. As a case studies, the outbreaks that occurred in Christchurch in 2010 and Wellington in 2018 were analysed in this investigation. Multiple dominant subtypes were present in the samples within each of those outbreaks according to consensus sequence analysis. By applying NGS analysis to the same samples *C. hominis* IbA10G2 was identified as the subtype present in all the samples from the outbreak that occurred in Christchurch in 2010. The samples from the outbreak in Wellington in 2018 were significantly more diverse with 5 different dominant subtypes being identified within the 9 samples involved in that outbreak. NGS analysis showed that once again *C. hominis* IbA10G2 was the subtype common to all the samples in that outbreak and showed that two other *C. hominis* subtypes (IgA20 and IfA12G1R5) were common across the majority of the samples involved in that study (6/9). This provides a

common species (*C. hominis*) link and a common subtype (IbA10G2) linking samples from that outbreak. Further work applying NGS analysis to outbreaks occurring in other countries could better our understanding of the epidemiology of this parasite and give a better understanding of its transmission patterns.

In conclusion, our study highlights the benefits of NGS analysis in the classification and characterisation of *Cryptosporidium* samples; capturing the broad genetic diversity present in individuals infected with the parasite, allowing the identification of shared subtypes in epidemiologically linked cases. It also showed the ability of this technique to identify emerging variants of *Cryptosporidium*, which advances our understanding of the epidemiology of this parasite in New Zealand and could help inform public health institutions as they seek to develop better strategies to combat this pervasive disease.

5 An *in vitro* platform for studying inter-species infectivity of *Cryptosporidium spp.*

5.1 Abstract

Cryptosporidium is a widespread protozoan pathogen that causes diarrhoea in the majority of vertebrate hosts. Previous studies have shown that different species can produce different symptoms in humans infected with this parasite. In addition, particular genotypes within species are more transmissible than others. However, the mechanisms underpinning these differences are not fully understood. Despite the burden of cryptosporidiosis on public health, there are few treatments and no vaccines available. The development of new and efficient *in vitro* systems for the culture and manipulation of *Cryptosporidium* would be a useful tool that would spur advances in our understanding of the disease mechanisms of this parasite and the development of effective therapies. This study lays the foundation of a novel *in vitro* assay for the analysis of the infectivity of *Cryptosporidium parvum* and *Cryptosporidium hominis* utilising the COLO-680N cell line and flow cytometry. It suggests that the assay is effective with and without the use of a fluorescent antibody.

5.2 Introduction

Cryptosporidiosis is a globally ubiquitous disease caused by infection with the parasite *Cryptosporidium* affecting humans, domestic animals and wildlife. The effects of the disease can be more severe in immunocompromised individuals and children under five years of age, which has led to *Cryptosporidium* being classified as the second most common cause of diarrhoea in the aforementioned age group (Savioli et al., 2006b). Currently, nitazoxanide is the only drug approved by the Food and Drug Administration of the United States (FDA) for the treatment of cryptosporidiosis (Manjunatha et al., 2016). However, this drug only partially alleviates the symptoms of the disease, further, there are no effective vaccines.

This study focuses on the infectivity of *Cryptosporidium hominis* and *Cryptosporidium parvum* in humans. The main mode of transmission of this parasite is the faecal-oral route, however, recent evidence has shown that the infection can be respiratory in humans and animals, invade the pancreatic and biliary systems, and in rarer cases, lead to cerebral pathologies and cancer (Audebert et al., 2020; Gaber et al., 2020; Sponseller et al., 2014). The most common symptoms of infection are acute diarrhoea and abdominal pain (Bones et al., 2019). However, symptoms such as nausea, vomiting, fever, nutrient malabsorption and growth retardation (in children) have been reported in immunocompetent individuals (Feng & Xiao, 2017; Tumwine et al., 2003); severe malabsorption syndrome leading to mortality can occur in immunocompromised individuals (Chalmers & Katzer, 2013). Of the 38 currently accepted species of *Cryptosporidium*, *C. hominis* and *C. parvum* are the two responsible for most infections in humans. Previous studies have found that the species of *Cryptosporidium* a person is infected with can have an effect on the symptoms or sequelae they experience, for example, eye pain and persistent headaches are symptoms associated with *C. hominis* infections but not with *C. parvum* (Hunter et al., 2004a). More work needs to be done to understand the mechanisms underpinning these differences in clinical presentations between species of *Cryptosporidium*.

Furthermore, the evidence from recent studies suggests that different genotypes within species can make different contributions in varied settings. For instance, one study showed that the virulent *C. hominis* subtype family Ib is the main causative agent of cryptosporidiosis in high-income countries (HICs), such as those in Europe and North America, and is responsible for most outbreaks worldwide (Khan et al., 2018b). In addition, the hyper-transmissible genotype of *C. parvum*, IIaA15G2R1, has been widely reported in the literature (Xiao, 2010). Focusing on waterborne transmission of *Cryptosporidium*, a study by Gilchrist et al., (2020) suggests that the bottleneck created by the dry season in places such as Bangladesh could result in the selection of mutations that lead to an increase in infectivity of the parasite over time.

The majority of the discoveries highlighted above were made using *in vivo* systems. These systems, when compared to *in vitro* culture and manipulation, are relatively more complex and require expensive setups and frequent monitoring in a laboratory

environment. Moreover, they require the use of animals in experiments. The lack of efficient *in vitro* systems for the study of *Cryptosporidium* is largely due to the difficulty of culturing this parasite in cell lines. The main cell lines that have been used for the culture of the parasite are Caco-2 and HCT-8, nevertheless, infection of these cell lines causes rapid death of the host cell and does not allow for the progression of the parasite through its complete life cycle (Manjunatha et al., 2016). Also, these cell lines show a preference for the culture of *C. parvum* over *C. hominis*. For this reason, more complex setups, such as the use of hollow fibre technology to augment cell culture have been proposed and implemented (Morada et al., 2016).

The development of efficient and long-term *in vitro* systems for the culture and manipulation of *Cryptosporidium* would be a useful tool to help fill the gaps in our understanding of the difference in host-pathogen interaction between and within species of the parasite, as outlined above. With regards to the epidemiology of cryptosporidiosis, questions remain as to whether contact rates or species/genotype determine case numbers in outbreaks or sporadic events. Despite the limitations of current *in vitro* techniques, over 25 presumptive virulence factors have been characterised in myriad processes connected to adhesion, invasion, locomotion, and proliferation (Casadevall & Pirofski, 2001), however, their roles are not yet fully understood. Some effort has been made to address the gap in capability. Transfection of *C. parvum* sporozoites using CRISPR/Cas9 technology has been reported by Vinayak et al., (2015).

Flow cytometry is a laser-based method used to analyse the physical and chemical characteristics of particles. Identification is usually achieved through the use of fluorescent antibodies. However, because this technique is particularly efficient at distinguishing particles from the noise or other artefacts the use of a fluorescent antibody is not always required. Previous studies have used flow cytometry for the quantification and identification *Cryptosporidium* oocysts with and without fluorescent antibodies (Shams et al., 2016). Vitelli et al., (2021) provide a good background of the past and current uses of flow cytometry.

Recently, Miller et al., (2018) characterised a new cell line, COLO-680N, which allows for the long-term cultivation of *Cryptosporidium* through its entire life cycle. COLO-680N is a

human oesophageal carcinoma epithelial cell line. This is remarkable because most of the previously described cell lines used to culture *Cryptosporidium* are derived from intestinal cells, however, as previously mentioned in the article by Josse et al., (2019), it is now understood that the parasite can infect multiple systems within the body so casting a wide net to find more suitable cell lines has the potential to greatly benefit the field.

In this study, infectivity is defined as the capacity of a pathogen to infect a susceptible host and complete its life cycle. By taking advantage of recent advances in the culture of *Cryptosporidium* the foundation of an *in vitro* system capable of assessing the infectivity of *C. parvum* and *C. hominis* has been developed. The system uses the COLO-680N cell line to culture the parasite, then the widely used *Cryptosporidium*-specific fluorescent dye, Sporo-glo™, in conjunction with flow cytometry is employed to assess levels of infectivity in *Cryptosporidium* species.

5.3 Methods

5.3.1 *Cryptosporidium* samples

The *C. parvum* and *C. hominis* oocysts used in this study were obtained from infected humans in New Zealand. The anonymised faecal samples were sent to our laboratory from diagnostic labs all across New Zealand due to our collaboration with the Ministry of Health. All faecal samples were stored at 4°C before use. The species and genotype of each sample was determined using PCR at the glycoprotein 60 (*gp60*) locus and Sanger sequencing.

5.3.2 Purification of oocysts

Cryptosporidium oocysts were purified using a modification of the methods described by Meloni & Thompson (1996). Briefly, 0.5%, 1.0% and 2% (w/v) Ficoll 400 (Merck KGaA, Darmstadt, Germany) solutions were prepared in PBS and stored at 4°C. To prepare the

gradient 1.5ml of Ficoll solution was layered using a pipette in a 2 ml safe-lock tube with the 2% solution at the bottom. 500 µl of oocyst solution was layered on top of the cold Ficoll gradient (4°C) and the tubes were centrifuged for 20 min at 1,500 × g at room temperature (RT). The interface between 0.5% and 1% was transferred to a 15 ml centrifuge tube, made up to 15 ml with cold PBS (4°C) and washed by centrifuging for 5 min at 2,000 × g at 4°C. After washing, the supernatant was removed leaving 1 ml of the PBS. The remainder containing the purified oocysts were transferred to a new 1.5 ml microcentrifuge tube. 15 µl of an antibiotic solution composed of 5 mg/ml gentamycin, 4 mg/ml lincomycin and 10 mg/ml ampicillin was added to the purified oocysts suspension before storage at 4°C.

5.3.3 Excystation of oocysts

Two methods of oocyst excystation were employed in this study. Initially, excystation of oocysts was performed according to the method outlined by Rasmussen et al., (1993). This method was selected based on a comparative study carried out by Pecková et al., (2016) and involves the incubation of oocysts in an excystation solution composed of 0.75% taurocholic acid and 0.25% trypsin in PBS for 45 min at 37°C.

A second method was implemented because the one outlined above proved to be inefficient. This method involved the incubation of oocysts at 37°C in an excystation solution composed of 0.8% taurocholic acid in PBS for 2 hrs (Petry & Harris, 1999).

5.3.4 Cell Culture

Human ileocecal colorectal adenocarcinoma (HCT-8) cells (ATCC, CCL244) and Human oesophageal squamous cell carcinoma (COLO-680N) cells (DSMZ Germany, ACC182) were maintained in 75 cm^2 tissue culture flasks incubated at 37°C in a humidified incubator with 5% CO_2 (Laurent et al., 1997; Miller et al., 2018). For HCT-8 cells the growth medium consisted of RPMI 1640 with L-glutamine (Thermo Fisher Scientific, Waltham, Massachusetts, United States) supplemented with 10% horse serum, 100 U/ml of penicillin, 100 µg/ml of streptomycin and 250 ng/ml of amphotericin B. These were passaged every 2-3 days, or when the cells reached 80% confluency, using 0.25% trypsin-

EDTA (Thermo Fisher Scientific, Waltham, Massachusetts, United States) to lift the cells off the flask. COLO-680N cells were maintained in a growth medium consisting of RPMI 1640 supplemented with 10% foetal bovine serum (FBS), 100 U/ml of penicillin, 100 µg/ml of streptomycin and 250 ng/ml of amphotericin B. This cell line was cultured and maintained according to the guidelines provided by Jossé et al. (2019). Counting for both cell lines was carried out by mixing 10 µl of cell suspension with 50 µl trypan blue and 40 µl PBS in a 1.5 ml microcentrifuge tube. Approximately 15 µl of this mixture was added to a haemocytometer and used to measure the concentration of cells in the suspension.

5.3.5 <u>Assessment of Immune Response from COLO-680N cells</u>

5×10^4 COLO-680N cells were seeded into an 8 well chamber slide (Thermo Fisher Scientific, Waltham, Massachusetts, United States) and left to grow for 24 hrs at 37°C in a humidified incubator with 5% CO_2. Once the cells had grown to >70% confluency, half of the wells were inoculated with 4.25×10^5 *C. parvum* sporozoites each, while the rest were left untreated as controls. Then the chamber slide was returned to the incubator for another 12 hrs. 3 µg/ml Brefeldin A (Invitrogen, Carlsbad, California, United States) was added to each well to block protein transport and encourage the accumulation of cytokines and chemokines within the cell, then the chamber slide was allowed to accumulate for another 12 hrs before staining. Before staining, all the cell culture medium was removed from each well using a pipette, then the cells were washed once with 300 µl of PBS. The cells were fixed by adding 2% paraformaldehyde in PBS to each well and allowing them to incubate for 20 min at room temperature (RT). Post-incubation the liquid was removed and 150 µl of 0.1% Triton® X-100 (Invitrogen, Carlsbad, California, United States) was added to each well and left to incubate for 10 min at RT. After this incubation, the liquid was removed from each well and the cells were washed with 300 µl of 1% bovine serum albumin (BSA) blocking solution. To begin the staining process, 60 µl DAPI solution (2 mg/ml) was added to each well and left to incubate for 30 min at RT in the dark. Then the cells were washed twice with 300 µl BSA, as previously described. After washing, the cells were stained with 100 µl of Anti-Human IL-8 APC (eBioscience, San Diego, California, United States) and left to incubate for 45 min at RT in the dark. After staining, the cells were washed twice with BSA. All the liquid was discarded from each well and the chamber removed. Mounting medium was added

dropwise to the slide and then covered with a cover slip. The slide was delivered to Massey Microscopy and Imaging Centre for fluorescence microscopy, and the images were processed using ImageJ software (Schneider et al., 2012).

5.3.6 *Salmonella typhimurium* culture

Salmonella typhimurium was cultured according to the procedure outlines by Abernathy et al., (2013). Briefly, overnight, *S. typhimurium* grown at 37°C in Luria-Bertani (LB) broth was sub-cultured for 3 h in pre-warmed (37°C) LB broth before challenge to ensure log-phase cultures. COLO-680N cells were challenged with *S. typhimurium* at a multiplicity of infection (MOI) of 100 by replacing media with *Salmonella* infectious media. Immediately upon challenge, plates were centrifuged at 500 × g for 5 min and placed in a CO_2 incubator and allowed to incubate for 3 hrs before harvest.

5.3.7 Infectivity Assay

To begin, cultures of HCT-8 and COLO-680N cells were passaged and seeded onto 12-well plates at a concentration of 2.5×10^5 cells/well. The cells were allowed to proliferate for 24 hrs at 37°C in a humidified incubator with 5% CO_2. 24 hrs later, oocysts purified from human isolates were excysted according to the method outlined above. *C. parvum* and *C. hominis* oocysts were excysted at a multiplicity of infection (MOI) of 40 (2.5×10^6 excysted oocysts/well) and 50 (3.2×10^6 excysted oocysts/well) for each species. The MOI took into account that each oocyst contains 4 sporozoites. The excysted oocysts were spun down at 13,200 × g for 3 min and resuspended in growth medium. Then 100 μl of the excysted oocysts suspension, corresponding to the amount of excysted oocysts required to reach the desired MOI, was added to the relevant wells in each 12-well plate. The plates were spun at 188 × g for 7 min to encourage invasion then returned to the incubator for 48 hrs to allow for the proliferation of the parasites within the cells. 48 hrs later, the growth medium was removed from each well, then wells were washed twice with 1 ml of PBS (Thermo Fisher Scientific, Waltham, Massachusetts, United States) to remove the excess sporozoites, oocysts and oocyst shells. All washes were collected and stored for further analysis of parasite numbers.

To harvest the cells, 300 µl 0.25% trypsin-EDTA was added to each well, then the plates were returned to the incubator for 13 min to encourage dissociation. Following this, 700 µl of growth medium was added to each well to inactivate the 0.25% trypsin-EDTA then transferred to 1.5 ml microcentrifuge tubes and spun for 3 min at 400 × g at RT. After centrifuging, the supernatant was discarded, and the cells were resuspended in 1 ml of PBS. Before the washing and harvesting of cells began, 2.5×10^6 oocysts from each species was excysted, washed and resuspended in PBS.

To prepare the samples for staining and further analysis all samples were fixed. All the cell samples were spun down at 400 × g for 3 mins at RT, the supernatant was discarded, the samples were resuspended in 100 µl of eBioscience™ Fixable Viability Dye eFluor™ 780 (Thermo Fisher Scientific, Waltham, Massachusetts, United States), and were incubated for 30 min at 4°C in the dark. The addition of this dye before fixing allows for the measurement of the number of viable cells in each sample (Figure D.1.). After incubation, the cell samples were spun down at the aforementioned settings and washed twice in 500 µl of PBS. All samples, including the sporozoites were spun down (cells at 400 × g for 3 min at RT, sporozoites and oocysts at 12,100 × g for 3 min at RT), resuspended in 500 µl of 0.22 µm filter sterilised FluoroFix Buffer (BioLegend, San Diego, California, United States), and incubated for 30 min in the dark at RT. Following incubation, all samples were spun down, the supernatant discarded, and washed once in 500 µl of PBS. After the final wash, all samples were spun down, resuspended in 500 µl of PBS and stored at 4°C, protected from light, before further manipulation and analysis, which usually occurred within 24 – 48 hrs.
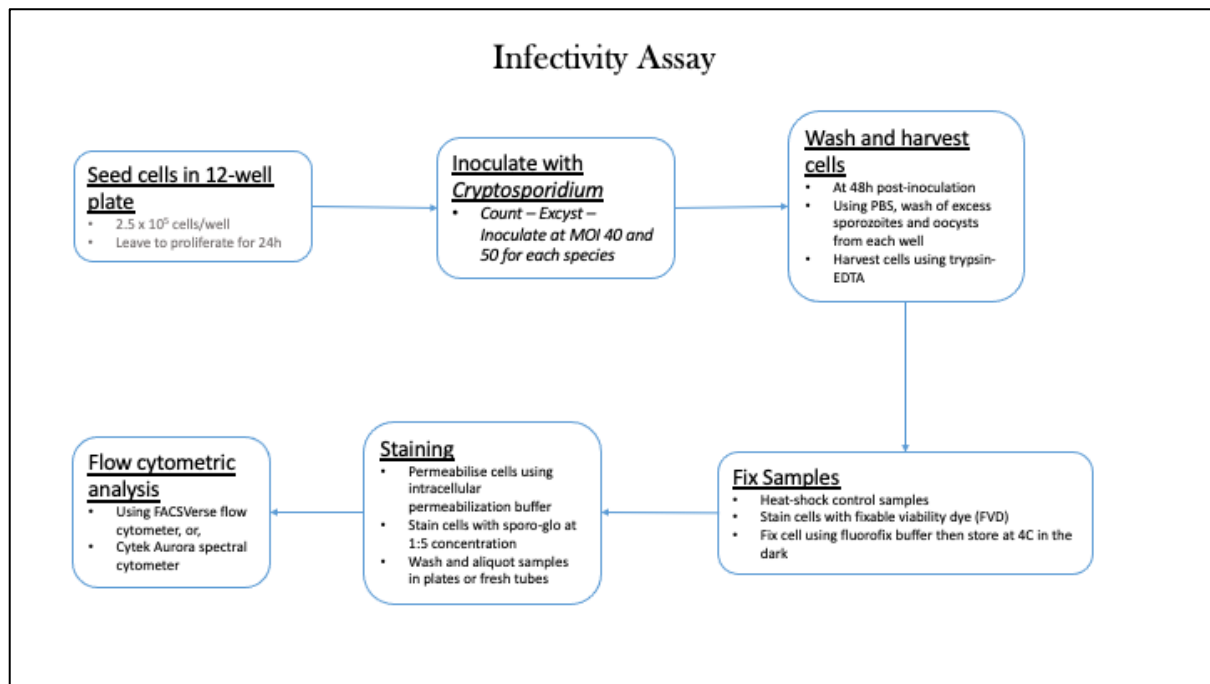
Figure 5.1. A flow diagram showing the procedure for the infectivity assay. The main processes carried out in each step of the infectivity assay are shown in each bubble.

5.3.8 Flow Cytometry

All fixed samples were put through flow cytometric analysis, but only a specific subset was stained (unstained infected and uninfected cells, oocysts and sporozoites were used as controls). The relevant fixed samples from the infectivity assay due to be stained were spun down (cells at 800 × g for 3 min at RT, sporozoites and oocysts at 12,100 × g for 3 min at RT) and washed twice in 500 µl of 0.22 µm filter-sterilised 1X Intracellular Permeabilization Buffer (IPB) (BioLegend, San Diego, California, United States), taking care to disrupt the pellet between each wash. The centripetal force was increased to 800 × g for the cells post-fixing because they became lighter and easier to disrupt after being treated with a fixative. After washing, the supernatant was discarded, the samples were resuspended in 500 µl of 0.22 µm filter-sterilised blocking buffer (5% Fetal Calf Serum in PBS), and incubated for 30 min in the dark at RT. After incubation in blocking buffer, the samples were spun down, the supernatant discarded and washed twice in 200 µl of IPB. After the final wash in IPB the samples were resuspended in 200 µl of Sporo-Glo™ (anti-Spor FITC) (Waterborne Inc., New Orleans, Louisiana, United States), an antibody targeting the intracellular life cycle stages of *C. parvum* (Figure 1.1.) conjugated to fluorescein isothiocyanate (FITC). The antibody was diluted to from its stock

concentration (1X) to a concentration of 1:16 in IPB to reduce the amount of background binding of the antibody to uninfected cells. Then the samples were incubated for 1 hr in the dark at RT. Following incubation in anti-Spor FITC, the samples were spun down (as above), the supernatant discarded, and washed twice in PBS. After the final wash of the stained samples, all samples to be analysed by flow cytometry were spun down and resuspended in 400 μl of PBS. 200 μl of each sample was aliquoted into two separate V-bottomed 96-well plates. Unstained and uninfected cells were used a negative-infection control, *S. typhimurium*-infected cells were included as positive infection controls, sporozoites stained with the anti-Spor FITC were included as a positive stain control, and heat-shocked (dead) cells were included as a viability control. Samples of PBS, IPB and blocking buffer were included as reagent controls. The plates were delivered to Joanna Roberts at FlowJoanna (Palmerston North, New Zealand) for analysis. The plates were analysed using a FACSVerse™ flow cytometer (BD biosciences, San Jose, California, United States), In a repeat experiment, they were analysed using a Cytek® Aurora spectral cytometer (Cytek biosciences, Fremont, California, United States) (Figures D.19 – D.29). The results from the FACSVerse™ were analysed using the FlowJo® software (BD biosciences, San Jose, California, United States), and the results from the Aurora were analysed using the SpectroFlo® software (Cytek biosciences, Fremont, California, United States).

5.4 Results

5.4.1 Comparison of HCT-8 and COLO-680N cell lines after infection with *Cryptosporidium*

Initially, the infectivity assay and flow cytometry analysis was conducted using the HCT-8 cell line and *C. parvum* oocysts. This was done because HCT-8 is the most common cell line used for the *in vitro* investigations of *Cryptosporidium*, and *C. parvum* is the species found to work best with this cell line. The data from the flow cytometer showed that the number of viable cells in infected cultures was too low to make any assessment of the

success of infection i.e., the low number of cells meant there was little for the fluorescent dye to bind to, resulting in a low signal beyond the limit of detection (Table 5.1.).

Table 5.1. Total counts of viable HCT-8 cells detected in *Cryptosporidium*-infected samples. VOL (µl) indicates the volume of the sample that was entered into the flow cytometer. The number of viable mammalian cells within that volume is shown and the number of cells/ml in the original sample is used to assess viability.

CELL TREATMENT	VIABLE MAMMALIAN CELL COUNTS	VOL (µL)	CELLS/ML
HCT-8 INFECTED MOI-1 (A)	3.00	48	62.50
HCT-8 INFECTED MOI-1 (B)	9.00	48	187.50
HCT-8 INFECTED MOI-1 (C)	0.00	48	0.00
HCT-8 INFECTED MOI-2.6 (A)	3.00	48	62.50
HCT-8 INFECTED MOI-2.6 (B)	13.00	48	270.83
HCT-8 INFECTED MOI-2.6 (C)	10.00	48	208.33
HCT-8 INFECTED MOI-2.6 (D)	4.00	47	85.11
HCT-8 INFECTED MOI-2.6 (E)	11.00	47	234.04
HCT-8 INFECTED MOI-2.6 (F)	8.00	48	166.67
HCT-8 HEAT-SHOCKED	1148.00	18	63777.78
HCT-8 HEAT-SHOCKED	792.00	16	49500.00
HCT-8 UNTREATED	1648.00	46	35826.09
MEAN	**304.08**	**43**	**12531.78**

Initial tests using the same procedure as with HCT-8 cells was conducted using the COLO-680N cell line. A two-fold increase in the number of viable cells (going by the means from table 5.1. and 5.2.) was observed, though a higher dose of *C. parvum* and *C. hominis* oocysts was used (Table 5.2.). Initially, during the infectivity assay, excess oocysts and sporozoites were washed off from each well 24 hrs post-inoculation and the growth medium was replaced. This premature change of medium caused increased mortality. The number of viable cells was significantly increased when the procedure was altered and the excess oocysts and sporozoites were washed off and collected at the point of harvesting and fixation.

Table 5.2. Total counts of viable COLO-680N cells in *Cryptosporidium*-infected samples. Noticeably more viable cells when compared to HCT-8.

CELL TREATMENT	VIABLE MAMMALIAN CELL COUNTS	VOL (µL)	CELLS/ML
PARVUM-INFECTED MOI 8	79.00	48	1645.83
PARVUM-INFECTED MOI 8	751.00	48	15645.83
PARVUM-INFECTED MOI 8	230.00	48	4791.67
PARVUM-INFECTED MOI 40	174.00	48	3625.00
PARVUM-INFECTED MOI 40	88.00	48	1833.33
PARVUM-INFECTED MOI 40	747.00	48	15562.50
HOMINIS-INFECTED MOI 8	700.00	48	14583.33
HOMINIS-INFECTED MOI 8	709.00	48	14770.83
HOMINIS-INFECTED MOI 8	607.00	48	12645.83
HOMINIS-INFECTED MOI 40	235.00	48	4895.83
HOMINIS-INFECTED MOI 40	416.00	48	8666.67
HOMINIS-INFECTED MOI 40	249.00	48	5187.50
UNTREATED	4723.00	48	98395.83
MEAN	746.77	48	15557.69

5.4.2 <u>Detection of immune response from COLO-680N</u>

To further verify that the COLO-680N cell line was suitable for our system i.e., that infection elicits an immune response in the host cells, an immunofluorescence assay was conducted. An analysis of previous literature on the subject of host immune responses as a result of *Cryptosporidium* infection identified interleukin-8 (IL-8) as a suitable candidate to use to assess the immune response from COLO-680N cells (Deng et al., 2004; Di Genova & Tonelli, 2016; Kothavade, 2011). IL-8 is a well-documented proinflammatory cytokine, produced by a wide variety of cells, and its fluorescent antibodies are easy to procure and handle. The studies referenced above showed an increased IL-8 response to infection (determined by quantitative RT-PCR and ELISA) in human epithelial cell lines such as HCT-8, murine intestinal cells and xenografts. Microscopic analysis of COLO-680N cells infected with *C. parvum* and stained with an anti-IL-8 fluorescent antibody showed increased secretion of IL-8 compared to controls (Figure 5.2.).
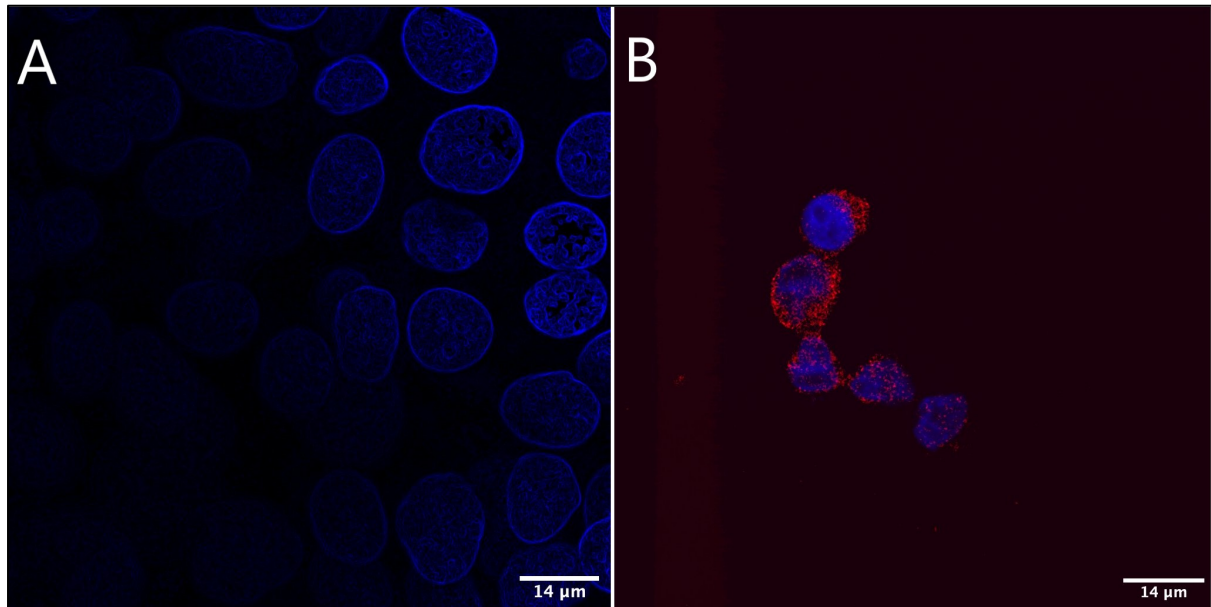
Figure 5.2. Confocal microscopy image of *C. parvum*-infected COLO-680N cells expressing IL-8. Uninfected cells are displayed in A, infected cells are displayed in B. The nucleus is stained blue with DAPI, and the red/orange dots represent IL-8 molecules tagged with a fluorescent antibody. 72 days old *C. parvum* IIaA18G2R1 excysted oocysts were used for infection in this experiment. Addition of Brefeldin A 12 hrs before fixing resulted in some cell death and shape distortion in some cells. Scale bar represents 14 μm.

5.4.3 Characterisation of *Cryptosporidium* infection using flow cytometry

A previous study demonstrated the possibility of identifying and quantifying *Cryptosporidium* oocysts by flow cytometry without the aid of a fluorescent antibody (Sonzogni-Desautels et al., 2019). Samples were prepared and submitted for analysis to establish the efficacy of our system and prepare for flow cytometric analysis of the infectivity assay control. Figure 5.3. shows the gating strategy for the identification cells and *Cryptosporidium* life cycle stages. Heat-shocked cells were used to create a baseline allowing for the distinguishing dead from viable cells (Figure D.1.). After exclusion of debris and dead cells, a gate covering a population of viable cells was established (Figure 5.3.A). Unstained oocysts and sporozoites were used to establish the characteristics of the extracellular life cycle stages of *Cryptosporidium* (Figure 5.3.B-C). Gating of the stages was made possible due to their distinct and regular size and shape. During this analysis a population of what appeared to be the shells of excysted oocysts were identified and

gated as well, and as expected, there was evidence of some spontaneous excystation seen by the population of sporozoites present in the oocyst sample. The oocysts were stored at 4°C after purification from the faecal samples. The sudden temperature change when they were retrieved for the flow cytometric assay triggered excystation in a small proportion of the oocysts. The anti-Spor FITC antibody was used in this study to identify sporozoites and other intracellular life cycle stages of *Cryptosporidium*. Figure 5.3.D shows the flow cytometric plot for samples containing stained. They are clearly gated due to their forward scatter (FSC) and side scatter (SSC) characteristics in the FITC channel, and the antibody has bound to populations of sporozoites, oocysts, and oocysts shells. Figure 5.4. shows classification of cell and parasite life cycle stages in an infected sample. The cells form a population distinct from the extracellular life cycle stages (sporozoites, oocysts, and oocyst shells) of the parasite. These extracellular life cycle stages are to be expected due to the inefficient nature of *in vitro* excystation and the impossibility of removing absolutely all oocysts from the sample before harvesting and fixing, despite multiple washes. Further evidence of our gating and noise exclusion strategies can be found in Figure D.2. – D.6.
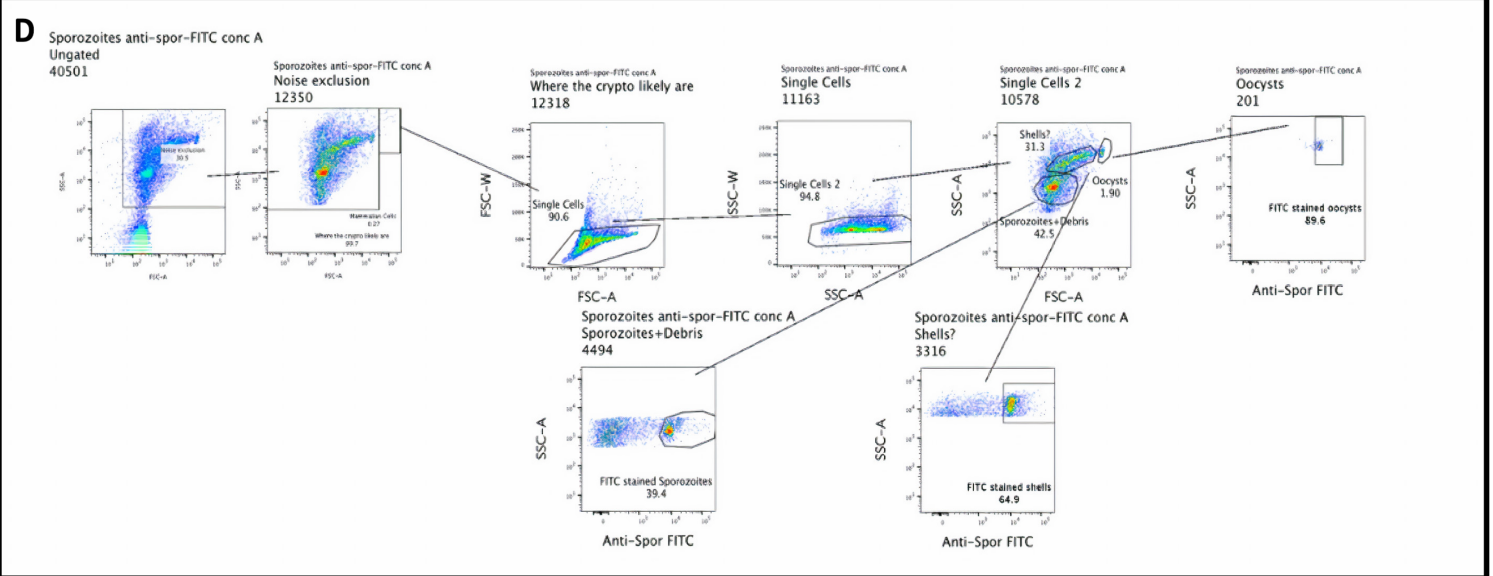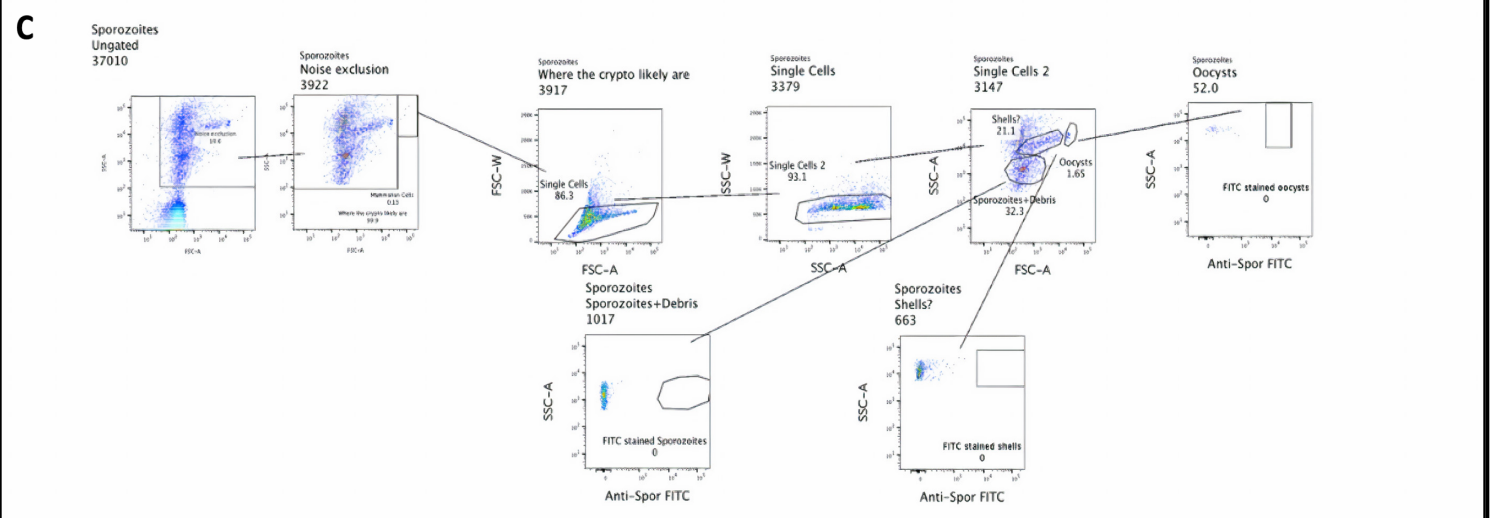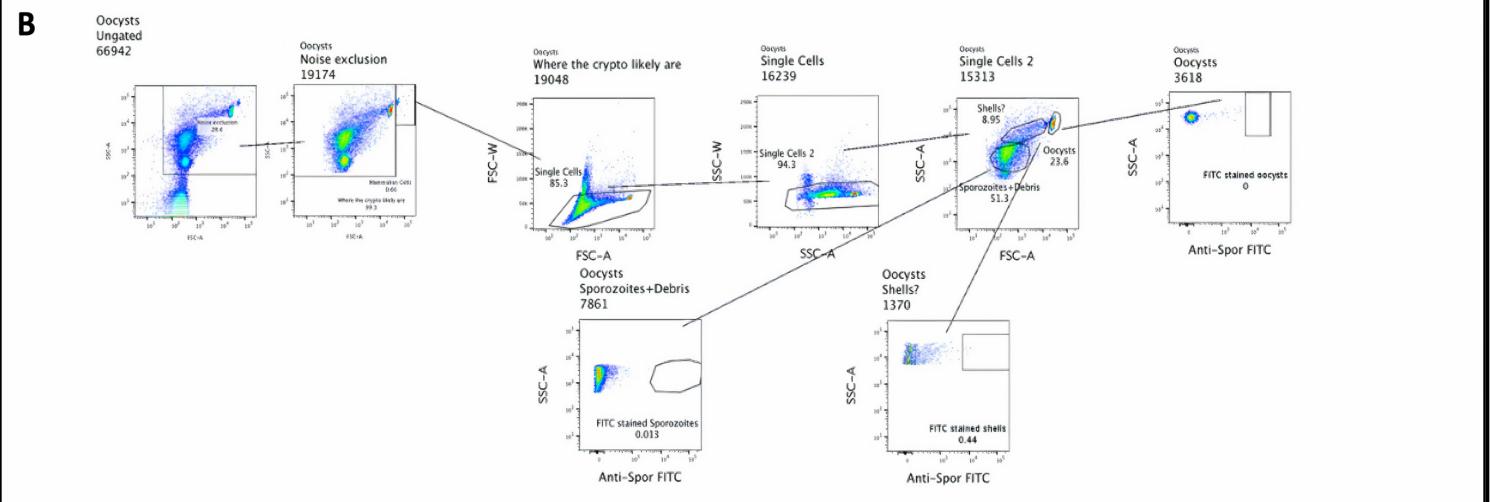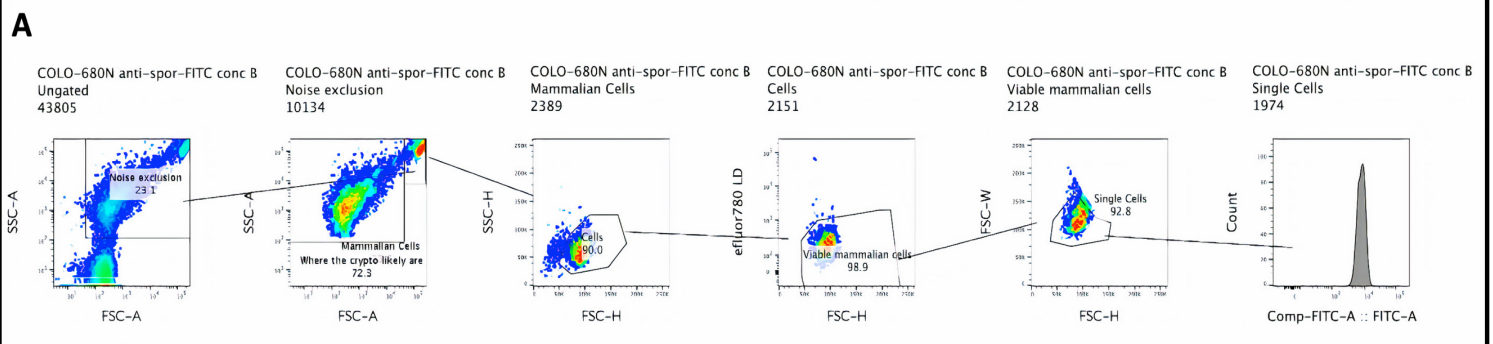
Figure 5.3. Classification of cell and particle types based on log forward scatter and log side scatter. To characterise both *Cryptosporidium* samples and mammalian cells, two gating hierarchies were generated – one for the *Cryptosporidium* alone samples and one for the mammalian cells. (A) Gating hierarchy for mammalian cells. The signature of buffers was used to exclude noise, then dead cells were excluded to keep only viable cells in gated region. (B) Gating hierarchy for *Cryptosporidium* oocysts. Unstained oocysts were used in this acquisition. Some sporozoites and oocysts shells were detected in this sample. This was possibly due to spontaneous excystation brought on by movement of sample from cold storage to warmer laboratory environment. (C) Gating hierarchy for unstained sporozoites. Some oocysts present in this sample because excystation is never 100% effective. (D) Gating hierarchy for anti-Spor FITC-stained sporozoites. This sample shows evidence of anti-Spor FITC-stained sporozoites, oocyst shells, and intact oocysts.
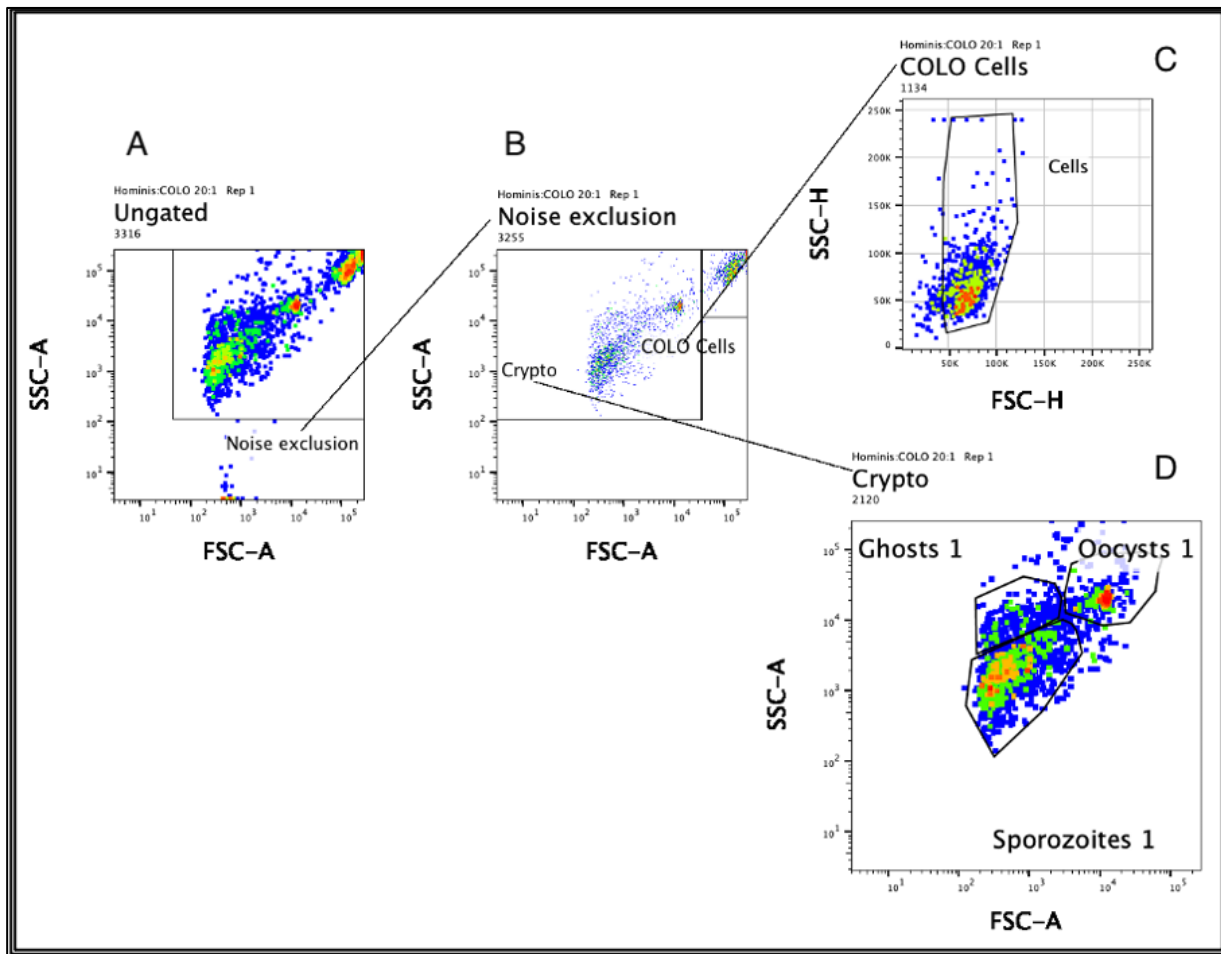
Figure 5.4. Detection of *Cryptosporidium* in infected COLO-680N cells. *C. hominis*-infected COLO-680N sample used as an example to show identification of COLO-680N cells (COLO) and *Cryptosporidium* life cycle stages on scatter dot plots. This figure shows the system was able to distinguish between COLO cells (C)and *Cryptosporidium* life cycle stages (D) in a single sample (A-B). N.B. the COLO gate in the middle panel is the box on the top right corner.

5.4.4 Quantification of infectivity using anti-Spor FITC antibody

Figure 5.5. shows the flow cytometric results from the infectivity assay. The cells infected with *C. hominis* sporozoites at an MOI of 8 (Figure 5.5.F) did not have a noticeable population of cells expressing the FITC antibody (approximately 0 – 1%) so infection could not be reliably quantified in all 3 replicates of that treatment. However, when the

dose was increased (Figure 5.5.G), it was possible to distinguish small but noticeable populations of *C. hominis*-infected cells – approximately 4 to 7% - expressing the FITC antibody. When cells were infected with *C. parvum,* noticeable populations of anti-Spor FITC expressing cells were found in both treatments (MOI 8 and 40) (Figure 5.5. D & E). Approximately 5 – 7% of cells in each of the triplicates expressed the FITC antibody at MOI 8 and 13 – 21% cells when the dose was increased to an MOI of 40. The positive infection control (*S. typhimurium*-infected cells) (Figure D.27.) showed no expression of the anti-Spor FITC antibody. The results of the experiment when tried at MOI 50 are shown in Figures D.13 - D.17. Table 5.3. shows the amount of variation in expression of the FITC antibody between replicates. Possible reasons for the variation observed include differing age of oocsts used and differences the efficiency of excystation between *Cryptosporidium* isolates.
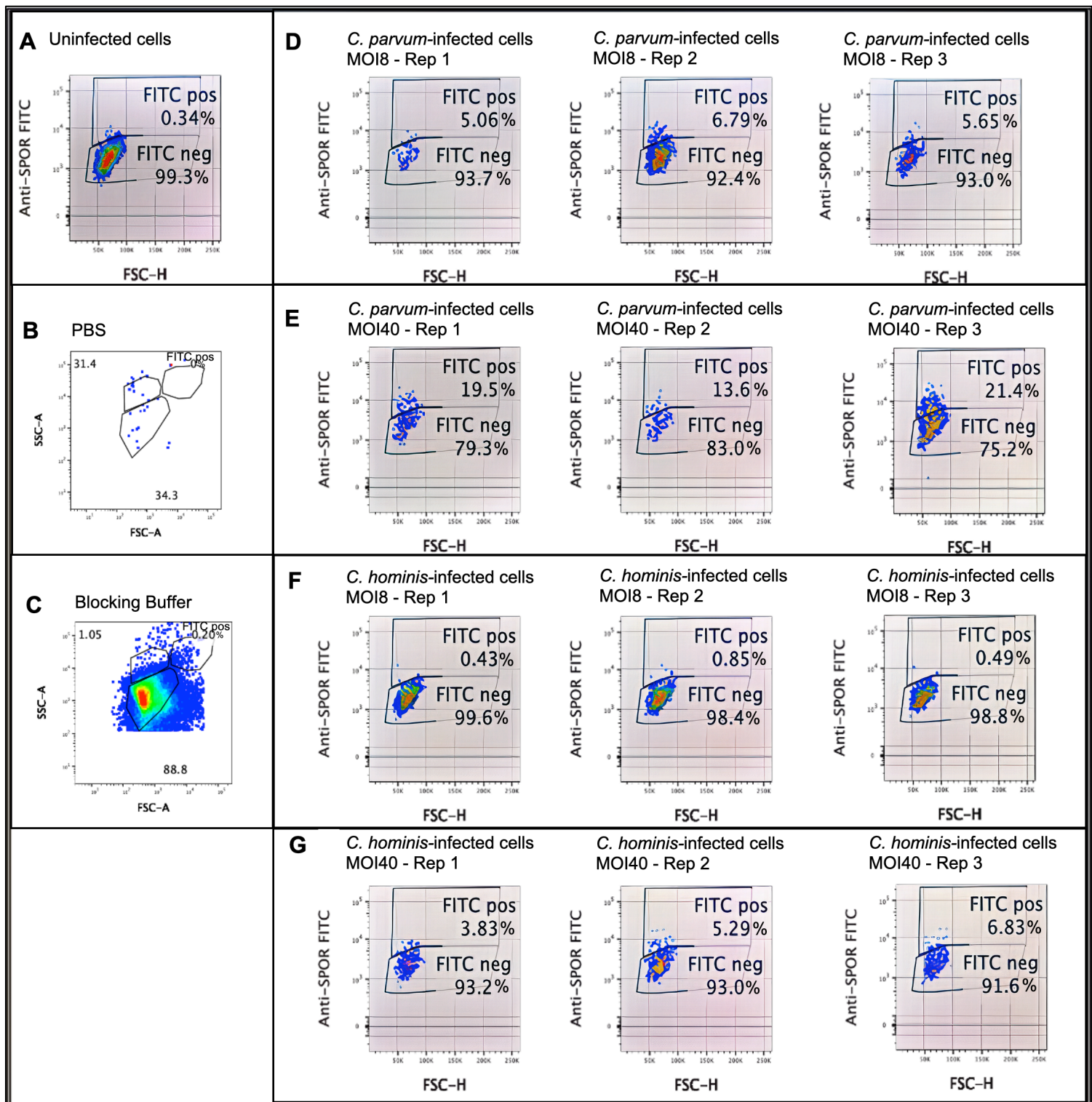
Figure 5.5. *C. parvum*- and *C. hominis*-infected cultures both contain a sub-population of cells with increased FITC fluorescent signal. This sub-population becomes more abundant at higher rates of infection and is more prevalent with *C. parvum* than *C. hominis*. 154-day old *C. parvum* IIdA24G1 and 97 day old *C. hominis* IgA17 excysted oocysts were used for infection in this experiment. Uninfected cells and buffer only controls are shown (A-C). At MOI 8 *C. parvum*-infected cells (B) show a low but noticeable

population of infected cells (5 – 7%) expressing the FITC signal across all replicates, and at MOI 40 (C) this population increased (13 – 21%). At MOI 8 (D) there is negligible evidence of *C. hominis*-infected cells expressing the FITC signal, however, at MOI 40 (E) the population of cells expressing this signal is increased (3-7%).

Table 5.3. Capturing variability in expression of anti-Spor FITC between replicates. This table shows the percentage of cells the expressing the anti-Spor FITC antibody in each replicate and at each MOI the experiment was tested.

	C. hominis			*C. parvum*		
	MOI 8	MOI 40	MOI 50	MOI 8	MOI 40	MOI 50
Rep 1	0.43%	3.83%	5.93%	5.06%	19.50%	6.67%
Rep 2	0.85%	5.29%	7.62%	6.79%	13.60%	8.18%
Rep 3	0.49%	6.83%		5.65%	21.40%	
Rep 4		3.83%			7.09%	
Rep 5		4.61%			6.16%	
Average	**0.59%**	**4.88%**	**6.78%**	**5.83%**	**13.55%**	**7.43%**
Standard Deviation	**0.23%**	**1.25%**	**1.20%**	**0.88%**	**6.95%**	**1.07%**

5.4.5 Quantification of infectivity using novel signal

An in-depth analysis of the spectral pattern in infected cells suggested the presence of a noticeable signal in the near infra-red wavelength after blue laser (488 nm) excitation, detectable in the PeCy7 (and PerCP) channel of the flow cytometer. To investigate this, all samples and controls were re-analysed in both channels (N.B. these channels do not pick up on the anti-Spor FITC antibody used in this study). A dose-responsive increase in this novel signal (hereafter referred to as SigM) in both *C. hominis* and *C. parvum*-infected cells was observed. It was noticeably more pronounced than the signal produced by the anti-Spor FITC antibody (Figure 5.6.). In the samples infected with *C. hominis* at an MOI of 8 approximately 2.8 – 3.8% of cells were positive for SigM (Figure 5.6.F), at an MOI of 40 approximately 14 – 18% were positive for SigM (Figure 5.6.G). In the *C. parvum*-infected cells, at an MOI of 8, approximately 13 – 20% of cells were positive for SigM

(Figure 5.6.D), and at an MOI of 40, approximately 34 – 37% of cells were positive for SigM (Figure 5.6.E). The signal was not observed in any of the controls used (Figure 5.6.A-C), or *Salmonella typhimurium*-infected cells (Figure D.27.).
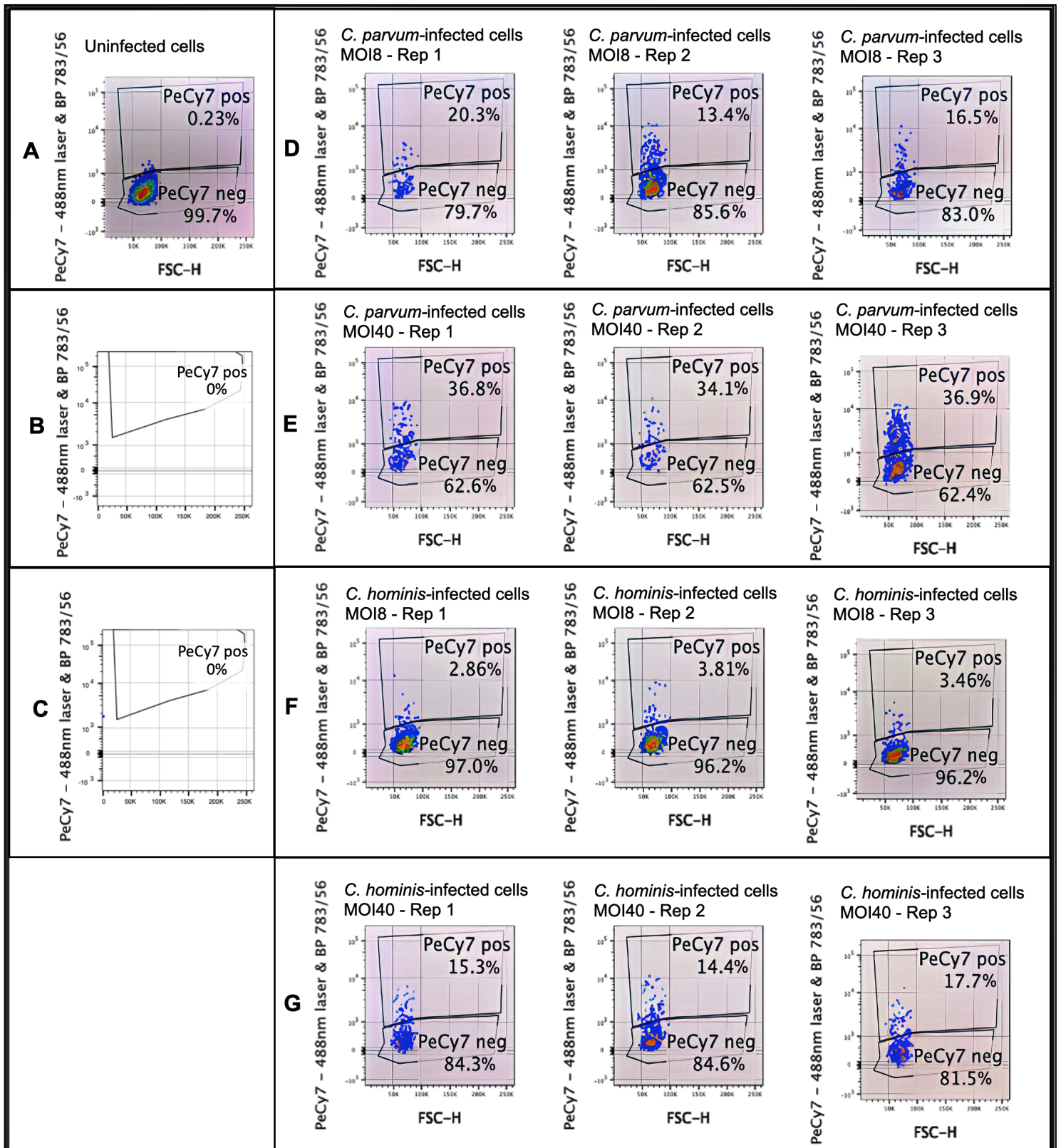
Figure 5.6. *C. parvum*- and *C. hominis*-infected cultures both contain a sub-population of cells with increased fluorescent signal in PeCy7 channel. 154-day old *C. parvum* IIdA24G1 and 97 day old *C. hominis* IgA17 excysted oocysts were used for infection in this experiment. Using the signal collected in the detector for light between 750nm and

811nm excited by the 488nm laser, a sub- population of cells is detectable that change in frequency corresponding to infectivity rates of *Cryptosporidium* and are more abundant in *C. parvum*-infected cultures.

The experiment was repeated to verify that the signal could be recreated. Figure 5.7. shows a comparison of the signal from the FITC antibody and the novel sigM in *C. hominis* – infected cells at an MOI of 40. Here again the FITC antibody is expressed by a low proportion of cells, approximately 3 - 5% of cells in both replicates. However, sigM is expressed by a higher proportion of cells (9 – 11% of cells in both replicates). The rest of the results from this iteration are shown in figures D.7. – D.17. Furthermore, when the experiment was repeated using a spectral cytometer (Aurora flow cytometer, Cytek, United States), which is capable of detecting and analysing multiple wavelengths of light at the same time, a population of infected cells expressing both the FITC antibody and sigM can be observed (Figure 5.8.). Due to the higher resolution of this machine a higher percentage of infected cells are positive for sigM alone (approximately 16%). The rest of the results from the experiment utilising the spectral cytometer are shown in figures D.18. – D.27.
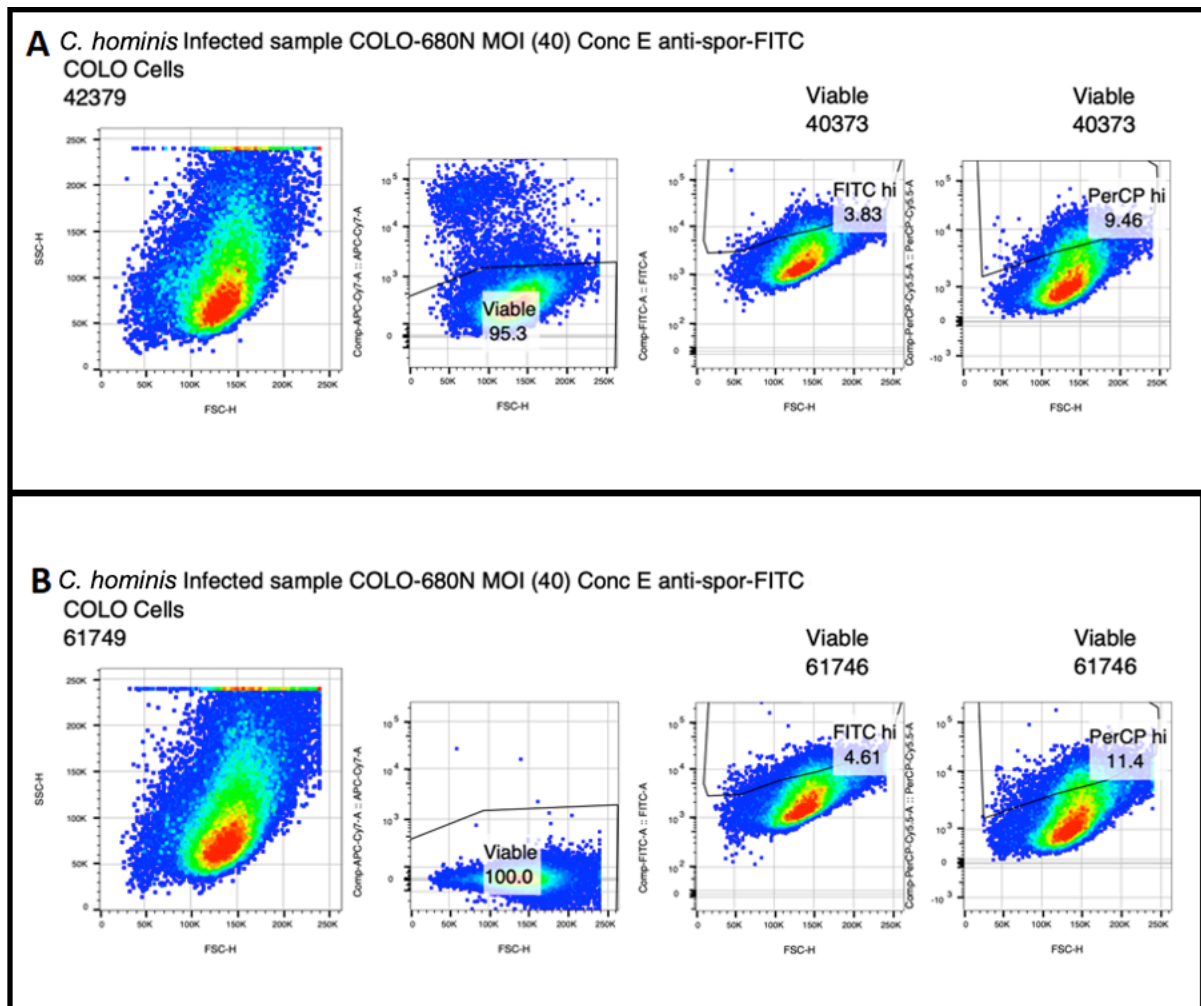
Figure 5.7. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in infected cells. Cells were infected with *C. hominis* at an MOI of 40. 160-day old *C. parvum* IIaA18G3R1 and 143-day old *C. hominis* IbA10G2 excysted oocysts were used for infection in this experiment. In each row the first panel displays the forward and side scatter characteristics of the sample, the second panel shows the percentage viable cells in that sample, the third panel shows the percentage of the viable cells expressing the FITC antibody, and the fourth panel shows the percentage of viable cells expressing sigM.
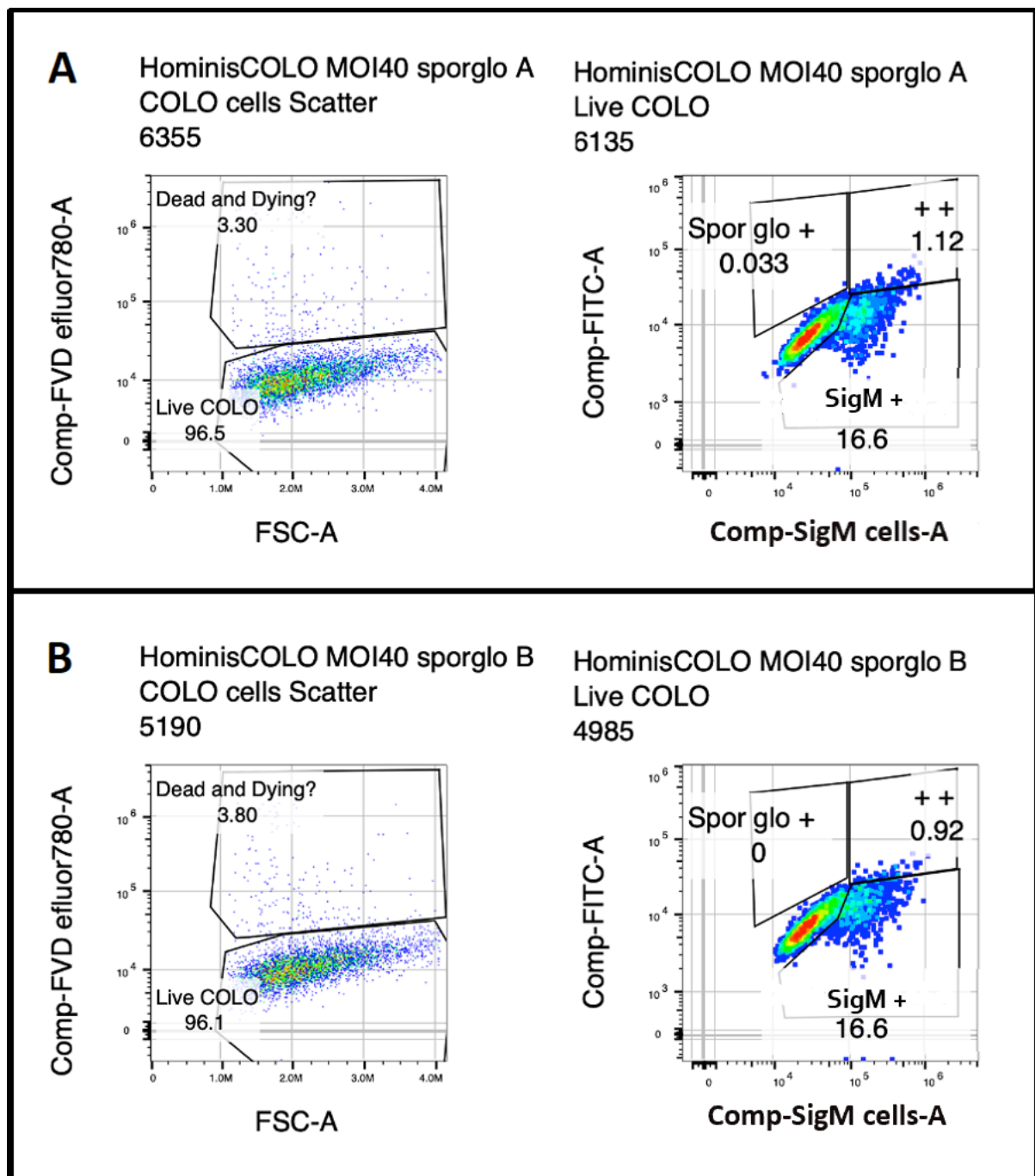
Figure 5.8. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in stained infected cells using the spectral cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. hominis* at an MOI of 40. 205-day old *C. hominis* IgA17 excysted oocsts were used for infection in this experiment. Viability displayed on the left panels with the percentage of live and dead cells shown. Fluorophore expression on the right panels with the percentage of cells expressing each signal shown.

5.4.6 Analysis of spectral signature from sigM-positive cells

To discount the possibility that the novel signal discovered in this study wasn't an artifact of autofluorescence, the spectral signatures of uninfected and infected cells were analysed. The spectral cytometer proved to be a particularly useful tool in this endeavour. Cells expressing sigM from infected cultures are substantially brighter in many channels compared to cells from uninfected cultures. The experiment outlined below used 193-day old *C. hominis* IgA20 and 184 day old *C. parvum* IIaA18G3R1 excysted oocysts for infection. Figure 5.10. shows sigM spectral signature from cells in *C. hominis*-infected cultures is many times brighter in absolute terms than spectral signature from uninfected cells.
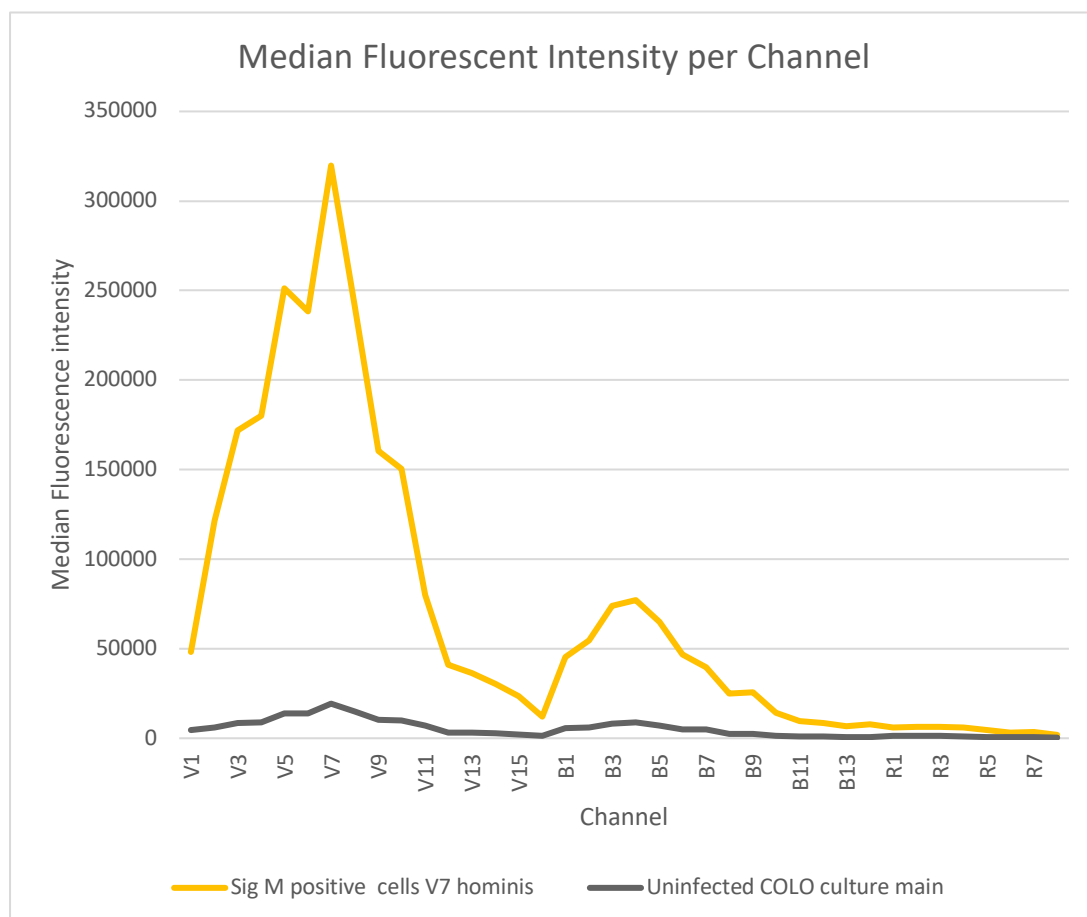
Figure 5.9. Comparison of spectral signature from uninfected cells and infected cells expressing sigM. Spectral signature of uninfected cells (grey) is far less bright than spectral signature of Sig M cells from *C. hominis* infected cultures (gold).

When brightness is very different, a comparison of spectral signature is less straight forward. Normalising the spectral signature allows a comparison of uniqueness between spectral signatures. Figure 5.11. shows a comparison of normalised spectral signatures between uninfected cells and *C. hominis*-infected cells expressing sigM. It shows that even after normalisation the spectral signatures are very different.
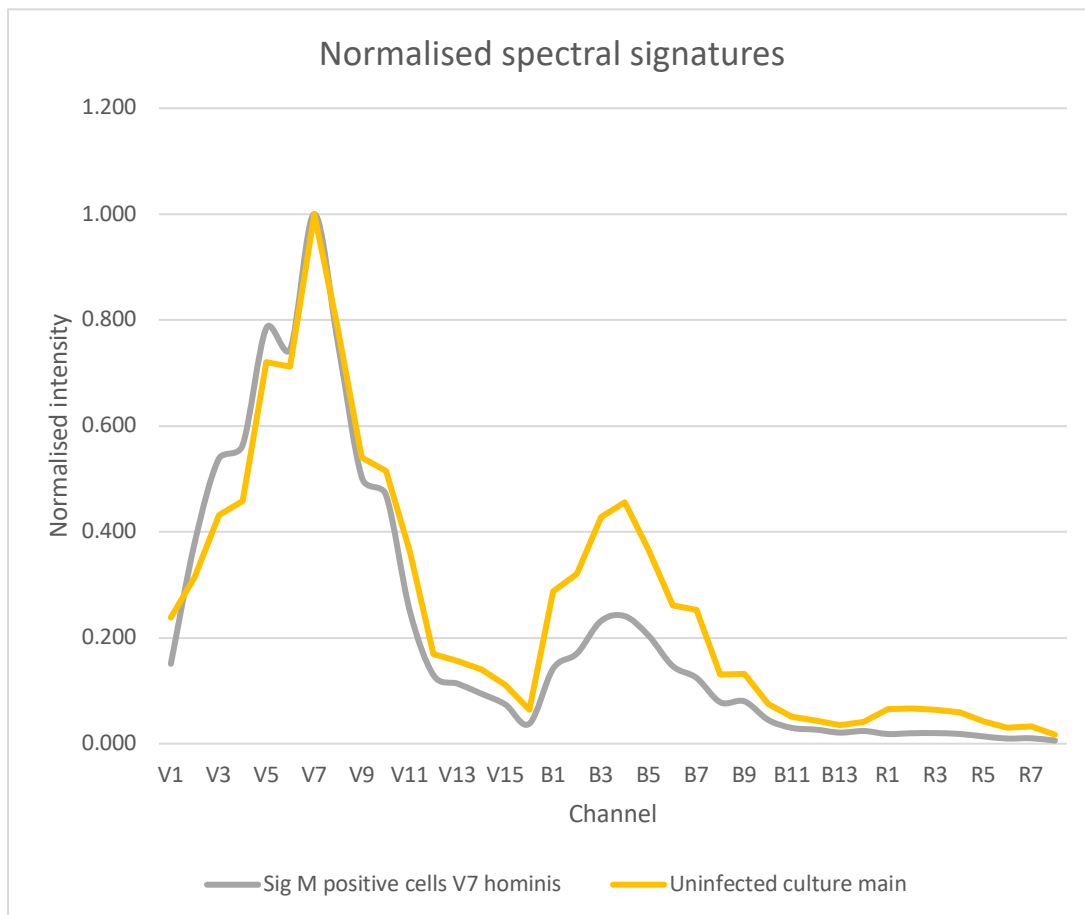
Figure 5.10. Comparison of normalised spectral signatures from uninfected cells and infected cells expressing sigM. Spectral signature of uninfected cells (gold) has a different normalized spectral signature to Sig M cells from *C. hominis* infected cultures (grey).

The spectral signatures from *C. hominis*- and *C. parvum*-infected cells expressing sigM were similar (Figure 5.12.) showing that this phenomenon was not unique to *C. hominis*-infected cells.
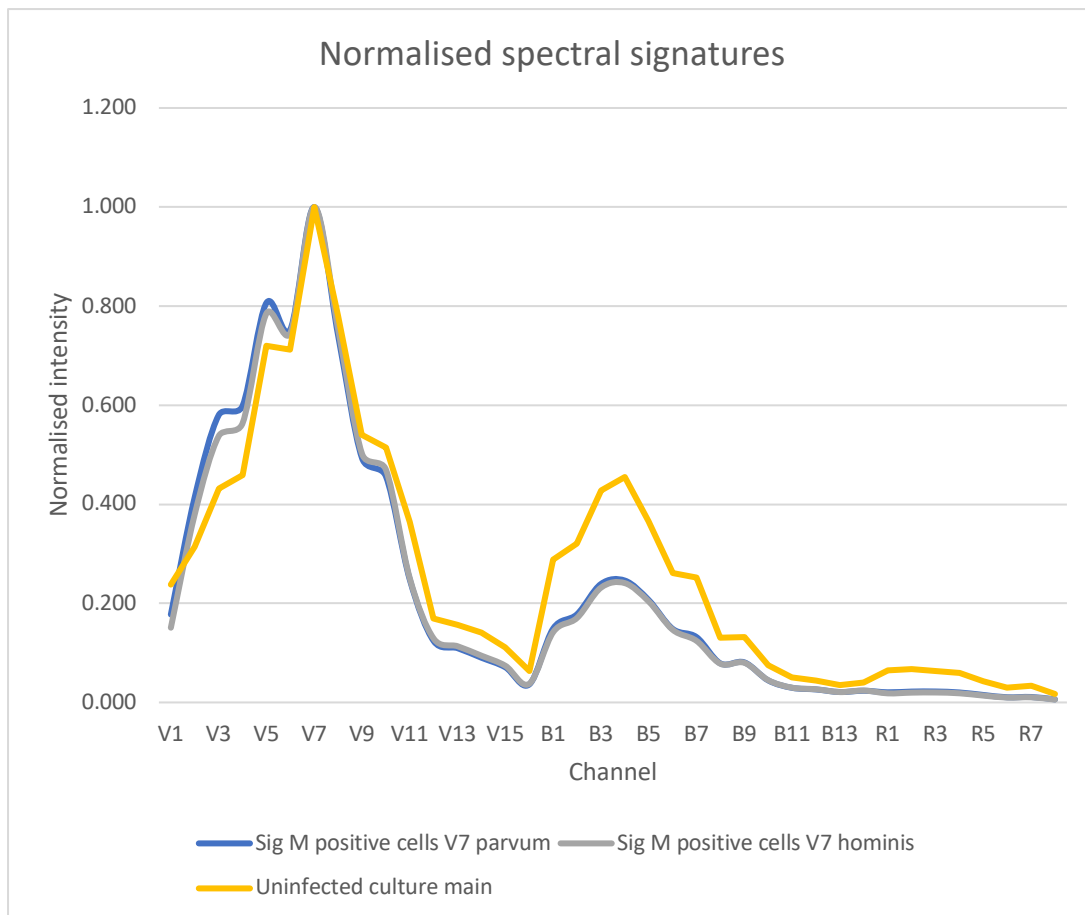
Figure 5.11. Analaysis of spectral signatures from *C. parvum*- and *C. hominis*-infected cells. Spectral signature for sigM from *parvum* infected and *hominis* infected COLO cells is strikingly similar (blue and grey). Uninfected COLO cells shown for comparison (gold).

Flow cytometers can analyse the spectral signatures of each cell in a sample in turn, this is how they are able to distinguish between populations with different characteristics within the same sample. *C. parvum*- and *C. hominis*-infected cultures contain populations of cells not expressing sigM. An analysis of the spectral signatures of these cells showed that they were identical to the spectral signatures of cells from the uninfected cultures (Figure 5.13.).
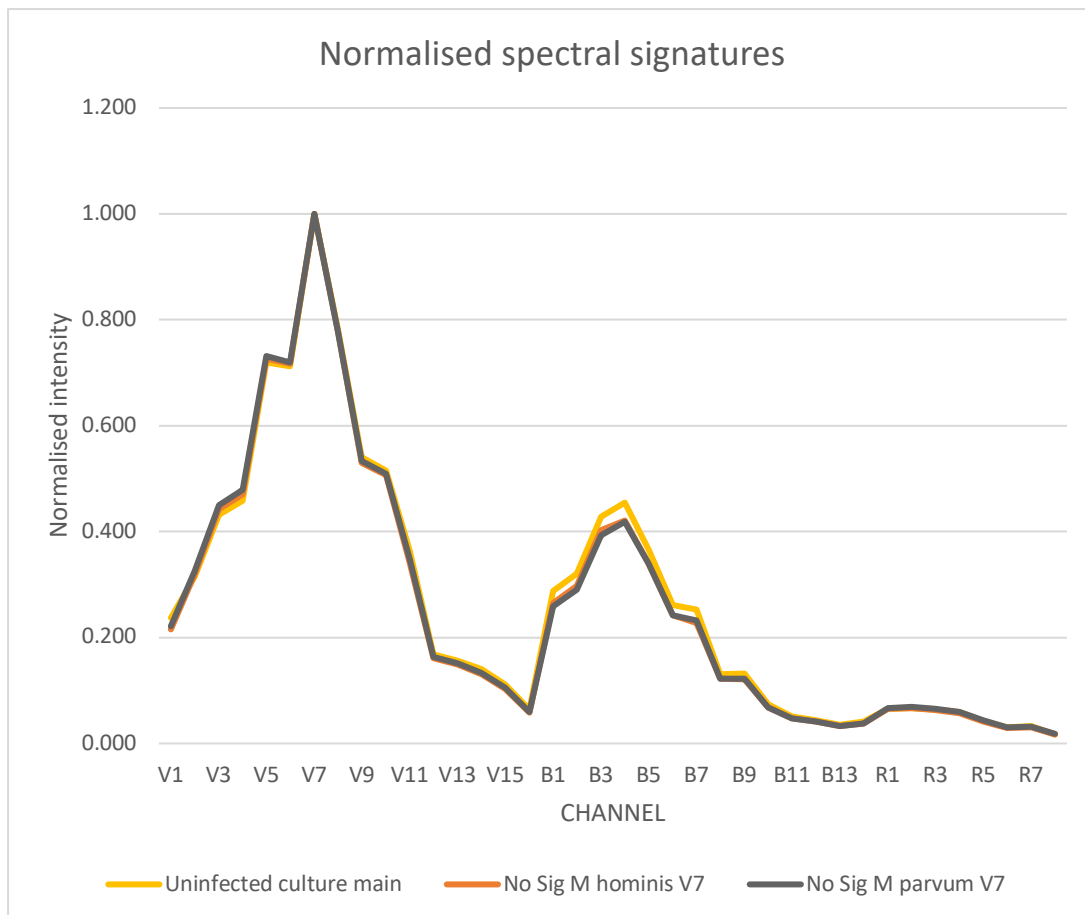
Figure 5.12. Analysis of cell from infected cultures not expressing sigM. *C. parvum-* and *C. hominis-*infected cultures contain cells without Sig M (mauve(obscured) and grey). The spectral signature of these cells is close to identical to that from uninfected cultures (gold)

5.5 Discussion

Most cell lines capable of culturing *Cryptosporidium* only allow limited progression through its life cycle and die after about 3 – 4 days post-inoculation (Karanis, 2018b). Some studies have reported the complete development of the life cycle of *Cryptosporidium* in the widely used HCT-8 cell line, however these results are difficult to replicate and require the use of a specialised maintenance medium for effective culture of the parasite (N. S. Hijjawi et al., 2001). This has hampered our understanding of the pathogenic mechanisms of the parasite and the development of effective therapies. Recently, a prominent limitation of *in vitro* cell culture of *Cryptosporidium* was alleviated through the characterisation of the COLO-680N cell line, which allows for the long-term

cultivation of the parasite without the need for specialised media, or costly, complicated mechanical structures such as hollow fibre culture systems (Bones et al., 2019). Due to these features, this cell line was utilised in this study for the cultivation of both *C. hominis* and *C. parvum*. Compared to HCT-8, COLO-680N maintains at least two-fold more viable cells at 48 hrs post-inoculation (Table 5.1. & 5.2.), suggesting it is more suitable for culture of *Cryptosporidium*.

Utilising flow cytometry in conjunction with the readily available anti-Spor FITC antibody it was possible to detect low levels of infection in cells at 48 hrs post-inoculation with *Cryptosporidium*. The results in the *C. hominis*-infected cells were not significant at low MOIs. This is possibly because no or few cells were infected, nonetheless, it was possible to detect a perceptible proportion of infected cells expressing the fluorophore at an MOI of 40. The anti-Spor FITC antibody was generated against *C. parvum* and has been validated against the parasite (Boxell et al., 2008). It has also been used for the detection of other *Cryptosporidium*, including *C. hominis* (N. Hijjawi et al., 2010; Karanis, 2018a). However, there are no binding efficacy studies, and the detection a flow cytometer signal, even at low values, suggested there might be *C. hominis* infection not detected by the anti-Spor FITC, and it is possible that the anti-*Cryptosporidium* polyclonal antibodies generated against *C. parvum* bind with lower efficacy to *C. hominis* antigens. Future studies could investigate the binding efficacy of multiple concentrations of the fluorophore on fixed quantities of oocysts from multiple species of *Cryptosporidium*, to verify if it shows a bias towards one species over another.

A clearer distinction was achieved in *C. parvum*-infected cells wherein up to 7% of cells at an MOI of 6 and 21% of cells at an MOI of 40 expressed the fluorescent antibody. These results suggest that while it is not possible to compare infectivity between species using the anti-Spor FITC antibody, it may be possible to use the antibody in conjunction with the high resolution of flow cytometry to assess the infectivity of multiple genotypes within species when a high infective dose of sporozoites is used. Previous studies assessed the infectivity of *Cryptosporidium* using *in vitro* techniques but due to the limitations of working with the HCT-8 cell line and cell-free medium they assessed only the factors affecting the establishment of infection, focusing on extracellular life cycle stages (King, Hoefel, Lim, Robinson, & Monis, 2009; King, Keegan, Robinson, & Monis,

2011). A possible reason for the low expression of the anti-Spor FITC antibody in infected cells could be due to the low concentration of the antibody used in this study. The fluorophore was diluted 1:16 because due to the high sensitivity of the flow cytometer some background staining was detected in uninfected cells. For this reason, a low concentration of the fluorophore was used to allow accurate distinction of infected and uninfected populations, and to reduce the chance of non-specific binding thereby reducing the chance of false positive results. An improvement of the staining procedure, possibly by repeating the anti-Spor FITC staining step multiple times for more effective binding, could alleviate this limitation. Another possible reason for the low FITC signal observed in this study could be that the infection of the cells was unsuccessful. For this reason, *in situ* staining of infected cells and microscopic imaging would be required to verify the presence of the intracellular life cycle stages of *Cryptosporidium* in the cells.

The results from the novel signal (SigM) detected in the near infra-red wavelength after blue laser using the flow cytometer highlight the potential of the *in vitro* system outlined in this study. SigM was expressed by a noticeably higher population of both *C. parvum* and *C. hominis*-cells when compared to the anti-Spor FITC antibody and was not expressed by any of the negative or positive infection controls. This suggests that the detected signal might be specific to COLO-680N cells infected with *Cryptosporidium*. In addition, when the *C. parvum*-infected cells are taken into account at the same MOI, the percentage of cells expressing sigM between the two species is quite similar. This suggests that sigM is not preferentially expressed by one species over the other, thereby suggesting that it is a useful marker for *Cryptosporidium* infection in the COLO-680N cell line. Validation of this signal was limited by the fact that a substantial amount of the sigM-positive cells could not be recovered through fluorescence activated cell sorting to conduct *in situ* staining of infectious foci using the anti-Spor FITC antibody. This limitation could be solved by upscaling the experiment from the 12-well plates to a T25 cell culture flask, which would allow for the recovery of more cells for analysis and the subsequent microscopic identification of *Cryptosporidium* life cycle stages in the cells expressing sigM. Due to the low expression of the anti-Spor FITC antibody observed in this study, such validation, possibly with the inclusion of qPCR quantitation, would be necessary to verify the successful infection and long-term cultivation of the parasite in this platform. However, an analysis of the spectral signature of infected cells expressing

sigM suggest that it is distinct from that of uninfected cells both in brightness and signal characteristics. Other fluorophores are available for the detection of *Cryptosporidium*, however, most of these are specific for the detection of the extracellular life cycle stages of the parasite (Barugahare et al., 2011; Warnecke et al., 2003). The anti-Spor FITC antibody used in this study is still the most widely used off-the-shelf fluorophore for the detection of the intracellular life cycle stages of the parasite in cell culture. Further work should involve a within-species comparison using multiple genotypes to find out if there are any differences in infectivity between genotypes of *C. parvum* and *C. hominis.*

The oocysts used in this study were anonymously sourced directly from human patients. While it is possible to order oocysts from a supplier, this was done to maintain the level of genetic diversity of *Cryptosporidium* present in any human infection. A limitation of this strategy was the difficulty of excysting some isolated oocysts. Older oocysts showed a reduced efficiency of excystation, and this had an effect on the level of infection observed. As shown in Chapter 4, infections in humans rarely display only one subtype of the parasite, so the benefit of replicating infections in humans using this *in vitro* platform outweigh the limitations of the temporal deterioration of oocyst viability. This limitation could be overcome through collaboration with diagnostic labs. An understanding of the relative infectivity of different species and genotypes of *Cryptosporidium* would be beneficial to public health authorities when trying to attribute the source of an outbreak.

In conclusion, this study presents the foundation of an *in vitro* assay potentially capable of assessing the infectivity of *C. parvum* and *C. hominis* in the COLO-680N cell line and suggests that it is possible to do so without the use of a fluorescent antibody. Following further validation and refinement this system has the potential to serve as a platform for the testing of new molecules and drugs for the treatment of cryptosporidiosis and provide new insights into the disease mechanisms of this pathogen.

6 Comparative gene expression analysis of *Cryptosporidium* species

6.1 Abstract

Understanding of gene expression of *Cryptosporidium* during infection could advance development of therapeutics to combat the widespread enteric disease, cryptosporidiosis. In recent years, great advances have been made using *in vitro* techniques to characterise the gene expression of the parasite during various stages of its life cycle. However, these investigations have been conducted using cell lines that do not reliably reproduce the full life cycle of the parasite. This study uses the COLO-680N cell line, which allows for the complete progression of the life cycle of *Cryptosporidium*, and NanoString nCounter analysis technology to assess the expression of genes at various timepoints during infection. In addition, the expression of twelve potential drug targets was assessed, with the aim of contributing to the development of new therapeutics for the treatment of cryptosporidiosis.

6.2 Introduction

Cryptosporidiosis is a disease affecting humans and animals across all economic settings worldwide. It is caused by the parasite *Cryptosporidium*, which causes self-limiting diarrhoea in infected individuals, usually lasting approximately 1-2 weeks post-infection (Su, Jin, Wu, et al., 2019). *Cryptosporidium* is one of the most common causes of diarrheal disease in the world, and while it is self-limiting in healthy individuals it is potentially fatal in the immunocompromised and the young (<5 years old). It is estimated that globally approximately 83,000 children under 5 years old died due to cryptosporidiosis in the year 2011 (Lanata et al., 2013). The real figure is probably higher due to the inefficient disease reporting mechanisms in some countries. Due to the previously unappreciated public health significance of the disease, in 2004 the World Health Organisation (WHO) added cryptosporidiosis to its Neglected Diseases Initiative (Savioli et al., 2006a). As yet, there are no vaccines against cryptosporidiosis and apart from the

broad spectrum antiparasitic drug nitazoxanide there is a dearth of effective treatments for this disease.

Cryptosporidium is transmitted mainly through the faecal-oral route and ingestion of contaminated water is the most common mode by which this parasite is spread in humans (Wuhib et al., 1994). The two species responsible for the majority of infections in humans are *C. hominis*, thought to be specific to humans, and *C. parvum,* which has a wider host range and is responsible for most zoonotic transmission of this parasite. Oocysts are the environmentally resistant infectious stage of the *Cryptosporidium*. They are able to withstand treatment with disinfectants like chlorine and bleach and are capable of surviving for up to 24 weeks at 20°C (Cacciò & Widmer, 2013). When oocysts are ingested, they travel through the gastrointestinal tract until they arrive at the distal ileum. Within each oocyst are 4 sporozoites (Figure 1.1.), and when the oocysts reach the distal ileum these sporozoites are released in a process termed excystation. The sporozoites invade a cell and form a parasitophorous vacuole within the plasma membrane in which they conduct the rest of their life cycle. Following invasion, the parasite progresses through the asexual phase of its life cycle, which involves the proliferation of merozoites by type I meronts. The development of type II meronts initiates the sexual phase of the life cycle, which eventually results in the production of two forms of oocysts: thin-walled oocysts that maintain the infection in the host and thick-walled oocysts that are released into the environment through the faeces of the host organism.

Cryptosporidiosis has no pathognomonic symptoms, nevertheless, the most common symptom of the disease is severe diarrhoea. Other symptoms can occur as a result of the disease, ranging from nausea and vomiting to joint and eye pain (Cacciò & Widmer, 2013). Some studies have linked specific species of *Cryptosporidium* to specific symptoms (Hunter et al., 2004a), but no definitive associations have been made due to our lack of understanding of the disease mechanisms of the parasite and the difficulty of effectively manipulating the parasite *in vitro*. With regards to *in vitro* manipulation, the cell lines most frequently used for the culture of *Cryptosporidium* are HCT-8 and Caco-2. While some success has been had in the complete and long-term cultivation of the parasite in these cell lines (N. S. Hijjawi et al., 2001; Tandel et al., 2019; Winkworth et al., 2008),

other studies have had limited success recreating these results and have only successfully achieved partial progression of the life cycle of this parasite, terminating in the asexual phase (Bones et al., 2019). This is the basis on which much of our knowledge of the processes and properties of the parasite is founded. Other techniques for the culture of the parasite exist, such as the use of 3D culture systems and animal models, but these require complex and expensive setups that are not conducive to large scale drug development (Karanis, 2018b).

In an effort to ascertain if any differences in gene expression between species could account for the differences in symptoms and severity of disease reported by previous studies, and to assess the expression of potential drug targets (Baragaña et al., 2019; Castellanos-Gonzalez et al., 2019; Manjunatha et al., 2017; Mfeka et al., 2020; Su, Jin, Wu, et al., 2019; Xu et al., 2019; Zhang et al., 2019), a panel of genes from previous transcriptomic studies (Lippuner et al., 2018a; Matos et al., 2019a) were selected and their expression in the COLO-680N cell line was tested. COLO-680N is a recently characterised cell line that allows for the progression of the entire life cycle of *Cryptosporidium* and allows for the long-term culture of the parasite without the requirement of maintenance medium. Previous studies have shown differences in gene expression depending on the organism or cell line used to culture this parasite. Taking into account the aforementioned benefits of COLO-680N over other cell lines, this study sought to assess the expression of *Cryptosporidium* and host genes at multiple timepoints during infection and compare it to data from previous studies, to gain a better understanding of the expression genes at various stages of the life cycle of this pathogen.

6.3 Methods

6.3.1 <u>Samples</u>

C. parvum and *C. hominis* were purified from anonymised human faecal samples that are routinely collected by the Hopkirk Research Institute under a contract with the New Zealand Ministry of Health. Identification of these samples was done using nested PCR and Sanger sequencing of a fragment of the glycoprotein 60 (*gp60*) gene using published methods (Garcia-R et al., 2020; Garcia–R et al., 2017b). At the time of the experiment only the most recent samples were included to prevent a drop off in the efficiency of excystation (see Chapter 5 discussion).

6.3.2 <u>Infection of cell monolayers</u>

COLO-680N (DSMZ, ACC182) cells were infected with *C. parvum* and *C. hominis* according to published protocols (Jossé et al., 2019). In 12-well culture plates, COLO-680N cells were grown to approximately 70% confluence in growth medium consisting of RPMI 1640 supplemented with 10% foetal bovine serum (FBS), 100 U/ml of penicillin, 100 µg/ml of streptomycin and 250 ng/ml of amphotericin B. Oocysts were excysted using a modification of the methods outlined by Rasmussen et al., (1993). The modification involved the incubation of oocysts in an excystation solution composed of 0.8% w/v taurocholic acid in PBS for 2 h. The excysted oocysts from each species were used to infect the cells at a multiplicity of infection (MOI) of 80, meaning approximately 80 sporozoites to each cell (calculated by the approximate number of cells in each well at 70% confluence). The plates were incubated at 37°C in a humidified incubator with 5% CO_2 post-inoculation (p.i.) and left for 48 h. There were 3 samples from each species and the experiment was repeated once under these conditions and once again with multiple lengths of incubation p.i. – 24h, 48h, 96h and 120 h As a positive-infection control *Salmonella typhimurium* was cultured according to established protocols (Abernathy et al., 2013) and used to infect cells that had been suspended in antibiotic-free medium at a concentration of 2.5 x 10^7 CFU/well. These were incubated for a maximum of 24 h p.i.

because the rapid proliferation of *S. typhimurium*. Untreated COLO-680N cells were used as negative controls.

6.3.3 RNA isolation

Following incubation, excess sporozoites and oocysts were removed from each well by washing in 500 μl of PBS. The cells were harvested at all time points specified above using 0.25% trypsin-EDTA and resuspended in 500 μl of PBS prior to RNA extraction. A RNeasy mini kit (Qiagen, Hilden, Germany) was used for RNA extraction and the protocol was executed according to the manufacturer's instructions. The quantity of RNA in each sample was measured by a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, United States) then each sample was diluted where necessary, by using a CentriVap® (Labconco, Kansas City, Missouri, United States) complete vacuum concentrator to consolidate the RNA, then resuspending in RNAse free water to a maximum RNA concentration of 128.5 ng/μl. The samples were stored at -80°C prior to sample prep and NanoString™ (NanoString, Seattle, Washington, United States) analysis.

6.3.4 mRNA detection using nCounter

mRNA detection was conducted using the nCounter (NanoString) platform. The panel used consisted of 144 genes: 40 human genes frequently expressed in human cell lines; 104 *Cryptosporidium* genes extracted from analysis of previous RNAseq studies, with 48 thought to be expressed intracellularly and 48 extracellularly according to data gleaned from a study by Matos et al., (2019) (Table E.1. - E.3.) and cross-referenced with a study by Lippuner et al., (2018) (Table E.4.). The extra 8 *Cryptosporidium* genes were selected based on analysis of potential drug targets from recent studies (Baragaña et al., 2019; Castellanos-Gonzalez et al., 2019; Manjunatha et al., 2017; Mfeka et al., 2020; Su, Jin, Wu, et al., 2019; Xu et al., 2019; Zhang et al., 2019). The gene IDs from the study were cross-

referenced with data from the CryptoDB database (http://cryptodb.org), from which the corresponding mRNA sequences were extracted. A full list of gene IDs included in this study can be found in table E.1.

Gene expression analysis was performed using the nCounter Analysis System (NanoString Technologies Inc., Seattle, WA). The use of NanoString technology enables RNA expression analysis from either purified RNA or directly from cell lysates without further RNA purification (Malkov et al., 2009) or amplification. The method uses molecular barcodes on gene-sequence-specific probes and single-molecule imaging to count RNA copies (Geiss et al., 2008). Briefly, multiplexed probes were designed with 2 sequence-specific probes for each gene of interest. The capture probe was coupled to biotin as an affinity tag. The reporter probe was coupled to a colour-coded tag. Each target molecule of interest is identified by the unique colour code generated by the ordered fluorescent tags on the reporter probe. The level of expression was measured by counting the number of codes for each mRNA using digital imaging. This allows the analysis of multiple genes from the same sample (multiplexing) using a customized set of probes with distinct bar codes, called a ProbeSet. RNA was hybridised with the ProbeSets according to the manufacturer's instructions (nCounter Elements XT Reagents User Manual; NanoString MAN-10086-01 June 2018). Briefly, RNA samples were thawed on ice. Samples were hybridised by adding 8 µL of MasterMix and 7 µL of RNA per each tube of a 12-tube strip immediately before placing the strip at 67°C for 22 h. After hybridization, samples were transferred to the nCounter Prep Station which automatically removed excess probe and aligned and immobilized the probe-target complexes in the nCounter cartridge. Sample cartridges were placed in the nCounter Digital Analyzer which counted and tabulated colour codes on the surface of the cartridge for each target molecule. Data were retrieved from the Analyzer as raw data (Reporter Code Count, RCC) files.

6.3.5 Data analysis

The raw reporter code counts were retrieved from the analyser in a tabulated data file (RCC) and imported into the nSolver Analysis software, version 4.0 (https://www.nanostring.com/products/analysis-solutions/ncounter-analysis-

solutions/), for analysis. A Reporter Library File (RLF) specific to our CodeSet was delivered to us by the manufacturer when the RNA panel was ordered, it contained fundamental information such as the assignment of probe to gene. This file was used by the nSolver software to execute its quality control (QC) program on the samples using these parameters: fields of view registration <75% (imaging QC); binding density outside of 0.1 – 2.25 range (binding density QC); positive control R^2 value <0.95 (positive control linearity QC); and 0.5fM positive control ≤2 SD above the mean of negative controls (positive control limit of detection QC). All samples passed the QC, however, there were limit of detection QC flags present in multiple samples due to a low level of detection of *Cryptosporidium*-specific genes. Positive controls (spiked by the NanoString Company in the Code-set) were used for correcting assay efficiency. Negative controls were used to filter out microRNAs with expression at noise level. Median normalization was performed to normalize across samples using all housekeeping genes and heat-maps were used for data visualisation according to established protocols (Yu et al., 2019).

6.4 Results

6.4.1 Comparative expression of mRNA in *Cryptosporidium*-infected samples collected 48 h post-inoculation

Initially, this study set out to analyse the relative expression of genes from our panel in samples infected with either *C. hominis* or *C. parvum* compared to the respective sporozoites alone; untreated cells and *S. typhimurium*-infected cells were used as controls. The NanoString nCounter analysis system hybridizes two probes (capture and reporter) with unique barcodes directly onto the RNA target without amplification, cDNA or library preparation. This allows for the direct counting of each target molecule using an automated fluorescence microscope, with no need for reads per kilobase of transcript per million mapped reads (RPKM), fragments per kilobase of transcript per million mapped reads (FPKM), or transcript per kilobase million (TKM) values (Eastel et al., 2019; Malkov et al., 2009; Yu et al., 2019). There was no post-processing normalisation or fold difference analysis due to the internal corrections used by the NanoString

nCounter system. NanoString, unlike RNAseq does not necessitate the use of fold differences due to its ability to directly measure a broad range of mRNA expression levels without cDNA synthesis and amplification steps (Urrutia et al., 2016). Analysis of the mRNA levels from cells collected 48 h post-inoculation with *Cryptosporidium* is shown in the heatmap in Figure 6.1. Compared to the host genes, the parasite genes were expressed at very low levels in infected cells. Also, the mRNA counts of parasite genes in the infected cells and the controls (uninfected cells and *S. typhimurium*-infected cells) are quite similar so no comparisons could be made. The raw counts of mRNA (Table E.5. & E.6.) provide a clearer picture of this lack of variation. The sporozoites showed relatively high expression of all the parasite genes compared to the infected cells (Figure 6.1.). This is particularly notable because 48 of the parasite genes were selected based on their intracellular expression characteristics. The sporozoites also showed practically no significant expression of host genes, which verifies the specificity of the panel. From the results there appeared to be a higher expression of parasite genes in the *C. parvum* sporozoites compared to the *C. hominis* sporozoites. This was possible due to differences in efficiency of excystation: 5×10^6 oocysts of each species were excysted to produce those samples. Counting of the sporozoites post-excystation showed that the *C. parvum* sample produced 10.5×10^6 sporozoites, while the *C. hominis* sample produced only 1×10^6 sporozoites.
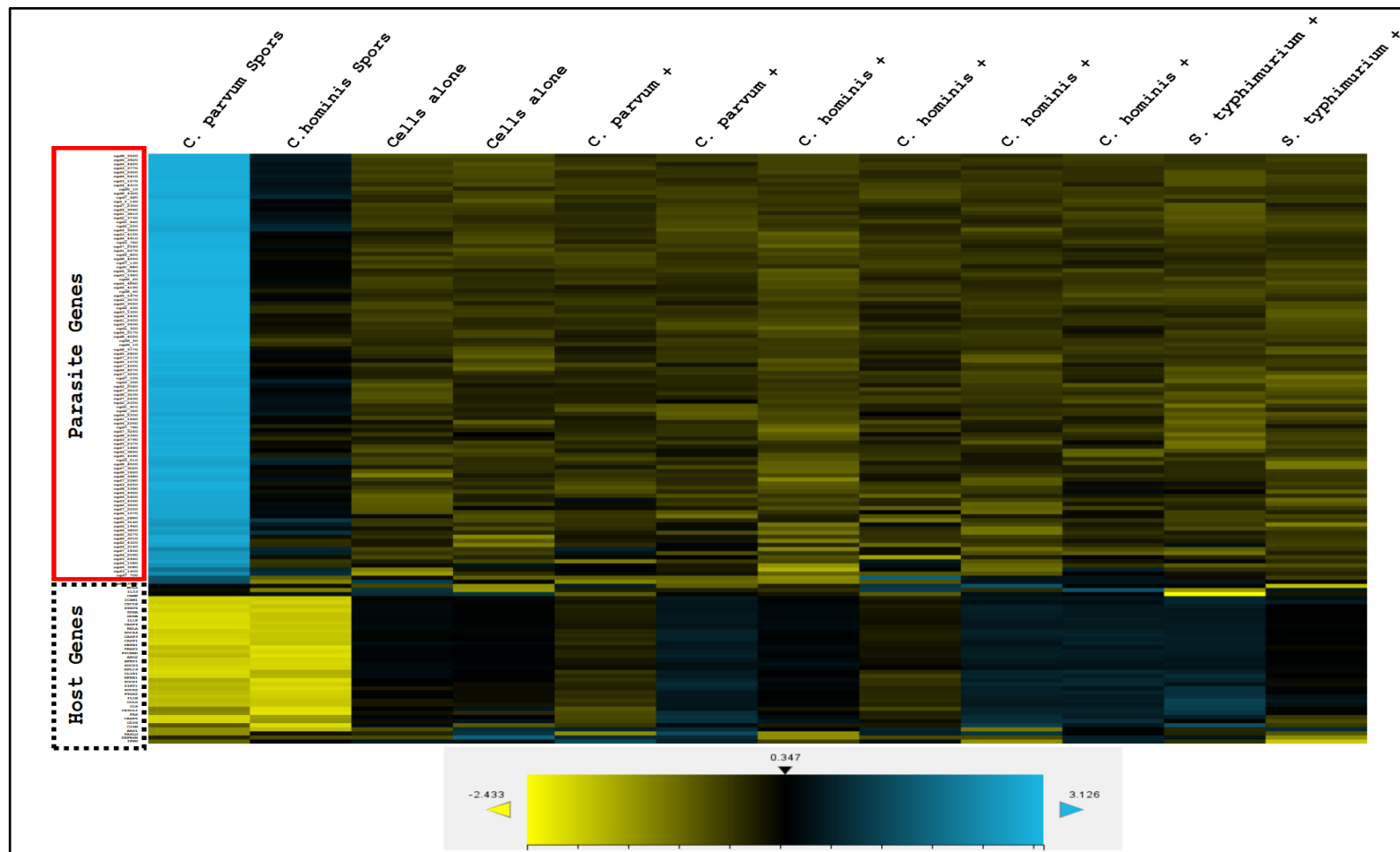
Figure 6.1. Heatmap showing the relative abundance of mRNA transcripts of parasite and host genes in each sample from the 48 h experiment. (+) signifies infected cells. mRNA transcript abundance is calculated using Euclidean distance - the distance between two samples or genes is calculated as the square root of the sum of squared differences in their log count values. Full list of genes is shown Table E.1. and raw counts are shown in Table E.5.
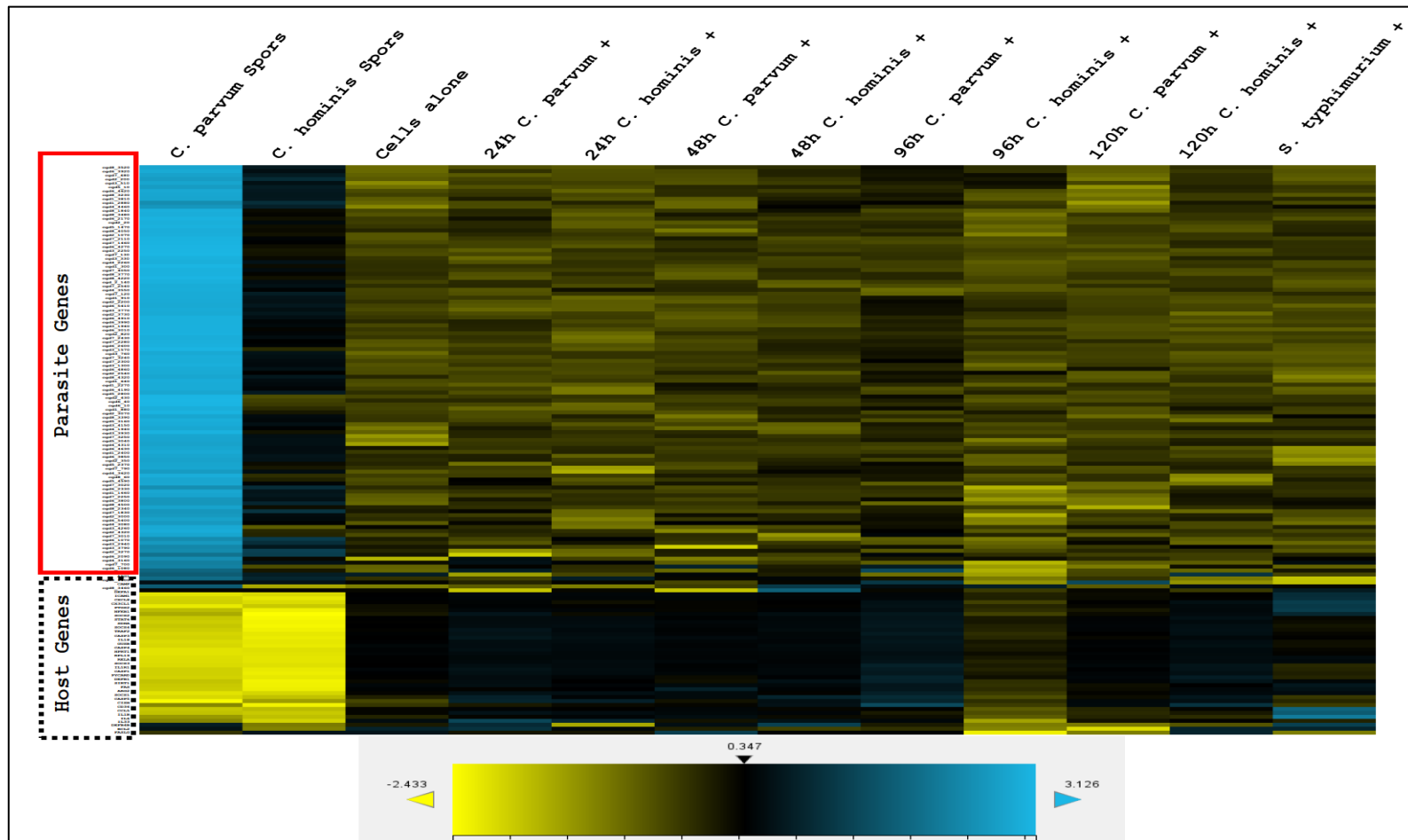
Figure 6.2. Heatmap showing the relative abundance of mRNA transcripts of parasite and host genes in each sample from the time series experiment. (+) signifies infected cells. mRNA transcript abundance is calculated using Euclidean distance - the distance between two samples or genes is calculated as the square root of the sum of squared differences in their log count values. Full list of genes shown in Table E.1. and raw counts are shown in Table E.6.

6.4.2 Comparative expression of mRNA from timeseries samples

A time series was conducted to ascertain if any differences in gene expression could be captured during an extended period of infection. Samples were collected at 24 h, 48 h, 96 h and 120 h post-inoculation. Similar to the previous experiment, no significant variation was observed in the expression of *Cryptosporidium*-specific genes at any time point in infected cells (Figure 6.2). Additionally, the raw counts for the parasite genes were low across all the cell samples (Table E.6.), so no comparison could be made. Once again, the sporozoite-only samples showed the highest levels of expression of parasite genes and the lowest expression of host genes.

6.4.3 Analysis of expression of potential drug targets and genes of interest

Table 6.1. shows a list of genes of interest that were included in the panel based on analysis of previous studies that identified them as potential drug targets. Heatmaps showing the relative expression of these genes in infected cells from the 48 h experiment and the time series are shown in Figures 6.3. and 6.4. All these genes were expressed at high levels in sporozoites compared to infected cells. Particularly, cgd2_3730, a glutathione S-transferase protein of interest, and cgd5_4590, a MEDLE-2 gene involved in parasite invasion, which previous studies showed is expressed at high levels in sporozoites was also among the genes showing the highest levels of expression in sporozoites in both the 48 h and time series experiments. Cgd3_1400, a pyrophosphate-fructose-6-phospate-1-phospho-transferase identified as a potential vaccine candidate by a previous study (Panda & Mahapatra, 2018a), appeared to be the gene of interest expressed at the lowest level in sporozoites (Figures 6.3. & 6.4.). Compared to the other genes of interest cgd3_1400 was expressed at a relatively higher level in infected cells, particularly the samples collected at 48 h and 96 h post-infection. Previous studies suggest this particular gene is mostly expressed in the plasma membrane and involved in the metabolic pathways of the parasite. No significant variation was observed in the expression of the other genes of interest for any inferences to be made.

Table 6.1. List of genes of interest and potential drug/vaccine targets based on review of current literature. Gene ID and Orthologue Groups extracted from the cryptoDB database (https://cryptodb.org/cryptodb/app/)

Gene ID	Gene Description	Orthologue Group	Reference Literature
cgd2_3730	Glutathione_S-transferase_C-terminal_domain_containing_protein	OG5_180098	(Mfeka et al., 2020)
cgd3_1400	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase	OG6_106222	(Panda & Mahapatra, 2018a)
cgd3_2940	Phospholipase D/Transphosphatidylase	OG6_106810	(Panda & Mahapatra, 2018a)
cgd4_1940	Nucleoside diphosphate kinase	OG6_100304	(Castellanos-Gonzalez et al., 2019)
cgd4_4460	Dihydrofolate reductase-thymidylate synthase	OG6_101427	(Gibbons et al., 1998)
cgd5_3160	Actin	OG5_126595	
cgd5_4590	MEDLE_gene_family_protein (MEDLE-2)	OG5_161402	(B. Li et al., 2017)
cgd7_1830	UTP--glucose-1-phosphate uridylyltransferase	OG6_109569	(Panda & Mahapatra, 2018a)
cgd7_3020	Rhomboid-like protein	OG6_100562	(Yang et al., 2016)
cgd7_480	L-lactate/malate_dehydrogenase	OG5_126911	(Dhal et al., 2018)
cgd7_700	N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase	OG6_101964	(Panda & Mahapatra, 2018b)
cgd8_4500	Phosphatidylinositol 3-/4-kinase	OG6_101543	(Manjunatha et al., 2017)
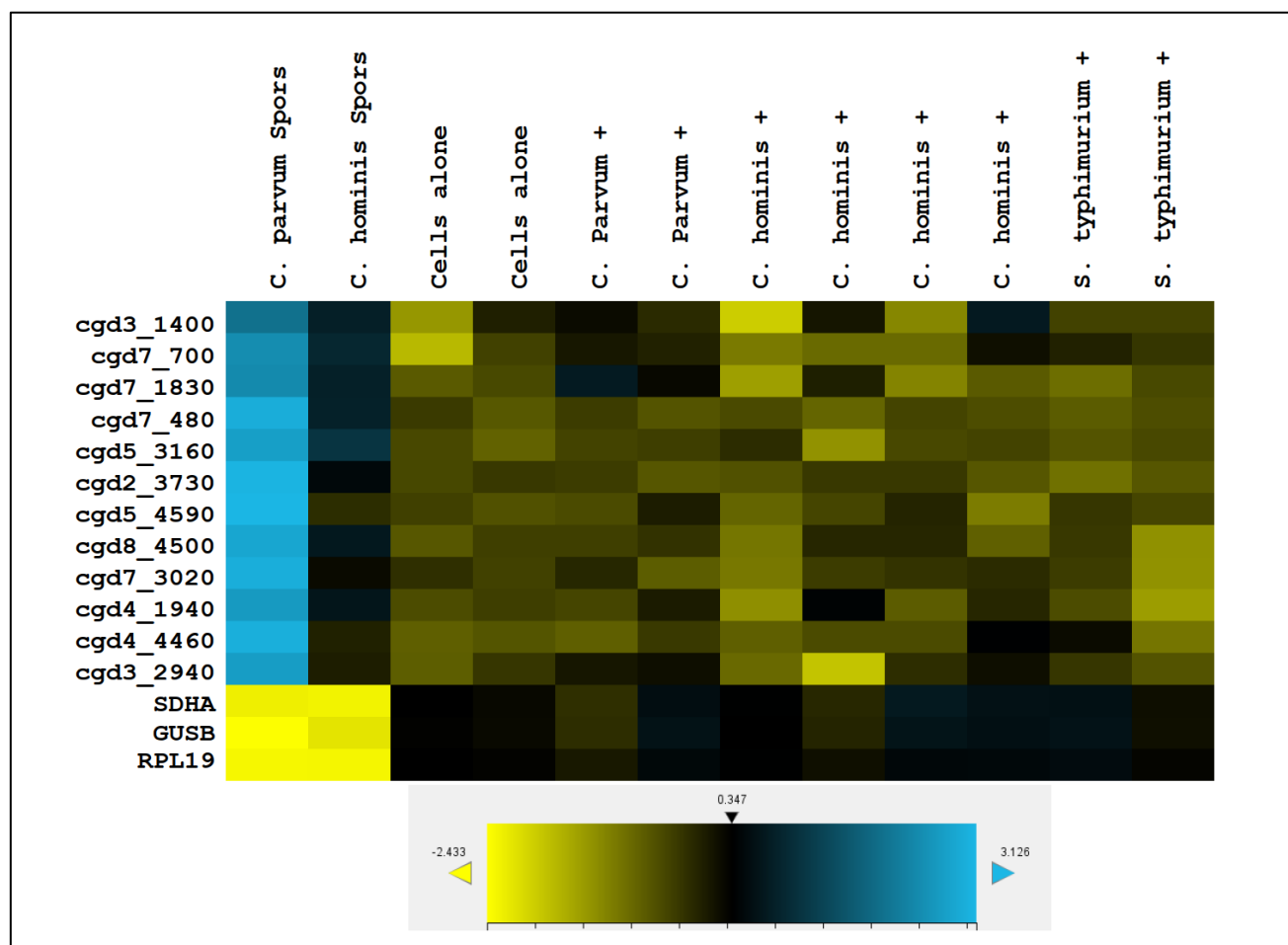
Figure 6.3. Heatmap showing the relative abundance of mRNA transcripts of the 12 parasite genes of interest and reference host genes in each sample from the 48 h experiment. (+) signifies infected cells. mRNA transcript abundance is calculated using Euclidean distance - the distance between two samples or genes is calculated as the square root of the sum of squared differences in their log count values.
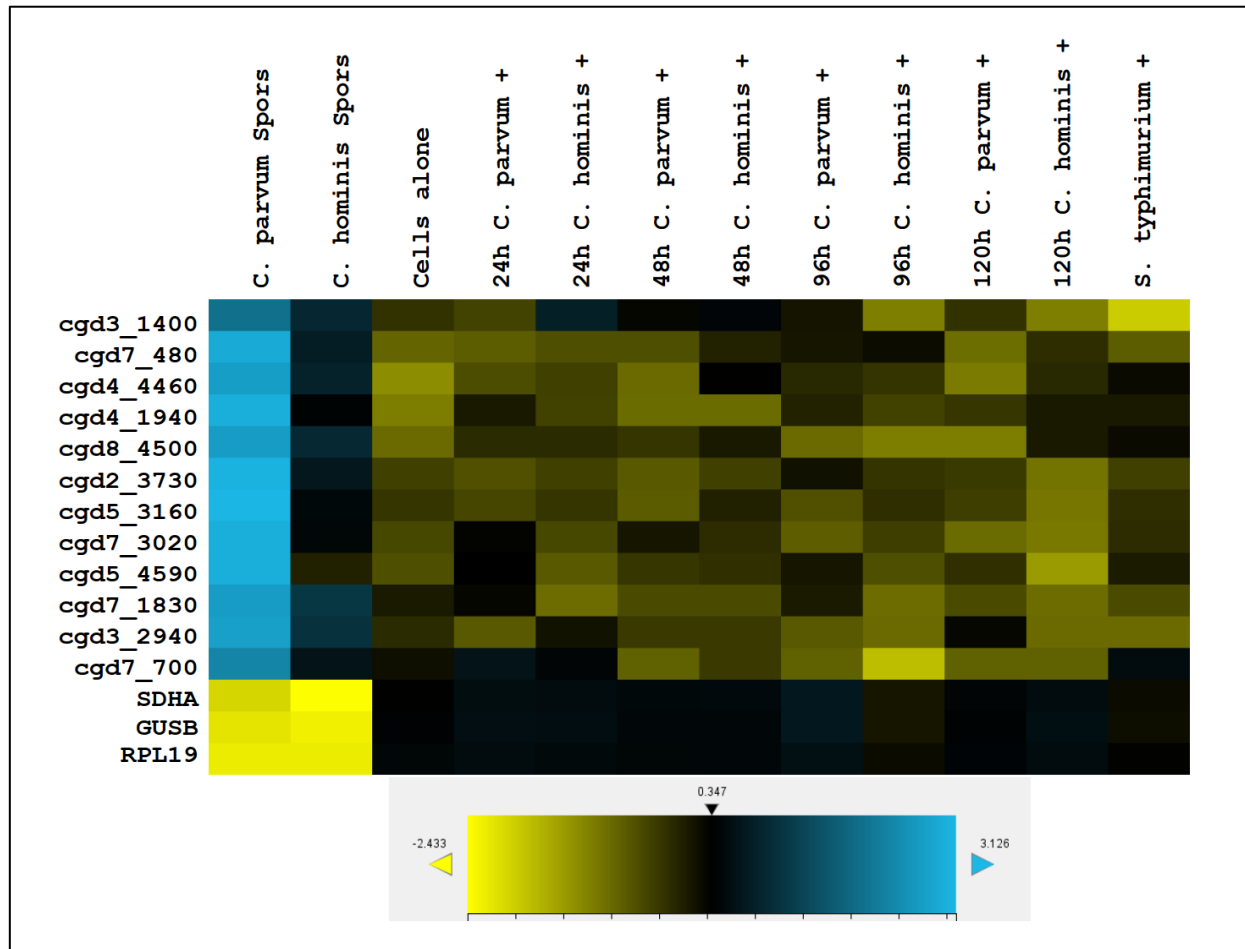
Figure 6.4. Heatmap showing the relative abundance of mRNA transcripts of the 12 parasite genes of interest and reference host genes in each sample from the time series experiment. (+) signifies infected cells. mRNA transcript abundance is calculated using Euclidean distance - the distance between two samples or genes is calculated as the square root of the sum of squared differences in their log count values.

6.5 Discussion

The most significant and oft-repeated limitation in the development of effective treatments and vaccines against cryptosporidiosis is the lack of efficient *in vitro* platforms for the manipulation of the parasite. Nevertheless, in recent years and despite limited options a lot of progress has been made in the understanding of the transcriptome of *Cryptosporidium* during infection. Most of these studies have relied on the HCT-8 cell line, which some studies have had limited success when attempting to achieve complete and long-term cultivation of the parasite (Karanis, 2018b). Recently, the COLO-680N cell line was characterised and has proved successful in allowing the full progression of the life cycle of *Cryptosporidium* without the usual premature death of the cells (Miller et al., 2018). For this reason, it is important to re-assess the data gained from the HCT-8 using COLO-680N. The argument for this is that because COLO-680N allows for the complete development of the parasite without the need for specialised media or mechanical structures, such as hollow fibre technology, more insight into the expression of genes during infection can be gained. In addition, COLO-680N is equally efficient at culturing *C. hominis* as it is with *C. parvum*, and since most studies have utilised *C. parvum* because of its relative ease of culture this provides a platform where the main species infecting humans can also be studied (see Chapter 5).

In this study data from a study by Matos et al., (2019) was analysed and the top 50 genes expressed intracellularly and extracellularly were extracted, from these 48 of the former and 48 of the latter were selected due to size restrictions on our NanoString panel. Using NanoString the expression of these genes in the COLO-680N cell line was assessed, comparing it to sporozoites of both *C. parvum* and *C. hominis*. The raw counts showed very low levels of the expression of *Cryptosporidium*-specific genes in all the infected cells, approaching the limit of detection of the nCounter machine. Similar to a data from a similar study conducted by Lippuner et al., (2018) all the parasite genes were expressed at higher levels in the sporozoites compared to infected cells. The differences in the expression of parasite genes between *C. parvum* and *C. hominis* sporozoites could be due to relatively low efficiency of excystation of the *C. hominis* oocysts (a common problem when using oocysts purified from human faecal samples), or lower specificity of mRNA

probes because most studies on the transcriptome of *Cryptosporidium* have been conducted using *C. parvum*.

Two significant limitations that might account for these results will be explained: Firstly, although the infectious dose of excysted oocysts was high (an MOI of 80), *Cryptosporidium* does not uniformly infect all cells they come in contact with *in vitro*, they tend to localise in a subset of cells and leave most untouched (see Chapter 5). This may be the reason for the low concentration of parasite RNA in our samples. Furthermore, a previous study by Lippuner et al., (2018) analysed the transcriptome of the parasite during infection using both *in vitro* and *in vivo* platforms. Analysis of the data from that study showed significant differences in the most highly expressed genes between both culture platforms. This suggests that the host exerts significant pressure on which genes are expressed by the parasite. This is to be expected given that *Cryptosporidium* relies on the host for most of its metabolic processes (Cacciò & Widmer, 2013). Therefore, *Cryptosporidium* cultured in different cell lines may show differing patterns of gene expression. COLO-680N is an oesophageal cell line while HCT-8 is a colorectal cell line, and while both cell lines are epithelial the organ of origin might produce differences in gene expression. Both these limitations could be solved in future studies by increasing the infectious dose and amplification of the *Cryptosporidium* RNA to allow for accurate analysis of the transcriptome of the parasite intracellularly and extracellularly. The most likely reason for the low expression of parasite genes in this study could be that the infection was only marginally or not at all successful. Alternatives methods for the verification of infection, such as *in situ* staining and qPCR were not employed, so it could not be definitively stated that infection was successful in this study. Such verification of successful infection would have to be the first step carried out for the development of this system. Due to the low levels of expression observed in this study, no inferences could be made on this point.

Twelve particular genes identified in previous studies as potential drug targets were also analysed in this study. Of note, MEDLE-2 (cgd5_4590) was found to be highly expressed in sporozoites, which verifies its classification as a gene crucial to parasite invasion (Su et al., 2019). Cgd3_1400, a pyrophosphate-fructose-6-phospate-1-phospho-transferase, showed the lowest level of expression in sporozoites. Analysis of the expression of this

gene in the infected cells showed relatively higher levels of expression at the 48 h and 96 h time points when compared to all the other *Cryptosporidium*-specific genes (Figure 6.3. & 6.4.). Considering the low levels of expression observed in the infected cells this suggests that cgd3_1400 might be a strong vaccine candidate, as was stated in a previous study.

Further research should analyse the complete transcriptome of the parasite in the COLO-680N cell line across its life cycle, using multiple *Cryptosporidium* species that have been found in humans. Also, validation of infection and the implementation of an amplification step after RNA purification might provide better results and allow for the comparison of the expression of parasite genes. This would improve our understanding and advance the identification of suitable drug targets, advancing the testing of novel drug compounds in this promising *in vitro* platform.

7 General Discussion

7.1 Background

This thesis sought to understand the role that genetic diversity plays in cryptosporidiosis and giardiasis, and what the implications of such a role would be for the epidemiology of these diseases in New Zealand. Chapter 1 provided a general introduction to the taxonomy and life cycle of *Cryptosporidium* and *Giardia*, and outlines the current understanding of the infectivity, transmission and risk factors associated with these organisms. In preparation for the studies carried out in this thesis, Chapter 1 also provides an introduction to the characteristics of outbreaks of cryptosporidiosis and giardiasis, highlighting the public health significance of these diseases that exact such a heavy toll on children and immunocompromised individuals. Lastly, Chapter 1 provides an overview of the currently available *in vitro* platforms for the culture and manipulation of both parasites, shining a light on the limitations that hinder the understanding of the disease mechanisms of these parasites and the development of effective therapies.

7.2 Genetic Diversity

Chapters 3 and 4 are based on the hypothesis that NGS is capable of revealing the genetic diversity, undetectable by consensus sequencing technologies (Sanger), present within faecal samples from patients affected by cryptosporidiosis or giardiasis. As shown in Chapter 1, recent advances in molecular techniques have greatly improved our understanding of the taxonomy and epidemiology of these parasites. However, new improvements in molecular techniques are constantly being made and it is important to utilise these new methods to answer old questions. One of those questions is, are there subtypes of either parasite linking outbreaks that cannot be detected with conventional PCR and Sanger sequencing? Chapters 3 and 4 attempt to answer this question.

7.2.1 Genetic Diversity in *Giardia*

Chapter 3 focuses on genetic diversity in *Giardia*. A comparative analysis of three outbreaks and sporadic cases of giardiasis that occurred in New Zealand between 2010 and 2018 showed that NGS uncovered significant genetic diversity within host, previously undetected by Sanger sequencing. Varied subtypes of *G. intestinalis* were present in the majority of samples, however, unlike in the samples from cryptosporidiosis outbreaks, there were multiple samples in which only one subtype of *G. intestinalis* was present. The results showed that assemblage B was the most common variant of *Giardia* present in humans in New Zealand during the period of study. Taken together, these results suggest the dominant mode of transmission of *Giardia* across the patients from which the samples in this study were isolated was anthroponotic. Further studies should look into employing such metabarcoding of *Giardia* on a wider dataset to allow for the definitive characterisation of the dominant mode of transmission of *Giardia* across all of New Zealand. In the case that it is verified that anthroponotic transmission is the dominant mode in the country, public health officials would then be able to develop strategies to mitigate this spread among humans. Such strategies could include education of citizens on the potential for spread of parasites in venues such communal swimming pools and water parks and the restriction of vet and farm workers from handling livestock when infected. The results also showed that the previously rare assemblages A and E are increasingly common in this country.

With regards to outbreaks, the results in Chapter 3 once again showed that using NGS it is possible to identify genetic links between samples from outbreaks. The outbreaks of giardiasis that occurred in Gisborne in 2014 and Hawke's Bay in 2015 served as case studies to illustrate this point. The results showed that *G. intestinalis* sub-assemblage BIV was shared between all samples within each outbreak. This overcame the difficulty presented by the results of Sanger sequencing that identified multiple dominant subtypes within a single outbreak.

7.2.2 Genetic Diversity in *Cryptosporidium*

Chapter 4 focuses on genetic diversity within *Cryptosporidium*. Taking samples from sporadic cases and outbreaks of cryptosporidiosis that occurred in New Zealand between 2010 and 2018, the investigation demonstrates that, in the majority of cases, *Cryptosporidium* infections in humans display multiple species and subtypes within a single host. NGS is preferable to Sanger sequencing for detection of this diversity. The results show that, the majority of infections in humans in New Zealand during the period of study involved both *C. hominis* and *C. parvum* and often contained multiple subtypes within these species. In addition, it suggests that the dominant mode of transmission of *Cryptosporidium* in New Zealand is anthroponotic, as evidenced by the presence of *C. hominis* in the majority of the samples tested. Another benefit to applying NGS to samples from cases of cryptosporidiosis is that it can serve as an early warning system capable of identifying emerging subtypes of *Cryptosporidium* in a country. As outlined in Chapter 4, previous studies and public health data have shown that the dominant subtype present in a country can change as time progresses. By employing NGS it possible to identify potential candidates early and gain new insight into the evolutionary biology of the parasite at the same time. The results suggest that *C. hominis* IfA12G1R5 and IgA20 are emerging subtypes in New Zealand.

As previously stated, one of the aims of this thesis was to identify subtypes linking samples between outbreaks of cryptosporidiosis. Using the outbreaks that occurred in Christchurch in 2010 and Wellington in 2018 as case studies, this study demonstrates that the ability of NGS to sequence multiple reads within a sample can be used to identify common subtypes in outbreaks that show multiple subtypes according to consensus sequencing. The results showed that within the outbreaks that occurred in both of the aforementioned outbreaks *C. hominis* IbA10G2 was the subtype shared by all samples.

7.2.3 Conclusions and future research

Taken together, these results show that the application of NGS can provide a better understanding of the epidemiology of cryptosporidiosis and giardiasis in New Zealand, linking outbreaks and identifying emerging subtypes. The samples used in this study were anonymised. Further work should include patient data to ascertain the contribution that travel, ethnicity and socioeconomic factors have on transmission patterns and

epidemiological outcomes, particularly during outbreaks. For example, by analysis the genetic diversity of patients within an outbreak it would be possible to ascertain if the incident was due to a subtype of *Cryptosporidium* or *Giardia* common to another country, thereby allowing the assessment of the relative contribution travel makes to the disease load within New Zealand. Or such analysis would show the relative contribution of zoonotic transmission by comparing the diversity and abundant subtypes found in humans and livestock. As New Zealand has a substantial livestock industry, such an analysis would be able to aid public health bodies in developing strategies to mitigate the spread of cryptosporidiosis and giardiasis from animals to humans and vice versa. MLST has proved to be the next step in the accurate classification of pathogens. Particularly in *Giardia,* it has been found that subtyping of a single isolate at different loci can result in the assignation of different subtypes. This is because, of the most widely used loci for the characterisation of *Giardia,* each one shows a different level of polymorphisms, in terms of substitutions per nucleotide. As such, an MLST approach is seen to be the most accurate available tool for the classification of *Giardia* isolates. So, the use of a single loci has the potential to overlook a substantial amount of genetic diversity of the parasite that might be present in a population. Application of MLST would allow for the definitive association of specific subtypes of each parasite with specific observed symptoms, without any disagreement or ambiguities in assigned subtypes, as can be the case in metabarcoding studies using only one locus. In addition, MLST can be incorporated into diagnostic laboratories, giving public health officials and epidemiologists accurate data on the subtypes of each parasite present within a population. Combined with WGS, application of this technology could provide more insight into the evolution of both parasites, the emergence of new dominant subtypes, and evolution of resistance and virulence factors. This will get rid of inconsistences in nomenclature (like in the case of *Giardia* assemblages), and, by providing more data to public health authorities, allow for outbreak management in real-time. Application of WGS in the diagnosis of *Cryptosporidium* and *Giardia,* aside from improving the tracking of outbreaks, would have the added benefit of allow for the monitoring of drug resistance markers in patients that don't respond to treatment. Also, a comparative analysis of the parasite and host genomes from asymptomatic vs symptomatic individuals could provide insight into the relative contribution specific subtypes of the parasites play in cases of mixed infections.

7.3 *In vitro* manipulation

The general introduction in Chapter 1 showed that there are limited treatment options available for patients suffering from cryptosporidiosis and giardiasis, and the clinical manifestation can differ depending on the subtype of the parasite with which in individual is infected. A significant limitation hampering the development of new therapies, and the understanding of the factors determining differences in symptoms between individuals, is the lack of efficient *in vitro* techniques for the culture and manipulation of *Cryptosporidium* and *Giardia*. This thesis sought to develop methods for culture to answer the question of whether contact rates or parasite subtypes determine case numbers during outbreaks and sporadic events.

7.3.1 *Giardia* purification

A factor impeding the development of an *in vitro* platform to assess the infectivity of *Giardia* is the lack of an efficient method for the purification of cysts from faecal samples. As outlined in Chapter 3, current methods do not adequately remove faecal contaminants and produce low yields of cysts. For this reason, in Chapter 2, outlines the development of a newly published method for the purification of *Giardia* cysts from faecal samples, overcoming the aforementioned limitations. The results show that my method eliminates the majority of the faecal matter in the sample and gives a 10-fold increase in the yield of cysts from a single gram of faeces. This method has the potential to aid in the development of *in vitro* systems for the manipulation of *Giardia*.

7.3.2 Infectivity of *Cryptosporidium*

One of the aims of this thesis was the development of an *in vitro* assay to assess what role the infectivity of specific species and subtypes of *Cryptosporidium* play on the size of outbreaks and transmission of the parasite. Chapter 5 outlines the steps taken to develop a system capable of assessing the infectivity of *C. hominis* and *C. parvum* in the COLO-680N cell line using flow cytometry. Using the Sporo-glo™ fluorescent antibody to detect

infection it was shown that when infected cells are analysed flow cytometrically it is possible to detect *C. parvum* as well as *C. hominis* infection in the COLO-680N cell line. However, low levels of infection were observed at low MOIs, possible reasons for this are discussed below (7.3.3.). In addition, chapter 5 shows the preliminary characterisation a new signal that could potentially be used to visually detect infection without the need for a fluorescent antibody. Assessment of controls suggests that this signal is specific to *Cryptosporidium* infection in the COLO-680N cell line.

The COLO-680N cell line encourages the progression of the complete life cycle of *Cryptosporidium*. This makes it preferable to the widely used cell lines, such as HCT-8 and Caco-2, which show mixed success in the cultivation of the full life cycle of the parasite and reduced viability when challenged with *Cryptosporidium*. In Chapter 6 I aimed to build upon the results of Chapter 5 by analysing the expression of parasite and host genes during the course of infection in an effort to see if it is a viable platform for assessing the genetic determinants that underpin the differences in infectivity observed in *Cryptosporidium* subtypes. The results revealed a low level of expression of parasite genes that was too close to the limit of detection achievable by NanoString technology. This meant that no comparisons between or within species could be made.

7.3.3 Conclusions and future research

A new method for the purification of *Giardia* cysts from faecal samples was developed with the aim of utilising it in studies of the infectivity of the various subtypes of *G. intestinalis*. However, due to time constraints, this experiment was not conducted. Considering the differing symptoms and associated sequelae observed in human patients suffering from giardiasis, it would be useful to conduct such an *in vitro* study to ascertains if any differences in infectivity exist between subtypes of *G. intestinalis*. Furthermore, the ability to produce a substantial amount of purified cysts will aid WGS studies were the presence of significant levels of contaminating genetic material would be a hinderance. As stated in the general introduction, previous studies have shown that assemblage A infections are associated with short acute infections, while assemblage B is associated

with chronic infections. A WGS analysis of each of these assemblages would highlight what areas of their genomes are under positive selection in each of these instances and give a better understanding of the factors influencing infectivity and transmission of *G. intestinalis*. To that end, recent studies have used single-cell whole genome amplification of to acquire WGS data for assemblages C and D from isolates extracted from dogs. This highlighted various genes related to host specificity and uncovered differences in heterozygosity between these assemblages. Furthermore, genomes for assemblages A, B and E, and *G. muris* are now available on GiardiaDB. By utilising such data provided by the application of WGS, public health officials would be able to properly attribute the source of an outbreak and track the spread of infection within a population, allowing more effective monitoring of small and large scale outbreaks of giardiasis.

G. intestinalis trophozoites undergo periodically changes in expression of a family of surface proteins termed variant-specific surface proteins (VSP). This mechanism of antigenic variation is employed by the parasite to evade an immune response by the host. It is thought that RNAi plays a role in this but much about this process, included the mechanisms that triggers this switching, remains to unknown. Further adding to the mystery around VSPs, it has been found that the VSP repertoire can differ between and within assemblages. The ability to purify large quantities of cysts with negligible contamination would aid the genomic and transcriptomic studies needed to properly understand the factors dictating this switch in expression of surface proteins. An understanding of this mechanism would identify new and effective drug targets and aid in the development of effective vaccines against this ubiquitous parasite.

The conclusions in Chapter 5 were hampered by lack of verification of the presence of infective foci or proliferation of parasite genetic material in the system. When one considers the low levels of infection, according to expression of the anti-Spor FITC marker, were observed in samples at low MOIs it could not be definitively stated that infection was successful. However, the titratable increase in expression of sigM and analysis of the spectral signatures of infected cells suggest that some level of infection was achieved. Some improvements to the *in vitro* system include, validation of sigM by extraction and staining of cells expressing the novel signal, and the testing of the binding efficacy of the anti-Spor FITC antibody between species to allow for accurate comparison

of levels of infectivity between multiple species and subtypes. If these improvements are made, this system could be used in conjunction with genomic techniques to definitively ascertain if any differences in infectivity exists between subtypes of *Cryptosporidium.* Also, if transcriptomic and proteomic techniques are incorporated, the system would allow for the understanding of what mechanisms underpinning the differences in infectivity observed between subtypes of the parasite. Such understanding of the mechanisms dictating the severity of the disease would aid public health officials in apportioning resources to the handling of outbreaks depending on the subtype of the parasite present i.e. if particularly infectious subtypes of the parasite are identified within a population appropriate measures can be taken to limit its spread, and patients can be given information on possible sequelae that might occur as a result of infection with a specific parasite.

The results from Chapter 6 showed that with some improvements, such as the amplification of *Cryptosporidium* RNA purified from infected cells, the technique I have refined could be capable of increasing our understanding of the expression of parasite genes during infection. Future studies should investigate the complete transcriptomic profile of *Cryptosporidium* during infection in the COLO-680N cell line.

An important step in discovery and development of drugs against any pathogen involves high-throughput screening (HTS) of compound librarys. Although, animal models exist for the proliferation of *Cryptosporidium*, such systems are not suitable for HTS. For HTS simple *in vitro* platforms are the preferred method, and currently, cultured cells propagated in two-dimensions (2D) on plastic surfaces optimized for tissue culture is the preferred method. For this reason, the progress outlined in this thesis could aid in the development of such a system, which would boost the development novel therapies and effective vaccines against *Cryptosporidium*. Due the fact that cryptosporidiosis can be severe in children and the immunocompromised, there will be a large market for any effective therapies once they are developed.

7.4 Conclusion

This thesis investigated the role the genetic diversity of *Cryptosporidium* and *Giardia* play in host-pathogen interactions. It demonstrates the benefits of NGS over Sanger sequencing in capturing the genetic diversity in humans infected with either parasite and in providing a clearer picture of the epidemiology of cryptosporidiosis and giardiasis in New Zealand. An *in vitro* assay with the potential to assess infection of *C. hominis* and *C. parvum* using flow cytometry and transcriptomic techniques was developed. These steps could inform public health strategies to combat these diseases and assist the development of new therapies to treat two of the most common causes of diarrhoea worldwide.

8 References

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

Abernathy, J., Corkill, C., Hinojosa, C., Li, X., & Zhou, H. (2013). *Deletions in the pyruvate pathway of Salmonella Typhimurium alter SPI1-mediated gene expression and infectivity* (Vol. 4). https://doi.org/10.1186/2049-1891-4-5

Abeywardena, H., Jex, A. R., Nolan, M. J., Haydon, S. R., Stevens, M. A., McAnulty, R. W., & Gasser, R. B. (2012). Genetic characterisation of *Cryptosporidium* and *Giardia* from dairy calves: Discovery of species/genotypes consistent with those found in humans. *Infection, Genetics and Evolution*, *12*(8), 1984–1993. https://doi.org/10.1016/j.meegid.2012.08.004

Adam, E. A., Yoder, J. S., Gould, L. H., Hlavsa, M. C., & Gargano, J. W. (2016). Giardiasis outbreaks in the United States, 1971-2011. In *Epidemiology and Infection* (Vol. 144, Issue 13, pp. 2790–2801). Cambridge University Press. https://doi.org/10.1017/S0950268815003040

Adam, R. D. (2021). *Giardia duodenalis*: Biology and Pathogenesis. *Clinical Microbiology Reviews*. https://doi.org/10.1128/CMR.00024-19

Adeyemo, F. E., Singh, G., Reddy, P., & Stenström, T. A. (2018). Methods for the detection of *Cryptosporidium* and *Giardia*: From microscopy to nucleic acid based tools in clinical and environmental regimes. *Acta Tropica*, *184*, 15–28. https://doi.org/10.1016/J.ACTATROPICA.2018.01.011

Afshin, B., Jafar, M., Esmaeel, F., & Reza, G. (2011). Introducing a simple and economical method to purify *Giardia lamblia* cysts. *African Journal of Biotechnology*, *10*(42), 8498–8501. https://doi.org/10.5897/AJB11.391

Alvarado, M. E., & Wasserman, M. (2006). Quick and efficient purification of *Giardia intestinalis* cysts from fecal samples. *Parasitology Research*, *99*(3), 300–302. https://doi.org/10.1007/s00436-006-0143-x

Alvarez-Pellitero, P., Quiroga, M. I., Sitjà-Bobadilla, A., Redondo, M. J., Palenzuela, O., Padrós, F., Vázquez, S., & Nieto, J. M. (2004). *Cryptosporidium scophthalmi* n. sp. (Apicomplexa: Cryptosporidiidae) from cultured turbot *Scophthalmus maximus*. Light and electron microscope description and histopathological study. *Diseases of Aquatic Organisms*, *62*(1–2), 133–145. https://doi.org/10.3354/DAO062133

Amadi, B., Mwiya, M., Musuku, J., Watuka, A., Sianongo, S., Ayoub, A., & Kelly, P. (2002). Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: A randomised controlled trial. *Lancet*, *360*(9343), 1375–1380. https://doi.org/10.1016/S0140-6736(02)11401-2

Audebert, C., Bonardi, F., Caboche, S., Guyot, K., Touzet, H., Merlin, S., Gantois, N., Creusy, C., Meloni, D., Mouray, A., Viscogliosi, E., Certad, G., Benamrouz-Vanneste, S., & Chabé, M. (2020). Genetic basis for virulence differences of various *Cryptosporidium parvum* carcinogenic isolates. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-64370-0

Baragaña, B., Forte, B., Choi, R., Hewitt, S. N., Bueren-Calabuig, J. A., Pisco, J. P., Peet, C., Dranow, D. M., Robinson, D. A., Jansen, C., Norcross, N. R., Vinayak, S., Anderson, M., Brooks, C. F., Cooper, C. A., Damerow, S., Delves, M., Dowers, K., Duffy, J., … Gilbert, I. H. (2019). Lysyl-tRNA synthetase as a drug target in malaria and cryptosporidiosis.

Proceedings of the National Academy of Sciences of the United States of America, *116*(14), 7015–7020. https://doi.org/10.1073/pnas.1814685116

Barker, I. K., & Carbonell, P. L. (1974). *Cryptosporidium agni* sp.n. from lambs, and *Cryptosporidium bovis* sp.n. from a calf, with observations on the oocyst. *Zeitschrift Für Parasitenkunde 1974 44:4*, *44*(4), 289–298. https://doi.org/10.1007/BF00366112

Bartelt, L. A., & Sartor, R. B. (2015). Advances in understanding *Giardia*: Determinants and mechanisms of chronic sequelae. *F1000Prime Reports*, *7*. https://doi.org/10.12703/P7-62

Barugahare, R., Dennis, M. M., Becker, J. A., & Šlapeta, J. (2011). Detection of *Cryptosporidium molnari* oocysts from fish by fluorescent-antibody staining assays for *Cryptosporidium* spp. affecting humans. *Applied and Environmental Microbiology*, *77*(5), 1878–1880. https://doi.org/10.1128/AEM.02691-10

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1195

Bingham, A. K., Jarroll, E. L., Meyer, E. A., & Radulescu, S. (1979). *Giardia* sp.: Physical factors of excystation in vitro, and excystation vs eosin exclusion as determinants of viability. *Experimental Parasitology*, *47*(2), 284–291. https://doi.org/10.1016/0014-4894(79)90080-8

Birky, C. W. (2010). *Giardia* Sex? Yes, but how and how much? *Trends in Parasitology*, *26*(2), 70–74. https://doi.org/10.1016/J.PT.2009.11.007

Bogan, J. E. (2019). Disinfection Techniques for *Cryptosporidium*. *Journal of Dairy & Veterinary Sciences*, *7*(4), 1–3. https://doi.org/10.19080/jdvs.2018.07.555718

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K.,

Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. In *Nature Biotechnology* (Vol. 37, Issue 8, pp. 852–857). Nature Publishing Group. https://doi.org/10.1038/s41587-019-0209-9

Bones, A. J., Jossé, L., More, C., Miller, C. N., Michaelis, M., & Tsaousis, A. D. (2019). Past and future trends of *Cryptosporidium* in vitro research. *Experimental Parasitology*, *196*, 28–37. https://doi.org/10.1016/J.EXPPARA.2018.12.001

Bouzid, M., Hunter, P. R., Chalmers, R. M., & Tyler, K. M. (2013). *Cryptosporidium* pathogenicity and virulence. *Clinical Microbiology Reviews*, *26*(1), 115–134. https://doi.org/10.1128/CMR.00076-12

Boxell, A., Hijjawi, N., Monis, P., & Ryan, U. (2008). Comparison of various staining methods for the detection of *Cryptosporidium* in cell-free culture. *Experimental Parasitology*, *120*(1), 67–72. https://doi.org/10.1016/j.exppara.2008.04.023

Braima, K., Zahedi, A., Oskam, C., Reid, S., Pingault, N., Xiao, L., & Ryan, U. (2019). Retrospective analysis of *Cryptosporidium* species in Western Australian human populations (2015–2018), and emergence of the *C. hominis* IfA12G1R5 subtype. *Infection, Genetics and Evolution*, *73*, 306–313. https://doi.org/10.1016/j.meegid.2019.05.018

Brynildsrud, O., Tysnes, K. R., Robertson, L. J., & Debenham, J. J. (2018). *Giardia duodenalis* in primates: Classification and host specificity based on phylogenetic analysis of sequence data. *Zoonoses and Public Health*, *65*(6), 637–647. https://doi.org/10.1111/ZPH.12470

Budu-Amoako, E., Greenwood, S. J., Dixon, B. R., Barkema, H. W., & McClure, J. T. (2011). Foodborne illness associated with *Cryptosporidium* and *Giardia* from livestock. In *Journal of Food Protection* (Vol. 74, Issue 11, pp. 1944–1955). Allen Press. https://doi.org/10.4315/0362-028X.JFP-11-107

Bushen, O. Y., Kohli, A., Pinkerton, R. C., Dupnik, K., Newman, R. D., Sears, C. L., Fayer, R., Lima, A. A. M., & Guerrant, R. L. (2007). Heavy cryptosporidial infections in children in northeast Brazil: comparison of *Cryptosporidium hominis* and *Cryptosporidium parvum*. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *101*(4), 378–384. https://doi.org/10.1016/J.TRSTMH.2006.06.005

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

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

Cacciò, S. M., & Widmer, G. (2013). *Cryptosporidium : parasite and disease*. Springer Vienna.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cama, V. A., Bern, C., Roberts, J., Cabrera, L., Sterling, C. R., Ortega, Y., Gilman, R. H., & Xiao, L. (2008). *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. *Emerging Infectious Diseases*, *14*(10), 1567–1574. https://doi.org/10.3201/eid1410.071273

Carey, C. M., Lee, H., & Trevors, J. T. (2004). Biology, persistence and detection of *Cryptosporidium parvum* and *Cryptosporidium hominis* oocyst. *Water Research*, *38*(4), 818–862. https://doi.org/10.1016/j.watres.2003.10.012

Carter, B. L., Chalmers, R. M., & Davies, A. P. (2020). Health sequelae of human cryptosporidiosis in industrialised countries: a systematic review. *Parasites & Vectors 2020 13:1*, *13*(1), 1–14. https://doi.org/10.1186/S13071-020-04308-7

Casadevall, A., & Pirofski, L. (2001). Host-Pathogen Interactions: The Attributes of

Virulence. *The Journal of Infectious Diseases*, *184*(3), 337–344. https://doi.org/10.1086/322044

Castellanos-Gonzalez, A., Martinez-Traverso, G., Fishbeck, K., Nava, S., & White, A. C. (2019). Systematic gene silencing identified *Cryptosporidium* nucleoside diphosphate kinase and other molecules as targets for suppression of parasite proliferation in human intestinal cells. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-48544-z

Cernikova, L., Faso, C., & Hehl, A. B. (2018). Five facts about *Giardia lamblia*. *PLoS Pathogens*, *14*(9), e1007250. https://doi.org/10.1371/journal.ppat.1007250

Chalmers, R. M., Campbell, B. M., Crouch, N., Charlett, A., & Davies, A. P. (2011). Comparison of diagnostic sensitivity and specificity of seven *Cryptosporidium* assays used in the UK. *Journal of Medical Microbiology*, *60*(11), 1598–1604. https://doi.org/10.1099/jmm.0.034181-0

Chalmers, R. M., & Katzer, F. (2013). Looking for *Cryptosporidium*: the application of advances in detection and diagnosis. *Trends in Parasitology*, *29*(5), 237–251. https://doi.org/10.1016/J.PT.2013.03.001

Chavez, M. A., & Jr, A. C. W. (2018). Novel treatment strategies and drugs in development for cryptosporidiosis. *Https://Doi.Org/10.1080/14787210.2018.1500457*, *16*(8), 655–661. https://doi.org/10.1080/14787210.2018.1500457

Checkley, W., White, A. C., Jaganath, D., Arrowood, M. J., Chalmers, R. M., Chen, X. M., Fayer, R., Griffiths, J. K., Guerrant, R. L., Hedstrom, L., Huston, C. D., Kotloff, K. L., Kang, G., Mead, J. R., Miller, M., Petri, W. A., Priest, J. W., Roos, D. S., Striepen, B., … Houpt, E. R. (2015). A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. In *The Lancet Infectious Diseases* (Vol. 15, Issue 1, pp. 85–94). Lancet Publishing Group. https://doi.org/10.1016/S1473-3099(14)70772-8

Čondlová, Š., Horčičková, M., Sak, B., Květoňová, D., Hlásková, L., Konečný, R., Stanko, M.,

McEvoy, J., & Kváč, M. (2018). *Cryptosporidium apodemi* sp. n. and *Cryptosporidium ditrichi* sp. n. (Apicomplexa: Cryptosporidiidae) in *Apodemus* spp. *European Journal of Protistology*, *63*, 1–12. https://doi.org/10.1016/J.EJOP.2017.12.006

Cooper, M. A., Adam, R. D., Worobey, M., & Sterling, C. R. (2007). Population Genetics Provides Evidence for Recombination in *Giardia*. *Current Biology*, *17*(22), 1984–1988. https://doi.org/10.1016/J.CUB.2007.10.020

Current, W. L., & Garcia, L. S. (1991). Cryptosporidiosis. In *Clinical Microbiology Reviews* (Vol. 4, Issue 3, pp. 325–358). American Society for Microbiology Journals. https://doi.org/10.1128/CMR.4.3.325

Current, W. L., Upton, S. J., & Haynes, T. B. (1986). The Life Cycle of *Cryptosporidium baileyi* n. sp. (Apicomplexa, Cryptosporidiidae) Infecting Chickens. *The Journal of Protozoology*, *33*(2), 289–296. https://doi.org/10.1111/j.1550-7408.1986.tb05608.x

Delahoy, M. J., Omore, R., Ayers, T. L., Schilling, K. A., Blackstock, A. J., Ochieng, J. B., Moke, F., Jaron, P., Awuor, A., Okonji, C., Juma, J., Farag, T. H., Nasrin, D., Panchalingam, S., Nataro, J. P., Kotloff, K. L., Levine, M. M., Oundo, J., Roellig, D. M., ... O'Reilly, C. E. (2018). Clinical, environmental, and behavioral characteristics associated with *Cryptosporidium* infection among children with moderate-to-severe diarrhea in rural western Kenya, 2008–2012: The Global Enteric Multicenter Study (GEMS). *PLOS Neglected Tropical Diseases*, *12*(7), e0006640. https://doi.org/10.1371/JOURNAL.PNTD.0006640

Deng, M., Rutherford, M. S., & Abrahamsen, M. S. (2004). Host intestinal epithelial response to *Cryptosporidium parvum*. *Advanced Drug Delivery Reviews*, *56*(6), 869–884. https://doi.org/10.1016/j.addr.2003.10.034

Dhal, A. K., Pani, A., Mahapatra, R. K., & Yun, S. I. L. (2018). In-silico screening of small molecule inhibitors against Lactate Dehydrogenase (LDH) of *Cryptosporidium parvum*. *Computational Biology and Chemistry*, *77*, 44–51. https://doi.org/10.1016/j.compbiolchem.2018.09.002

Di Genova, B. M., & Tonelli, R. R. (2016). Infection strategies of intestinal parasite pathogens and host cell responses. *Frontiers in Microbiology*, *7*(MAR), 1–16. https://doi.org/10.3389/fmicb.2016.00256

Dib, H. H., Lu, S. Q., & Wen, S. F. (2008). Prevalence of *Giardia lamblia* with or without diarrhea in South East, South East Asia and the Far East. In *Parasitology Research* (Vol. 103, Issue 2, pp. 239–251). https://doi.org/10.1007/s00436-008-0968-6

Dixon, B. R. (2021). *Giardia duodenalis* in humans and animals – Transmission and disease. *Research in Veterinary Science*, *135*, 283–289. https://doi.org/10.1016/J.RVSC.2020.09.034

Dong, S., Yang, Y., Wang, Y., Yang, D., Yang, Y., Shi, Y., Li, C., Li, L., Chen, Y., Jiang, Q., & Zhou, Y. (2020). Prevalence of *Cryptosporidium* Infection in the Global Population: A Systematic Review and Meta-analysis. *Acta Parasitologica*, *65*(4), 882–889. https://doi.org/10.2478/s11686-020-00230-1

Drummond, J. D., Boano, F., Atwill, E. R., Li, X., Harter, T., & Packman, A. I. (2018). *Cryptosporidium* oocyst persistence in agricultural streams -a mobile-immobile model framework assessment. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-22784-x

Eastel, J. M., Lam, K. W., Lee, N. L., Lok, W. Y., Tsang, A. H. F., Pei, X. M., Chan, A. K. C., Cho, W. C. S., & Wong, S. C. C. (2019). Application of NanoString technologies in companion diagnostic development. *Expert Review of Molecular Diagnostics*, *19*(7), 591–598. https://doi.org/10.1080/14737159.2019.1623672

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

Einarsson, E., Ma'ayeh, S., & Svärd, S. G. (2016a). An up-date on *Giardia* and giardiasis. *Current Opinion in Microbiology*, *34*, 47–52. https://doi.org/10.1016/j.mib.2016.07.019

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

Elliott, D. A., & Clark, D. P. (2000). *Cryptosporidium parvum* induces host cell actin accumulation at the host- parasite interface. *Infection and Immunity*, *68*(4), 2315–2322. https://doi.org/10.1128/IAI.68.4.2315-2322.2000

Elwin, K., Hadfield, S. J., Robinson, G., Crouch, N. D., & Chalmers, R. M. (2012). *Cryptosporidium viatorum* n. sp. (Apicomplexa: Cryptosporidiidae) among travellers returning to Great Britain from the Indian subcontinent, 2007–2011. *International Journal for Parasitology*, *42*(7), 675–682. https://doi.org/10.1016/J.IJPARA.2012.04.016

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

Erlandsen, S. L., & Bemrick, W. J. (1987). SEM evidence for a new species, *Giardia psittaci*. *The Journal of Parasitology*, *73*(3), 623–629. https://doi.org/10.2307/3282146

Erlandsen, S. L., Bemrick, W. J., Wells, C. L., Feely, D. E., Knudson, L., Campbell, S. R., Van Keulen, H., & Jarroll, E. L. (1990). Axenic culture and characterization of *Giardia ardeae* from the great blue heron (*Ardea herodias*). *Journal of Parasitology*, *76*(5), 717–724. https://doi.org/10.2307/3282988

Fayer, R., & Santín, M. (2009). *Cryptosporidium xiaoi* n. sp. (Apicomplexa: Cryptosporidiidae) in sheep (*Ovis aries*). *Veterinary Parasitology*, *164*(2–4), 192–200. https://doi.org/10.1016/J.VETPAR.2009.05.011

Fayer, R., Santín, M., & Macarisin, D. (2010). *Cryptosporidium ubiquitum* n. sp. in animals and humans. *Veterinary Parasitology*, *172*(1–2), 23–32. https://doi.org/10.1016/J.VETPAR.2010.04.028

Fayer, R., Santín, M., & Trout, J. M. (2008). *Cryptosporidium ryanae* n. sp. (Apicomplexa: Cryptosporidiidae) in cattle (*Bos taurus*). *Veterinary Parasitology*, *156*(3–4), 191–198. https://doi.org/10.1016/J.VETPAR.2008.05.024

Fayer, R., Santín, M., & Xiao, L. (2005). *Cryptosporidium bovis* n. sp. (Apicomplexa: Cryptosporidiidae) in cattle (*Bos Taurus*). *Journal of Parasitology*, *91*(3), 624–629. https://doi.org/10.1645/GE-3435

Fayer, R., Trout, J. M., & Jenkins, M. C. (1998). Infectivity of *Cryptosporidium parvum* oocysts stored in water at environmental temperatures. *Journal of Parasitology*, *84*(6), 1165–1169. https://doi.org/10.2307/3284666

Fayer, R., Trout, J. M., Xiao, L., Morgan, U. M., Lal, A. A., & Dubey, J. P. (2001). *Cryptosporidium canis* n. sp. from domestic dogs. *Journal of Parasitology*, *87*(6), 1415–1422.

Feely, D. E. (1988). Morphology of the Cyst of *Giardia microti* by Light and Electron Microscopy. *The Journal of Protozoology*, *35*(1), 52–54. https://doi.org/10.1111/j.1550-7408.1988.tb04075.x

Feng, Y., Ryan, U. M., & Xiao, L. (2018). Genetic Diversity and Population Structure of *Cryptosporidium*. *Trends in Parasitology*, 1–15. https://doi.org/10.1016/j.pt.2018.07.009

Feng, Y., & Xiao, L. (2011). Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clinical Microbiology Reviews*, *24*(1), 110–140. https://doi.org/10.1128/CMR.00033-10

Feng, Y., & Xiao, L. (2017). Molecular Epidemiology of Cryptosporidiosis in China. *Frontiers in Microbiology*, *8*, 1701. https://doi.org/10.3389/fmicb.2017.01701

Filice, F. P. (1952). Studies on the cytology and life history of a *Giardia* from the laboratory rat. *Univ Calif Publ Zool*, *57*, 53–146.

Gaber, M., Galal, L. A. A., Hassan, D., Badary, D. M., Mohamed, I. M. A., & Elossily, N. (2020). Evidences of brain and lung invasion of a local water *Cryptosporidium parvum* isolate in comparison to Iowa strain: serological and immunohistochemical cytokine evaluation. *Annals of Parasitology*, *66*(3), 311–318. https://doi.org/10.17420/ap6603.269

Garcia-R, J. C., Ogbuigwe, P., Pita, A. B., Velathanthiri, N., Knox, M. A., Biggs, P. J., French, N. P., & Hayman, D. T. S. (2021). First report of novel assemblages and mixed infections of *Giardia duodenalis* in human isolates from New Zealand. *Acta Tropica*, *220*, 105969. https://doi.org/10.1016/j.actatropica.2021.105969

Garcia-R, J. C., Pita, A. B., Velathanthiri, N., French, N. P., & Hayman, D. T. S. (2020). Species and genotypes causing human cryptosporidiosis in New Zealand. *Parasitology Research*, *119*(7), 2317–2326. https://doi.org/10.1007/s00436-020-06729-w

Garcia–R, J. C., French, N., Pita, A., Velathanthiri, N., Shrestha, R., & Hayman, D. (2017a). Local and global genetic diversity of protozoan parasites: Spatial distribution of *Cryptosporidium* and *Giardia* genotypes. *PLoS Neglected Tropical Diseases*, *11*(7), 1–20. https://doi.org/10.1371/journal.pntd.0005736

Garcia–R, J. C., French, N., Pita, A., Velathanthiri, N., Shrestha, R., & Hayman, D. (2017b). Local and global genetic diversity of protozoan parasites: Spatial distribution of *Cryptosporidium* and *Giardia* genotypes. *PLOS Neglected Tropical Diseases*, *11*(7), e0005736. https://doi.org/10.1371/journal.pntd.0005736

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., … Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology 2008 26:3*, *26*(3), 317–325. https://doi.org/10.1038/nbt1385

Gibbons, C. L., Gazzard, B. G., Ibrahim, M. A. A., Morris-Jones, S., Ong, C. S. L., & Awad-El-Kariem, F. M. (1998). Correlation between markers of strain variation in

Cryptosporidium parvum: Evidence of clonality. *Parasitology International*, *47*(2), 139–147. https://doi.org/10.1016/S1383-5769(98)00012-9

Gilchrist, C. A. (2020). Eukaryome Impact on Human Intestine Homeostasis and Mucosal Immunology. In *Eukaryome Impact on Human Intestine Homeostasis and Mucosal Immunology*. Springer International Publishing. https://doi.org/10.1007/978-3-030-44826-4

Gillin, F. D., Boucher, S. E., Rossi, S. S., & Reiner, D. S. (1989). *Giardia lamblia*: The roles of bile, lactic acid, and pH in the completion of the life cycle in vitro. *Experimental Parasitology*. https://doi.org/10.1016/0014-4894(89)90185-9

Gómez-Couso, H., Freire-Santos, F., Amar, C. F. L., Grant, K. A., Williamson, K., Ares-Mazás, M. E., & McLauchlin, J. (2004). Detection of *Cryptosporidium* and *Giardia* in molluscan shellfish by multiplexed nested-PCR. *International Journal of Food Microbiology*, *91*(3), 279–288. https://doi.org/10.1016/j.ijfoodmicro.2003.07.003

Grinberg, A., Biggs, P. J., Dukkipati, V. S. R., & George, T. T. (2013). Extensive intra-host genetic diversity uncovered in *Cryptosporidium parvum* using Next Generation Sequencing. *Infection, Genetics and Evolution*, *15*, 18–24. https://doi.org/10.1016/j.meegid.2012.08.017

Harris, J. R., & Petry, F. (1999). *Cryptosporidium parvum*: Structural Components of the Oocyst Wall. *The Journal of Parasitology*, *85*(5), 839. https://doi.org/10.2307/3285819

Hijjawi, N., Estcourt, A., Yang, R., Monis, P., & Ryan, U. (2010). Complete development and multiplication of *Cryptosporidium hominis* in cell-free culture. *Veterinary Parasitology*, *169*(1–2), 29–36. https://doi.org/10.1016/j.vetpar.2009.12.021

Hijjawi, N. S., Meloni, B. P., Morgan, U. M., & Thompson, R. C. A. (2001). Complete development and long-term maintenance of *Cryptosporidium parvum* human and cattle genotypes in cell culture. *International Journal for Parasitology*, *31*(10), 1048–1055. https://doi.org/10.1016/S0020-7519(01)00212-0

Hillman, A., Ash, A., Elliot, A., Lymbery, A., Perez, C., & Thompson, R. C. A. (2016). Confirmation of a unique species of *Giardia*, parasitic in the quenda (*Isoodon obesulus*). *International Journal for Parasitology: Parasites and Wildlife*, *5*(1), 110–115. https://doi.org/10.1016/J.IJPPAW.2016.01.002

Hlavsa, M. C., Roellig, D. M., Seabolt, M. H., Kahler, A. M., Murphy, J. L., McKitt, T. K., Geeter, E. F., Dawsey, R., Davidson, S. L., Kim, T. N., Tucker, T. H., Iverson, S. A., Garrett, B., Fowle, N., Collins, J., Epperson, G., Zusy, S., Weiss, J. R., Komatsu, K., … Xiao, L. (2017). Using Molecular Characterization to Support Investigations of Aquatic Facility–Associated Outbreaks of Cryptosporidiosis — Alabama, Arizona, and Ohio, 2016. *MMWR. Morbidity and Mortality Weekly Report*, *66*(19), 493–497. https://doi.org/10.15585/mmwr.mm6619a2

Holubová, N., Sak, B., Horčičková, M., Hlásková, L., Květoňová, D., Menchaca, S., McEvoy, J., & Kváč, M. (2016). *Cryptosporidium avium* n. sp. (Apicomplexa: Cryptosporidiidae) in birds. *Parasitology Research 2016 115:6*, *115*(6), 2243–2251. https://doi.org/10.1007/S00436-016-4967-8

Holubová, N., Tůmová, L., Sak, B., Hejzlarová, A., Konečný, R., McEvoy, J., & Kváč, M. (2020). Description of *Cryptosporidium ornithophilus* n. sp. (Apicomplexa: Cryptosporidiidae) in farmed ostriches. *Parasites & Vectors 2020 13:1*, *13*(1), 1–17. https://doi.org/10.1186/S13071-020-04191-2

Holubová, N., Zikmundová, V., Limpouchová, Z., Sak, B., Konečný, R., Hlásková, L., Rajský, D., Kopacz, Z., McEvoy, J., & Kváč, M. (2019). *Cryptosporidium proventriculi* sp. n. (Apicomplexa: Cryptosporidiidae) in Psittaciformes birds. *European Journal of Protistology*, *69*, 70–87. https://doi.org/10.1016/J.EJOP.2019.03.001

Hooshyar, H., Rostamkhani, P., Arbabi, M., & Delavari, M. (2019). *Giardia lamblia* infection: Review of current diagnostic strategies. In *Gastroenterology and Hepatology from Bed to Bench* (Vol. 12, Issue 1, pp. 3–12). Research Institute for Gastroenterology and Liver Diseases. https://doi.org/10.22037/ghfbb.v0i0.1414

Hoque, M. E., Hope, V. T., Scragg, R., Kjellström, T., & Lay-Yee, R. (2001). Nappy handling

and risk of giardiasis. *Lancet*, *357*(9261), 1017–1018. https://doi.org/10.1016/S0140-6736(00)04251-3

Horčičková, M., Čondlová, Š., Holubová, N., Sak, B., Květoňová, D., Hlásková, L., Konečný, R., Sedláček, F., Clark, M., Giddings, C., McEvoy, J., & Kváč, M. (2019). Diversity of *Cryptosporidium* in common voles and description of *Cryptosporidium alticolis* sp. n. and *Cryptosporidium microti* sp. n. (Apicomplexa: Cryptosporidiidae). *Parasitology*, *146*(2), 220–233. https://doi.org/10.1017/S0031182018001142

Horlock-Roberts, K., Reaume, C., Dayer, G., Ouellet, C., Cook, N., & Yee, J. (2017). Drug-Free Approach To Study the Unusual Cell Cycle of *Giardia intestinalis. MSphere*, *2*(5). https://doi.org/10.1128/msphere.00384-16

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

Huang, D. B., & White, A. C. (2006a). An Updated Review on *Cryptosporidium* and *Giardia*. *Gastroenterology Clinics of North America*, *35*(2), 291–314. https://doi.org/10.1016/j.gtc.2006.03.006

Huang, D. B., & White, A. C. (2006b). An Updated Review on *Cryptosporidium* and *Giardia*. In *Gastroenterology Clinics of North America* (Vol. 35, Issue 2, pp. 291–314). Elsevier. https://doi.org/10.1016/j.gtc.2006.03.006

Hunter, P. R., Hughes, S., Woodhouse, S., Nicholas, R., Syed, Q., Chalmers, R. M., Verlander, N. Q., & Goodacre, J. (2004a). Health Sequelae of Human Cryptosporidiosis in Immunocompetent Patients. *Clinical Infectious Diseases*, *39*(4), 504–510. https://doi.org/10.1086/422649

Hunter, P. R., Hughes, S., Woodhouse, S., Nicholas, R., Syed, Q., Chalmers, R. M., Verlander, N. Q., & Goodacre, J. (2004b). Health Sequelae of Human Cryptosporidiosis in Immunocompetent Patients. *Clinical Infectious Diseases*, *39*(4), 504–510. https://doi.org/10.1086/422649

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

Inácio, S. V., Zucatto, A. S., de Aquino, M. C. C., Oliveira, B. C. M., Bresciani, K. D. S., Widmer, G., de Brito, R. L. L., Neto, L. da S., Nakamura, A. A., Meireles, M. V., Carvalho, J. G. B., & Gomes, J. F. (2017). First description of *Cryptosporidium hominis* GP60 genotype IkA20G1 and *Cryptosporidium parvum* GP60 genotypes IIaA18G3R1 and IIaA15G2R1 in foals in Brazil. *Veterinary Parasitology*, *233*, 48–51. https://doi.org/10.1016/j.vetpar.2016.11.021

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

Institute of Environmental Science and Research Ltd (ESR). (2018b). *Annual Summary of Outbreaks in New Zealand 2017* (Issue ESR).

Iseki, M. (1979). *Cryptosporidium* felis sp. n.(Protozoa Eimeriorina) from the domestic cat. *Jpn. J. Parasitol.*, *28*, 285.

Jarroll, E. L., & Hoff, J. C. (1988). Effect of disinfectants on *Giardia* cysts. *Critical Reviews in Environmental Control*, *18*(1), 1–28. https://doi.org/10.1080/10643388809388341

Jenkins, M. B., Eaglesham, B. S., Anthony, L. C., Kachlany, S. C., Bowman, D. D., & Ghiorse, W. C. (2010). Significance of wall structure, macromolecular composition, and surface polymers to the survival and transport of *Cryptosporidium parvum* oocysts. *Applied and Environmental Microbiology*, *76*(6), 1926–1934. https://doi.org/10.1128/AEM.02295-09

Ježková, J., Horcicková, M., Hlásková, L., Sak, B., Dana Kvetonová, Novák, J., Hofmannová, L., McEvoy, J., & Kvác, M. (2016). *Cryptosporidium testudinis* sp. n., *Cryptosporidium ducismarci* Traversa, 2010 and *Cryptosporidium* tortoise genotype III (Apicomplexa: Cryptosporidiidae) in tortoises. *Folia Parasitologica*, *63*. https://doi.org/10.14411/fp.2016.035

Ježková, J., Limpouchová, Z., Prediger, J., Holubová, N., Sak, B., Konečný, R., Květoňová, D., Hlásková, L., Rost, M., McEvoy, J., Rajský, D., Feng, Y., & Kváč, M. (2021). *Cryptosporidium myocastoris* n. sp. (Apicomplexa: Cryptosporidiidae), the Species Adapted to the Nutria (*Myocastor coypus*). *Microorganisms 2021, Vol. 9, Page 813*, *9*(4), 813. https://doi.org/10.3390/MICROORGANISMS9040813

Ježková, J., Prediger, J., Holubová, N., Sak, B., Konečný, R., Feng, Y., Xiao, L., Rost, M., McEvoy, J., & Kváč, M. (2021). *Cryptosporidium ratti* n. sp. (Apicomplexa: Cryptosporidiidae) and genetic diversity of *Cryptosporidium* spp. in brown rats (*Rattus norvegicus*) in the Czech Republic. *Parasitology*, *148*(1), 84–97. https://doi.org/10.1017/S0031182020001833

Jirků, M., Valigurová, A., Koudela, B., Křížek, J., Modrý, D., & Šlapeta, J. (2008). New species of *Cryptosporidium Tyzzer*, 1907 (Apicomplexa) from amphibian host: Morphology, biology and phylogeny. *Folia Parasitologica*, *55*(2), 81–94. https://doi.org/10.14411/fp.2008.011

Jossé, L., Bones, A. J., Purton, T., Michaelis, M., & Tsaousis, A. D. (2019). A Cell Culture Platform for the Cultivation of *Cryptosporidium parvum*. *Current Protocols in Microbiology*, *53*(1), 1–14. https://doi.org/10.1002/cpmc.80

Karanis, P. (2018a). The truth about in vitro culture of *Cryptosporidium* species. In *Parasitology* (Vol. 145, Issue 7, pp. 855–864). Cambridge University Press. https://doi.org/10.1017/S0031182017001937

Karanis, P. (2018b). The truth about *in vitro* culture of *Cryptosporidium* species. *Parasitology*, *145*(7), 855–864. https://doi.org/10.1017/S0031182017001937

Katz, D. E., Heisey-Grove, D., Beach, M., Dicker, R. C., & Matyas, B. T. (2006). Prolonged outbreak of giardiasis with two modes of transmission. *Epidemiology and Infection*, *134*(5), 935–941. https://doi.org/10.1017/S0950268805005832

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond,

A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keegan, A., Daminato, D., Saint, C. P., & Monis, P. T. (2008). Effect of water treatment processes on *Cryptosporidium* infectivity. *Water Research*, *42*(6–7), 1805–1811. https://doi.org/10.1016/j.watres.2007.11.008

Keister, D. B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *77*(4), 487–488. https://doi.org/10.1016/0035-9203(83)90120-7

Khalil, I. A., Troeger, C., Rao, P. C., Blacker, B. F., Brown, A., Brewer, T. G., Colombara, D. V., De Hostos, E. L., Engmann, C., Guerrant, R. L., Haque, R., Houpt, E. R., Kang, G., Korpe, P. S., Kotloff, K. L., Lima, A. A. M., Petri, W. A., Platts-Mills, J. A., Shoultz, D. A., … Mokdad, A. H. (2018). Morbidity, mortality, and long-term consequences associated with diarrhoea from *Cryptosporidium* infection in children younger than 5 years: a meta-analyses study. *The Lancet Global Health*, *6*(7), e758–e768. https://doi.org/10.1016/S2214-109X(18)30283-3

Khan, A., Shaik, J. S., & Grigg, M. E. (2018a). Genomics and molecular epidemiology of *Cryptosporidium* species. *Acta Tropica*, *184*, 1–14. https://doi.org/10.1016/j.actatropica.2017.10.023

Khan, A., Shaik, J. S., & Grigg, M. E. (2018b). Genomics and molecular epidemiology of *Cryptosporidium* species. *Acta Tropica*, *184*, 1–14. https://doi.org/10.1016/J.ACTATROPICA.2017.10.023

Kim, K., Hong, W., & Lee, K. (2001). Disinfection characteristics of waterborne pathogenic protozoa *Giardia lamblia*. *Biotechnology and Bioprocess Engineering*, *6*(2), 95–99. https://doi.org/10.1007/BF02931953

King, B. J., Hoefel, D., Lim, S. P., Robinson, B. S., & Monis, P. T. (2009). Flow cytometric assessment of distinct physiological stages within *Cryptosporidium parvum*

sporozoites post-excystation. *Parasitology*, *136*(9), 953–966. https://doi.org/10.1017/S0031182009006519

King, B. J., Keegan, A. R., Robinson, B. S., & Monis, P. T. (2011). *Cryptosporidium* cell culture infectivity assay design. *Parasitology*, *138*(6), 671–681. https://doi.org/10.1017/S0031182011000217

King, P., Tyler, K. M., & Hunter, P. R. (2019). Anthroponotic transmission of *Cryptosporidium parvum* predominates in countries with poorer sanitation: a systematic review and meta-analysis. *Parasites & Vectors*, *12*(1), 16. https://doi.org/10.1186/s13071-018-3263-0

Kothavade, R. J. (2011). Challenges in understanding the immunopathogenesis of *Cryptosporidium* infections in humans. *European Journal of Clinical Microbiology and Infectious Diseases*, *30*(12), 1461–1472. https://doi.org/10.1007/s10096-011-1246-6

Kulda, J. (1978). Flagellates of the human intestine and of intestines of other species. *Parastic Protozoa*, *2*, 1–138.

Künstler, J. (1882). Sur cinq protozoaires parasites nouveaux. *Comptes Rendus Des Séances de L'académie Des Sciences, Paris*, *95*, 347–349.

Kváč, M., Havrdová, N., Hlásková, L., Daňková, T., Kanděra, J., Ježková, J., Vítovec, J., Sak, B., Ortega, Y., Xiao, L., Modrý, D., Chelladurai, J. R. J. J., Prantlová, V., & McEvoy, J. (2016). *Cryptosporidium proliferans* n. sp. (Apicomplexa: Cryptosporidiidae): Molecular and Biological Evidence of Cryptic Species within Gastric *Cryptosporidium* of Mammals. *PLOS ONE*, *11*(1), e0147090. https://doi.org/10.1371/JOURNAL.PONE.0147090

Kváč, M., Hofmannová, L., Hlásková, L., Květoňová, D., Vítovec, J., McEvoy, J., & Sak, B. (2014). *Cryptosporidium erinacei* n. sp. (Apicomplexa: Cryptosporidiidae) in hedgehogs. *Veterinary Parasitology*, *201*(1–2), 9–17. https://doi.org/10.1016/J.VETPAR.2014.01.014

Kváč, M., Kestřánová, M., Pinková, M., Květoňová, D., Kalinová, J., Wagnerová, P., Kotková, M., Vítovec, J., Ditrich, O., McEvoy, J., Stenger, B., & Sak, B. (2013). *Cryptosporidium scrofarum* n. sp. (Apicomplexa: Cryptosporidiidae) in domestic pigs (*Sus scrofa*). *Veterinary Parasitology*, *191*(3–4), 218–227. https://doi.org/10.1016/J.VETPAR.2012.09.005

Kváč, M., Vlnatá, G., Ježková, J., Horčičková, M., Konečný, R., Hlásková, L., McEvoy, J., & Sak, B. (2018). *Cryptosporidium occultus* sp. n. (Apicomplexa: Cryptosporidiidae) in rats. *European Journal of Protistology*, *63*, 96–104. https://doi.org/10.1016/J.EJOP.2018.02.001

Lake, I. R., Pearce, J., & Savill, M. (2008). The seasonality of human cryptosporidiosis in New Zealand. *Epidemiology and Infection*, *136*(10), 1383–1387. https://doi.org/10.1017/S0950268807009922

Lal, A., Dobbins, T., Bagheri, N., Baker, M. G., French, N. P., & Hales, S. (2016). Cryptosporidiosis Risk in New Zealand Children Under 5 Years Old is Greatest in Areas with High Dairy Cattle Densities. *EcoHealth*, *13*(4), 652–660. https://doi.org/10.1007/s10393-016-1187-8

Lanata, C. F., Fischer-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*(9), e72788. https://doi.org/10.1371/journal.pone.0072788

Lane, S., & Lloyd, D. (2002). Current trends in research into the waterborne parasite *Giardia*. In *Critical Reviews in Microbiology*. https://doi.org/10.1080/1040-840291046713

Laurent, F., Eckmann, L., Savidge, T. C., Morgan, G., Theodos, C., Naciri, M., & Kagnoff, M. F. (1997). *Cryptosporidium parvum* Infection of Human Intestinal Epithelial Cells Induces the Polarized Secretion of C-X-C Chemokines. In *INFECTION AND IMMUNITY* (Vol. 65, Issue 12).

Learmonth, J., Ionas, G., Pita, A., & Cowie, R. (2002). Seasonal shift in *Cryptosporidium parvum* transmission cycles in New Zealand. *Journal of Eukaryotic Microbiology, Supplement*, 34S-35S.

Learmonth, J. J., Ionas, G., Ebbett, K. A., & Kwan, E. S. (2004). Genetic characterization and transmission cycles of *Cryptosporidium* species isolated from humans in New Zealand. *Applied and Environmental Microbiology*, *70*(7), 3973–3978. https://doi.org/10.1128/AEM.70.7.3973-3978.2004

Leggett, H. C., Cornwallis, C. K., & West, S. A. (2012). Mechanisms of Pathogenesis, Infective Dose and Virulence in Human Parasites. *PLoS Pathogens*, *8*(2), e1002512. https://doi.org/10.1371/journal.ppat.1002512

Levine, N. D. (1961). Protozoan Parasites of Domestic Animals and of Man. *Protozoan Parasites of Domestic Animals and of Man.*

Li, B., Wu, H., Li, N., Su, J., Jia, R., Jiang, J., Feng, Y., & Xiao, L. (2017). Preliminary Characterization of MEDLE-2, a Protein Potentially Involved in the Invasion of *Cryptosporidium parvum*. *Frontiers in Microbiology*, *8*, 1647. https://doi.org/10.3389/fmicb.2017.01647

Li, X., Pereira, M. das G. C., Larsen, R., Xiao, C., Phillips, R., Striby, K., McCowan, B., & Atwill, E. R. (2015). *Cryptosporidium rubeyi* n. sp. (Apicomplexa: Cryptosporidiidae) in multiple Spermophilus ground squirrel species. *International Journal for Parasitology: Parasites and Wildlife*, *4*(3), 343–350. https://doi.org/10.1016/J.IJPPAW.2015.08.005

Lindsay, D. S., Upton, S. J., Owens, D. S., Morgan, U. M., Mead, J. R., & Blagburn, B. L. (2000). *Cryptosporidium andersoni* n. sp. (Apicomplexa: Cryptosporiidae) from cattle, *Bos taurus*. *Journal of Eukaryotic Microbiology*, *47*(1), 91–95. https://doi.org/10.1111/j.1550-7408.2000.tb00016.x

Lippuner, C., Ramakrishnan, C., Basso, W. U., Schmid, M. W., Okoniewski, M., Smith, N. C., Hässig, M., Deplazes, P., & Hehl, A. B. (2018a). RNA-Seq analysis during the life cycle

of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites. *International Journal for Parasitology*, *48*(6), 413–422. https://doi.org/10.1016/j.ijpara.2017.10.007

Lippuner, C., Ramakrishnan, C., Basso, W. U., Schmid, M. W., Okoniewski, M., Smith, N. C., Hässig, M., Deplazes, P., & Hehl, A. B. (2018b). RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites. *International Journal for Parasitology*, *48*(6), 413–422. https://doi.org/10.1016/J.IJPARA.2017.10.007

Liu, L., Fang, R., Wei, Z., Wu, J., Li, X., & Li, W. (2020). *Giardia duodenalis* induces apoptosis in intestinal epithelial cells via reactive oxygen species-mediated mitochondrial pathway in vitro. *Pathogens*, *9*(9), 1–14. https://doi.org/10.3390/pathogens9090693

Love, M. S., Beasley, F. C., Jumani, R. S., Wright, T. M., Chatterjee, A. K., Huston, C. D., Schultz, P. G., & McNamara, C. W. (2017). A high-throughput phenotypic screen identifies clofazimine as a potential treatment for cryptosporidiosis. *PLOS Neglected Tropical Diseases*, *11*(2), e0005373. https://doi.org/10.1371/journal.pntd.0005373

Luján, H. D., & Svärd, S. G. (2011). *Giardia*: A Model Organism. In *Springer Link* (Vol. 91). https://doi.org/10.1017/CBO9781107415324.004

Lutz, J. M., & Hardenburg, R. E. (1968). *The commercial storage of fruits, vegetables, and florist and nursery stocks* (Issue 66). US Department of Agriculture.

Lyu, Z., Shao, J., Xue, M., Ye, Q., Chen, B., Qin, Y., & Wen, J. (2018). A new species of *Giardia Künstler*, 1882 (Sarcomastigophora: Hexamitidae) in hamsters. *Parasites and Vectors*, *11*(1), 202. https://doi.org/10.1186/s13071-018-2786-8

Malkov, V. A., Serikawa, K. A., Balantac, N., Watters, J., Geiss, G., Mashadi-Hossein, A., & Fare, T. (2009). Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter™ Assay System. *BMC Research Notes 2009 2:1*, *2*(1), 1–9. https://doi.org/10.1186/1756-0500-2-80

Manjunatha, U. H., Chao, A. T., Leong, F. J., & Diagana, T. T. (2016). Cryptosporidiosis Drug Discovery: Opportunities and Challenges. *ACS Infectious Diseases*, *2*(8), 530–537. https://doi.org/10.1021/acsinfecdis.6b00094

Manjunatha, U. H., Vinayak, S., Zambriski, J. A., Chao, A. T., Sy, T., Noble, C. G., Bonamy, G. M. C., Kondreddi, R. R., Zou, B., Gedeck, P., Brooks, C. F., Herbert, G. T., Sateriale, A., Tandel, J., Noh, S., Lakshminarayana, S. B., Lim, S. H., Goodman, L. B., Bodenreider, C., … Diagana, T. T. (2017). A *Cryptosporidium* PI(4)K inhibitor is a drug candidate for cryptosporidiosis. *Nature*, *546*(7658), 376–380. https://doi.org/10.1038/nature22337

Matos, L. V. S., McEvoy, J., Tzipori, S., Bresciani, K. D. S., & Widmer, G. (2019a). The transcriptome of *Cryptosporidium* oocysts and intracellular stages. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-44289-x

Matos, L. V. S., McEvoy, J., Tzipori, S., Bresciani, K. D. S., & Widmer, G. (2019b). The transcriptome of *Cryptosporidium* oocysts and intracellular stages. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-44289-x

Matos, L. V. S., McEvoy, J., Tzipori, S., Bresciani, K. D. S., & Widmer, G. (2019c). The transcriptome of *Cryptosporidium* oocysts and intracellular stages. *Scientific Reports*, *9*(1), 7856. https://doi.org/10.1038/s41598-019-44289-x

McAuliffe, G., Bissessor, L., Williamson, D., Moore, S., Wilson, J., Dufour, M., Taylor, S., & Upton, A. (2017). Use of the EntericBio Gastro Panel II in a diagnostic microbiology laboratory: challenges and opportunities. *Pathology*, *49*(4), 419–422. https://doi.org/10.1016/J.PATHOL.2017.02.003

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Meloni, B. P., & Thompson, R. (1996). Simplified Methods for Obtaining Purified Oocysts from Mice and for Growing *Cryptosporidium parvum* In vitro. *The Journal of*

Parasitology, *82*(5), 757. https://doi.org/10.2307/3283888

Mfeka, M. S., Martínez-Oyanedel, J., Chen, W., Achilonu, I., Syed, K., & Khoza, T. (2020). Comparative analyses and structural insights of new class glutathione transferases in *Cryptosporidium* species. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-77233-5

Miller, C. N., Jossé, L., Brown, I., Blakeman, B., Povey, J., Yiangou, L., Price, M., Cinatl, J., Xue, W. F., Michaelis, M., & Tsaousis, A. D. (2018). A cell culture platform for *Cryptosporidium* that enables long-term cultivation and new tools for the systematic investigation of its biology. *International Journal for Parasitology*, *48*(3–4), 197–201. https://doi.org/10.1016/j.ijpara.2017.10.001

Mmbaga, B. T., & Houpt, E. R. (2017). *Cryptosporidium* and *Giardia* Infections in Children: A Review. *Pediatric Clinics of North America*, *64*(4), 837–850. https://doi.org/10.1016/j.pcl.2017.03.014

Morada, M., Lee, S., Gunther-Cummins, L., Weiss, L. M., Widmer, G., Tzipori, S., & Yarlett, N. (2016). Continuous culture of *Cryptosporidium parvum* using hollow fiber technology. *International Journal for Parasitology*, *46*(1), 21–29. https://doi.org/10.1016/j.ijpara.2015.07.006

Morgan-Ryan, U. M., Fall, A., Ward, L. A., Hijjawi, N., Sulaiman, I., Fayer, R., Andrew Thompson, R. C., Olson, M., Lal, A., & Xiao, L. (2002). *Cryptosporidium hominis* n. sp. (Apicomplexa: Cryptosporidiidae) from *Homo sapiens*. *Journal of Eukaryotic Microbiology*, *49*(6), 433–440. https://doi.org/10.1111/j.1550-7408.2002.tb00224.x

Morris, A., Robinson, G., Swain, M. T., & Chalmers, R. M. (2019). Direct Sequencing of *Cryptosporidium* in Stool Samples for Public Health. In *Frontiers in Public Health* (Vol. 7, p. 360). Frontiers Media S.A. https://doi.org/10.3389/fpubh.2019.00360

Nader, J. L., Mathers, T. C., Ward, B. J., Pachebat, J. A., Swain, M. T., Robinson, G., Chalmers, R. M., Hunter, P. R., van Oosterhout, C., & Tyler, K. M. (2019). Evolutionary genomics

of anthroponosis in *Cryptosporidium*. *Nature Microbiology*, 1. https://doi.org/10.1038/s41564-019-0377-x

Naghavi, M., Abajobir, A. A., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abera, S. F., Aboyans, V., Adetokunboh, O., Ärnlöv, J., Afshin, A., Agrawal, A., Kiadaliri, A. A., Ahmadi, A., Ahmed, M. B., Aichour, A. N., Aichour, I., Aichour, M. T. E., Aiyar, S., Al-Eyadhy, A., … Murray, C. J. L. (2017). Global, regional, and national age-sex specifc mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1151–1210. https://doi.org/10.1016/S0140-6736(17)32152-9

Ng, J. S. Y., Eastwood, K., Walker, B., Durrheim, D. N., Massey, P. D., Porigneaux, P., Kemp, R., McKinnon, B., Laurie, K., Miller, D., Bramley, E., & Ryan, U. (2012). Evidence of *Cryptosporidium* transmission between cattle and humans in northern New South Wales. *Experimental Parasitology*, *130*(4), 437–441. https://doi.org/10.1016/j.exppara.2012.01.014

Omarova, A., Tussupova, K., Berndtsson, R., Kalishev, M., & Sharapatova, K. (2018). Protozoan parasites in drinking water: A system approach for improved water, sanitation and hygiene in developing countries. In *International Journal of Environmental Research and Public Health* (Vol. 15, Issue 3). https://doi.org/10.3390/ijerph15030495

Ordóñez-Mena, J. M., McCarthy, N. D., & Fanshawe, T. R. (2018). Comparative efficacy of drugs for treating giardiasis: a systematic update of the literature and network meta-analysis of randomized clinical trials. *Journal of Antimicrobial Chemotherapy*, *73*(3), 596–606. https://doi.org/10.1093/jac/dkx430

Ortega-Pierres, M. G., Jex, A. R., Ansell, B. R. ., & Svärd, S. G. (2018). Recent advances in the genomic and molecular biology of *Giardia*. *Acta Tropica*, *184*, 67–72. https://doi.org/10.1016/j.actatropica.2017.09.004

Panda, C., & Mahapatra, R. K. (2018a). Identification of novel therapeutic candidates in *Cryptosporidium parvum*: An in silico approach. *Parasitology*, *145*(14), 1907–1916.

https://doi.org/10.1017/S0031182018000677

Panda, C., & Mahapatra, R. K. (2018b). Identification of novel therapeutic candidates in *Cryptosporidium parvum*: An in silico approach. *Parasitology*, *145*(14), 1907–1916. https://doi.org/10.1017/S0031182018000677

Parsons, M. B., Travis, D., Lonsdorf, E. V., Lipende, I., Roellig, D. M. A., Kamenya, S., Zhang, H., Xiao, L., & Gillespie, T. R. (2015). Epidemiology and Molecular Characterization of *Cryptosporidium* spp. in Humans, Wild Primates, and Domesticated Animals in the Greater Gombe Ecosystem, Tanzania. *PLoS Neglected Tropical Diseases*, *9*(2), e0003529. https://doi.org/10.1371/journal.pntd.0003529

Pavlasek, I., & Ryan, U. (2008). *Cryptosporidium varanii* takes precedence over *C. saurophilum*. *Experimental Parasitology*, *118*(3), 434–437. https://doi.org/10.1016/J.EXPPARA.2007.09.006

Pecková, R., Stuart, P. D., Sak, B., Květoňová, D., Kváč, M., & Foitová, I. (2016). Statistical comparison of excystation methods in *Cryptosporidium parvum* oocysts. *Veterinary Parasitology*, *230*, 1–5. https://doi.org/10.1016/J.VETPAR.2016.10.007

Petry, F., & Harris, J. R. (1999). Ultrastructure, fractionation and biochemical analysis of *Cryptosporidium parvum* sporozoites. *International Journal for Parasitology*, *29*(8), 1249–1260. https://doi.org/10.1016/S0020-7519(99)00080-6

pluriSelect - The Cell Separation Company. (n.d.). Retrieved February 3, 2020, from https://www.pluriselect.com/

Plutzer, J., Lassen, B., Jokelainen, P., Djurković-Djaković, O., Kucsera, I., Dorbek-Kolin, E., Šoba, B., Sréter, T., Imre, K., Omeragić, J., Nikolić, A., Bobić, B., Živičnjak, T., Lučinger, S., Stefanović, L. L., Kučinar, J., Sroka, J., Deksne, G., Keidāne, D., … Karanis, P. (2018). Review of *Cryptosporidium* and *Giardia* in the eastern part of Europe, 2016. In *Eurosurveillance* (Vol. 23, Issue 4). https://doi.org/10.2807/1560-7917.ES.2018.23.4.16-00825

Power, M. L., & Ryan, U. M. (2008). A New Species of *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) from Eastern Grey Kangaroos (*Macropus giganteus*). *Journal of Parasitology*, *94*(5), 1114–1117. https://doi.org/10.1645/GE-1508.1

Pumipuntu, N., & Piratae, S. (2018). Cryptosporidiosis: A zoonotic disease concern. *Veterinary World*, *11*(5), 681–686. https://doi.org/10.14202/vetworld.2018.681-686

Putignani, L., & Menichella, D. (2010). Global distribution, public health and clinical impact of the protozoan pathogen *Cryptosporidium*. *Interdisciplinary Perspectives on Infectious Diseases*, *2010*. https://doi.org/10.1155/2010/753512

Rasmussen, K. R., Larsen, N. C., & Healey, M. C. (1993). Complete development of *Cryptosporidium parvum* in a human endometrial carcinoma cell line. *Infection and Immunity*, *61*(4), 1482–1485.

Read, C. M., Monis, P. T., & Thompson, R. (2004). Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP. *Infection, Genetics and Evolution*, *4*(2), 125–130. https://doi.org/10.1016/J.MEEGID.2004.02.001

Ren, X., Zhao, J., Zhang, L., Ning, C., Jian, F., Wang, R., Lv, C., Wang, Q., Arrowood, M. J., & Xiao, L. (2012). *Cryptosporidium tyzzeri* n. sp. (Apicomplexa: Cryptosporidiidae) in domestic mice (*Mus musculus*). *Experimental Parasitology*, *130*(3), 274–281. https://doi.org/10.1016/J.EXPPARA.2011.07.012

Rivero, F. D., Saura, A., Prucca, C. G., Carranza, P. G., Torri, A., & Lujan, H. D. (2010). Disruption of antigenic variation is crucial for effective parasite vaccine. *Nature Medicine*, *16*(5), 551–557. https://doi.org/10.1038/nm.2141

Robertson, L. J., Clark, C. G., Debenham, J. J., Dubey, J. P., Kváč, M., Li, J., Ponce-Gordo, F., Ryan, U., Schares, G., Su, C., & Tsaousis, A. D. (2019). Are molecular tools clarifying or confusing our understanding of the public health threat from zoonotic enteric protozoa in wildlife? *International Journal for Parasitology: Parasites and Wildlife*. https://doi.org/10.1016/J.IJPPAW.2019.01.010

Robinson, G., Wright, S., Elwin, K., Hadfield, S. J., Katzer, F., Bartley, P. M., Hunter, P. R., Nath, M., Innes, E. A., & Chalmers, R. M. (2010). Re-description of *Cryptosporidium cuniculus* (Apicomplexa: Cryptosporidiidae): Morphology, biology and phylogeny. *International Journal for Parasitology*, *40*(13), 1539–1548. https://doi.org/10.1016/J.IJPARA.2010.05.010

Rose, J. B. (1997). Environmental ecology of *Cryptosporidium* and public health implications. In *Annual Review of Public Health* (Vol. 18, pp. 135–161). Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA . https://doi.org/10.1146/annurev.publhealth.18.1.135

Rousseau, A., La Carbona, S., Dumètre, A., Robertson, L. J., Gargala, G., Escotte-Binet, S., Favennec, L., Villena, I., Gérard, C., & Aubert, D. (2018). Assessing viability and infectivity of foodborne and waterborne stages (cysts/oocysts) of *Giardia duodenalis*, *Cryptosporidium* spp., and *Toxoplasma gondii*: a review of methods. *Parasite (Paris, France)*, *25*, 14. https://doi.org/10.1051/parasite/2018009

Rovid Spickler, A. (2005). *Giardiasis*.

Ryan, U., & Cacciò, S. M. (2013). Zoonotic potential of *Giardia*. In *International Journal for Parasitology* (Vol. 43, Issues 12–13, pp. 943–956). Pergamon. https://doi.org/10.1016/j.ijpara.2013.06.001

Ryan, U., Fayer, R., & Xiao, L. (2014). *Cryptosporidium* species in humans and animals: Current understanding and research needs. *Parasitology*, *141*(13), 1667–1685. https://doi.org/10.1017/S0031182014001085

Ryan, U., Hijjawi, N., Feng, Y., & Xiao, L. (2019). *Giardia*: an under-reported foodborne parasite. *International Journal for Parasitology*, *49*(1), 1–11. https://doi.org/10.1016/J.IJPARA.2018.07.003

Ryan, U. M., Monis, P., Enemark, H. L., Sulaiman, I., Samarasinghe, B., Read, C., Buddle, R., Robertson, I., Zhou, L., Thompson, R. C. A., & Xiao, L. (2004). *Cryptosporidium suis* n. sp. (Apicomplexa: Cryptosporidiidae) in pigs (*Sus scrofa*). *Journal of Parasitology*,

90(4), 769–773. https://doi.org/10.1645/GE-202R1

Ryan, U. M., Power, M., & Xiao, L. (2008). *Cryptosporidium fayeri* n. sp. (Apicomplexa: Cryptosporidiidae) from the red kangaroo (*Macropus rufus*). *Journal of Eukaryotic Microbiology*, *55*(1), 22–26. https://doi.org/10.1111/j.1550-7408.2007.00299.x

Ryan, U. M., Xiao, L., Read, C., Sulaiman, I. M., Monist, P., Lal, A. A., Fayer, R., & Pavlasek, I. (2003). A redescription of *Cryptosporidium galli* Pavlasek, 1999 (Apicomplexa: Cryptosporidiidae) from birds. *Journal of Parasitology*, *89*(4), 809–813. https://doi.org/10.1645/GE-74RI

Ryan, U., Paparini, A., Monis, P., & Hijjawi, N. (2016). It's official – *Cryptosporidium* is a gregarine: What are the implications for the water industry? In *Water Research* (Vol. 105, pp. 305–313). Elsevier Ltd. https://doi.org/10.1016/j.watres.2016.09.013

Ryan, U., Paparini, A., Tong, K., Yang, R., Gibson-Kueh, S., O'Hara, A., Lymbery, A., & Xiao, L. (2015). *Cryptosporidium huwi* n. sp. (Apicomplexa: Eimeriidae) from the guppy (*Poecilia reticulata*). *Experimental Parasitology*, *150*, 31–35. https://doi.org/10.1016/J.EXPPARA.2015.01.009

Saaed, F. M. A., & Ongerth, J. E. (2019). *Giardia* and *Cryptosporidium* in children with diarrhea, Kufra, Libya, a North African migration route city. *International Journal of Hygiene and Environmental Health*, *222*(5), 840–846. https://doi.org/10.1016/j.ijheh.2019.04.006

Samuelson, J., & Robbins, P. (2011). A simple fibril and lectin model for cyst walls of *Entamoeba* and perhaps *Giardia*. In *Trends in Parasitology* (Vol. 27, Issue 1, pp. 17–22). Elsevier Current Trends. https://doi.org/10.1016/j.pt.2010.09.002

Satcher, D., Program Office Stephen Thacker, E. B., Richard Goodman, D. A., Editor, M., Series Scott Wetterhall, M. F., Associate Editor, M., Hewitt, S. M., Managing Editor William Park, M. M., & Ford Morie M Higgins Beverly J Holland Peter M Jenkins, S. L. (1996). *CDC Surveillance Summaries* (Vol. 45).

Sauch, J. F. (1984). Purification of *Giardia muris* cysts by velocity sedimentation. In *Applied and Environmental Microbiology* (Vol. 48, Issue 2). https://doi.org/10.1128/aem.48.2.454-455.1984

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

Savioli, L., Smith, H., & Thompson, A. (2006b). *Giardia* and *Cryptosporidium* join the "Neglected Diseases Initiative." *Trends in Parasitology*, *22*(5), 203–208. https://doi.org/10.1016/j.pt.2006.02.015

Sayed, F. G., Hamza, A. I., Galal, L. A., Sayed, D. M., & Gaber, M. (2016). Virulence of geographically different *Cryptosporidium parvum* isolates in experimental animal model. *Annals of Parasitology*, *62*(3), 221–232. https://doi.org/10.17420/ap6203.56

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. In *Nature Methods* (Vol. 9, Issue 7, pp. 671–675). Nature Publishing Group. https://doi.org/10.1038/nmeth.2089

Seabolt, M. H., Konstantinidis, K. T., & Roellig, D. M. (2021). Hidden Diversity within Common Protozoan Parasites Revealed by a Novel Genomotyping Scheme. *Applied and Environmental Microbiology*, *87*(6), 1–17. https://doi.org/10.1128/aem.02275-20

Serradell, M. C., Saura, A., Rupil, L. L., Gargantini, P. R., Faya, M. I., Furlan, P. J., & Lujan, H. D. (2016). Vaccination of domestic animals with a novel oral vaccine prevents *Giardia* infections, alleviates signs of giardiasis and reduces transmission to humans. *Npj Vaccines 2016 1:1*, *1*(1), 1–11. https://doi.org/10.1038/npjvaccines.2016.18

Shams, S., Khan, S., Khan, A., Khan, I., Ijaz, M., & Ullah, A. (2016). Differential techniques used for detection of *Cryptosporidium* oocysts in stool specimens. *J Parasit Dis Diagnosis Ther*, *1*(1), 1–11.

Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A. A., Ren, B., Amir, A., Schwager, E., Crabtree, J., Ma, S., Abnet, C. C., Knight, R., White, O., & Huttenhower, C. (2017). Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nature Biotechnology 2017 35:11*, *35*(11), 1077–1086. https://doi.org/10.1038/nbt.3981

Slavin, D. (1955). *Cryptosporidium meleagridis* (sp. nov.). *Journal of Comparative Pathology*, *65*(3), 262–266. https://doi.org/10.1016/s0368-1742(55)80025-2

Smith, K. F., Goldberg, M., Rosenthal, S., Carlson, L., Chen, J., Chen, C., & Ramachandran, S. (2014). Global rise in human infectious disease outbreaks. *Journal of The Royal Society Interface*, *11*(101), 20140950. https://doi.org/10.1098/rsif.2014.0950

Snel, S. J., Baker, M. G., Kamalesh, V., French, N., & Learmonth, J. (2009). A tale of two parasites: The comparative epidemiology of cryptosporidiosis and giardiasis. *Epidemiology and Infection*, *137*(11), 1641–1650. https://doi.org/10.1017/S0950268809002465

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

Sonzogni-Desautels, K., Di Lenardo, T. Z., Renteria, A. E., Gascon, M.-A., Geary, T. G., & Ndao, M. (2019). A protocol to count *Cryptosporidium* oocysts by flow cytometry without antibody staining. *PLOS Neglected Tropical Diseases*, *13*(3), e0007259. https://doi.org/10.1371/journal.pntd.0007259

Sponseller, J. K., Griffiths, J. K., & Tzipori, S. (2014). The evolution of respiratory cryptosporidiosis: Evidence for transmission by inhalation. *Clinical Microbiology Reviews*, *27*(3), 575–586. https://doi.org/10.1128/CMR.00115-13

Su, J., Jin, C., Fei, J., Wu, H., Li, N., Guo, Y., Feng, Y., & Xiao, L. (2019). Differential expression of three *Cryptosporidium* species-specific MEDLE proteins. *Frontiers in Microbiology*, *10*(MAY). https://doi.org/10.3389/fmicb.2019.01177

Su, J., Jin, C., Wu, H., Fei, J., Li, N., Guo, Y., Feng, Y., & Xiao, L. (2019). Differential Expression of Three *Cryptosporidium* Species-Specific MEDLE Proteins. *Frontiers in Microbiology*, *10*(MAY), 1177. https://doi.org/10.3389/fmicb.2019.01177

Swaffer, B., Abbott, H., King, B., van der Linden, L., & Monis, P. (2018). Understanding human infectious *Cryptosporidium* risk in drinking water supply catchments. *Water Research*, *138*, 282–292. https://doi.org/10.1016/J.WATRES.2018.03.063

Tandel, J., English, E. D., Sateriale, A., Gullicksrud, J. A., Beiting, D. P., Sullivan, M. C., Pinkston, B., & Striepen, B. (2019). Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium parvum*. *Nature Microbiology*. https://doi.org/10.1038/s41564-019-0539-x

Thompson, & Monis, P. T. (2004). Variation in *Giardia*: Implications for taxonomy and epidemiology. In *Advances in Parasitology* (Vol. 58, pp. 69–137). https://doi.org/10.1016/S0065-308X(04)58002-8

Thompson, R. (2013). Parasite zoonoses and wildlife: One health, spillover and human activity. *International Journal for Parasitology*, *43*(12–13), 1079–1088. https://doi.org/10.1016/J.IJPARA.2013.06.007

Thompson, R., & Ash, A. (2016). Molecular epidemiology of *Giardia* and *Cryptosporidium* infections. *Infection, Genetics and Evolution*, *40*, 315–323. https://doi.org/10.1016/j.meegid.2015.09.028

Thompson, R. C. A., & Monis, P. T. (2011). Taxonomy of *Giardia* Species. In *Giardia* (pp. 3–15). Springer Vienna. https://doi.org/10.1007/978-3-7091-0198-8_1

Thompson, R. C. A., Palmer, C. S., & O'Handley, R. (2008). The public health and clinical significance of *Giardia* and *Cryptosporidium* in domestic animals. In *Veterinary Journal* (Vol. 177, Issue 1, pp. 18–25). W.B. Saunders. https://doi.org/10.1016/j.tvjl.2007.09.022

Thompson, R. C. A., & Polley, L. (2014). Parasitology and One Health. In *International*

Journal for Parasitology: Parasites and Wildlife (Vol. 3, Issue 3, pp. A1–A2). Australian Society for Parasitology. https://doi.org/10.1016/j.ijppaw.2014.09.002

Thompson, R., & Smith, A. (2011). Zoonotic enteric protozoa. *Veterinary Parasitology*, *182*(1), 70–78. https://doi.org/10.1016/J.VETPAR.2011.07.016

Thomson, S., Hamilton, C. A., Hope, J. C., Katzer, F., Mabbott, N. A., Morrison, L. J., & Innes, E. A. (2017). Bovine cryptosporidiosis: impact, host-parasite interaction and control strategies. In *Veterinary research* (Vol. 48, Issue 1, p. 42). BioMed Central. https://doi.org/10.1186/s13567-017-0447-0

Traversa, D. (2010). Evidence for a new species of *Cryptosporidium* infecting tortoises: *Cryptosporidium ducismarci*. *Parasites & Vectors 2010 3:1*, *3*(1), 1–4. https://doi.org/10.1186/1756-3305-3-21

Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner, R. C., Fullman, N., Thompson, R. L., Abajobir, A., Ahmed, M., Alemayohu, M. A., Alvis-Guzman, N., Amare, A. T., Antonio, C. A., Asayesh, H., Avokpaho, E., Awasthi, A., Bacha, U., Barac, A., … Mokdad, A. H. (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(17)30276-1

Tsui, C. K.-M., Miller, R., Uyaguari-Diaz, M., Tang, P., Chauve, C., Hsiao, W., Isaac-Renton, J., & Prystajecky, N. (2018). Beaver Fever: Whole-Genome Characterization of Waterborne Outbreak and Sporadic Isolates To Study the Zoonotic Transmission of Giardiasis. *MSphere*, *3*(2). https://doi.org/10.1128/MSPHERE.00090-18

Tumwine, J. K., Kekitiinwa, A., Nabukeera, N., Akiyoshi, D. E., Rich, S. M., Widmer, G., Feng, X., & Tzipori, S. (2003). *Cryptosporidium parvum* in children with diarrhea in Mulago Hospital, Kampala, Uganda. *American Journal of Tropical Medicine and Hygiene*, *68*(6), 710–715. https://doi.org/10.4269/ajtmh.2003.68.710

Tyzzer, E. E. (1907). A sporozoan found in the peptic glands of the common mouse.

Experimental Biology and Medicine, *5*(1), 12–13. https://doi.org/10.3181/00379727-5-5

Tyzzer, E. E. (1912). *Cryptosporidium parvum* (sp. nov.), a {Coccidium}. *Archiv Für Protistenkunde*, *26*, 394.

Urrutia, A., Duffy, D., Rouilly, V., Posseme, C., Djebali, R., Illanes, G., Libri, V., Albaud, B., Gentien, D., Piasecka, B., Hasan, M., Fontes, M., Quintana-Murci, L., Albert, M. L., Abel, L., Alcover, A., Astrom, K., Bousso, P., Bruhns, P., … Albert, M. L. (2016). Standardized Whole-Blood Transcriptional Profiling Enables the Deconvolution of Complex Induced Immune Responses. *Cell Reports*, *16*(10), 2777–2791. https://doi.org/10.1016/J.CELREP.2016.08.011

Vetterling, J. M., Jervis, H. R., Merril, T. G., & Sprinz, H. (1971). *Cryptosporidium wrairi* sp. n. from the Guinea Pig *Cavia porcellus*, with an Emendation of the Genus. *The Journal of Protozoology*, *18*(2), 243–247. https://doi.org/10.1111/j.1550-7408.1971.tb03315.x

Vinayak, S., Pawlowic, M. C., Sateriale, A., Brooks, C. F., Studstill, C. J., Bar-Peled, Y., Cipriano, M. J., & Striepen, B. (2015). Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature*, *523*(7561), 477–480. https://doi.org/10.1038/nature14651

Vitelli, M., Budman, H., Pritzker, M., & Tamer, M. (2021). Applications of flow cytometry sorting in the pharmaceutical industry: A review. *Biotechnology Progress*, e3146. https://doi.org/10.1002/BTPR.3146

Walderich, B., Müller, L., Bracha, R., Knobloch, J., & Burchard, G. D. (1997). A new method for isolation and differentiation of native *Entamoeba histolytica* and *E. dispar* cysts from fecal samples. *Parasitology Research*, *83*(7), 719–721. https://doi.org/10.1007/s004360050326

Wang, R. jun, Li, J. qiang, Chen, Y. cai, Zhang, L. xian, & Xiao, L. hua. (2018). Widespread occurrence of *Cryptosporidium* infections in patients with HIV/AIDS: Epidemiology,

clinical feature, diagnosis, and therapy. In *Acta Tropica* (Vol. 187, pp. 257–263). Elsevier B.V. https://doi.org/10.1016/j.actatropica.2018.08.018

Warnecke, M., Weir, C., & Vesey, G. (2003). Evaluation of an internal positive control for *Cryptosporidium* and *Giardia* testing in water samples. *Letters in Applied Microbiology*, *37*(3), 244–248. https://doi.org/10.1046/J.1472-765X.2003.01383.X

Widmer, G., Köster, P. C., & Carmena, D. (2020). *Cryptosporidium hominis* infections in non-human animal species: revisiting the concept of host specificity. In *International Journal for Parasitology* (Vol. 50, Issue 4, pp. 253–262). Elsevier Ltd. https://doi.org/10.1016/j.ijpara.2020.01.005

Winiecka-Krusnell, J., & Linder, E. (1998). Cysticidal effect of chlorine dioxide on *Giardia intestinalis* cysts. *Acta Tropica*. https://doi.org/10.1016/S0001-706X(98)00036-9

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. *Applied and Environmental Microbiology*, *74*(16), 5100–5105. https://doi.org/10.1128/AEM.00232-08

Wiser, M. F. (2010). *Protozoa and human disease*. Garland Science.

Wuhib, T., Silva, T. M. J., Newman, R. D., Garcia, L. S., Pereira, M. L. D., Chaves, C. S., Wahlquist, S. P., Bryan, R. T., Sousa, A. D. Q., De Queiroz, T. R. B. S., & Sears, C. L. (1994). Cryptosporidial and microsporidial infections in human immunodeficiency virus-infected patients in northeastern brazil. *Journal of Infectious Diseases*, *170*(2), 494–497. https://doi.org/10.1093/infdis/170.2.494

Xiao, L. (2010). Molecular epidemiology of cryptosporidiosis: An update. *Experimental Parasitology*, *124*(1), 80–89. https://doi.org/10.1016/j.exppara.2009.03.018

Xiao, L., & Feng, Y. (2017). Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. *Food and Waterborne Parasitology*, *8–9*(August), 14–32. https://doi.org/10.1016/j.fawpar.2017.09.002

Xu, R., Guo, Y., Li, N., Zhang, Q., Wu, H., Ryan, U., Feng, Y., & Xiao, L. (2019). Characterization of INS-15, a metalloprotease potentially involved in the invasion of *Cryptosporidium parvum*. *Microorganisms*, *7*(10). https://doi.org/10.3390/microorganisms7100452

Yang, Y., Xue, X., Yang, Y., Chen, X., & Du, A. (2016). Efficacy of a potential DNA vaccine encoding *Cryptosporidium baileyi* rhomboid protein against homologous challenge in chickens. *Veterinary Parasitology*, *225*, 5–11. https://doi.org/10.1016/j.vetpar.2016.05.024

Yu, L., Bhayana, S., Jacob, N. K., & Fadda, P. (2019). Comparative studies of two generations of NanoString nCounter System. *PLOS ONE*, *14*(11), e0225505. https://doi.org/10.1371/JOURNAL.PONE.0225505

Zahedi, A., Durmic, Z., Gofton, A. W., Kueh, S., Austen, J., Lawson, M., Callahan, L., Jardine, J., & Ryan, U. (2017). *Cryptosporidium homai* n. sp. (Apicomplexa: Cryptosporidiiae) from the guinea pig (*Cavia porcellus*). *Veterinary Parasitology*, *245*, 92–101. https://doi.org/10.1016/J.VETPAR.2017.08.014

Zahedi, A., Gofton, A. W., Jian, F., Paparini, A., Oskam, C., Ball, A., Robertson, I., & Ryan, U. (2017). Next Generation Sequencing uncovers within-host differences in the genetic diversity of *Cryptosporidium* gp60 subtypes. *International Journal for Parasitology*, *47*(10–11), 601–607. https://doi.org/10.1016/j.ijpara.2017.03.003

Zhang, S., Wang, Y., Wu, H., Li, N., Jiang, J., Guo, Y., Feng, Y., & Xiao, L. (2019). Characterization of a Species-Specific Insulinase-Like Protease in *Cryptosporidium parvum*. *Frontiers in Microbiology*, *10*(MAR), 354. https://doi.org/10.3389/fmicb.2019.00354

9 Appendices

9.1 Appendix A

Supplementary Material to Chapter 2

High-Yield Purification of *Giardia intestinalis* Cysts from Fecal Samples

Paul Ogbuigwe,[1,2] Anthony B. Pita,[1] Matthew A. Knox,[1] Niluka Velathanthiri,[1] and David T. S. Hayman[1]

[1]Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand
[2]Corresponding author: *P.Ogbuigwe@massey.ac.nz*

Giardia is an enteric protozoan parasite that causes gastroenteritis in all classes of vertebrates. It is ranked among the leading causes of death in children under 5 years of age. Giardiasis affects approximately 280 million people worldwide annually, a situation exacerbated by the low availability of effective treatments and the lack of a vaccine. In addition, the parasite is difficult to manipulate in in vitro environments, which hampers the development of effective disease management strategies. This article highlights the development of a method for the purification of viable *Giardia* cysts from fecal samples, verified by a trypan blue dye exclusion test. This protocol produces a 10-fold increase in yield over current methods. By combining sucrose flotation with gated filtration, the protocol significantly reduces the amount of debris in the purified cysts suspension. Cyst viability is verified by a trypan blue dye exclusion test. The ability to purify large quantities of *Giardia* from fecal samples could advance the development of effective treatments to target this worldwide prevalent parasite. © 2020 Wiley Periodicals LLC.

Basic Protocol: Purification of *Giardia* cysts from fecal samples
Support Protocol: Cyst viability test

Keywords: filtration • flotation • *Giardia* • purification

INTRODUCTION

Giardia intestinalis is a protozoan parasite that has the ability to infect the epithelial cells of the gastrointestinal tract of all classes of vertebrates. *Giardia* is one of the leading causes of diarrhea worldwide and can be lethal to immunocompromised individuals and children under 5 years of age (Dib, Lu, & Wen, 2008; Saaed & Ongerth, 2019). The main mode of transmission of this parasite in humans and other animals is through contaminated food and water (Efstratiou, Ongerth, & Karanis, 2017; Ryan, Hijjawi, Feng, & Xiao, 2019). Transmission among people is aided by the ability of the environmental stage of *Giardia*, the cyst, to resist chlorine and bleach, two of the most common chemicals used in water purification (Jarroll & Hoff, 1988; Kim, Hong, & Lee, 2001; Winiecka-Krusnell & Linder, 1998). Current methods of detection of *Giardia* include microscopy, immunoassays, and molecular assays (Adeyemo, Singh, Reddy, & Stenström, 2018). However, the inability to purify significant numbers of cysts from feces

CURRENT PROTOCOLS
A Wiley Brand

has hindered the progress of in vitro studies looking at the molecular mechanisms of this parasite. Here, we detail a new method for the purification of *Giardia* cysts from feces, which builds on past methods (Afshin, Jafar, Esmaeel, & Reza, 2011) but gives a 10-fold increase while reducing the amount of debris in the sample. We delineate protocols for purifying *Giardia* cysts from fecal samples using sucrose flotation to separate cysts from fecal matter, followed by gated filtration to purify the cysts from any remaining debris (Basic Protocol), and a dye-exclusion test to verify the viability of the purified cysts (Support Protocol).

Human fecal samples are used in this experiment and *Giardia* spp. that infect people are pathogens. The Guidelines for biosafety and practice can be found in the latest edition of *Biosafety in Microbiological and Biomedical Laboratories* (Current edition: 5th Edition) and refer to section VIII-C: Parasitic Agents. See *UNIT* 1A.*1* and other pertinent resources (*APPENDIX* 1B*)* for more information. Additionally, see Burnett et al., 2009.

STRATEGIC PLANNING

The entire protocol should be carried out in a fume hood and all necessary equipment, such as weighing scales and vortex mixers, should be transferred to the fume hood prior to the start of the experiment. Sucrose solutions should be made at least a day before the experiment and stored at 4°C.

BASIC PROTOCOL

PURIFICATION OF *GIARDIA* CYSTS FROM FECAL SAMPLES

Giardia cysts can be purified from fecal samples acquired from patients or animals that have been diagnosed with giardiasis. The fecal samples should be stored at 4°C prior to purification of cysts. This protocol details the steps for purifying *Giardia* cysts from fecal samples using sucrose flotation to separate the cysts from debris, and gated filtration to produce a clean, high-yield cyst suspension.

CAUTION: Giardia is a Biosafety Level 2 (BSL-2) pathogen. Please follow all appropriate guidelines and regulations for use and handling of pathogenic microorganisms. See Burnett et al., 2009 for more information. Human fecal samples are used in this experiment, so please follow all appropriate guidelines and regulations, including human ethics approvals, for the use and handling of human-derived materials. See Burnett et al., 2009 for more information.

Materials

Feces samples
0.1 M phosphate-buffered saline (PBS), pH 7.3
Tween 80 (Sigma Aldrich, cat. no. 9005-65-6)
Distilled H_2O
1.5 M and 0.85 M Sucrose solutions (see recipe)
0.1 mm diameter zirconia/silica beads (BioSpec, cat. no. 11079101Z)
Antibiotic solution (see recipe)
70% ethanol
Penicillin-Streptomycin (10,000 U/ml) (ThermoFisher, cat. no. 15140163)
Amphotericin B (ThermoFisher, cat. no. 15290018)

Wooden applicator (Sigma Aldrich, cat. no. Z406430) *or* transfer pipette
50-ml centrifuge tubes (Sigma Aldrich, cat. no. CLS430828)
Vortex mixer
40-μm cell strainers (PluriSelect, cat. no. 43-50040-51)

171

Benchtop centrifuge (refrigerated)
5-ml pipette
Autoclavable receptacle
20-μm PluriSelect cell strainers (PluriSelect, cat. no. 43-50020-03)
50-ml syringe
5-μm PluriSelect cell strainers (PluriSelect, cat. no. 43-50005-13)
1.5-ml microcentrifuge tubes
PluriSelect connector ring (Pluriselect, cat. no. 41-50000-03)
Hemocytometer
Coverslips
Inverted microscope
4°C refrigerator

Removal of large debris

1. In a fume hood, weigh out 1 g of feces using a wooden applicator or transfer pipette (depending on the solidity of the sample).

2. Dissolve the feces in 20 ml refrigerated PBS in a 50-ml centrifuge tube and mix thoroughly using a vortex mixer at maximum speed.

3. Place a 40-μm cell strainer on top of a new opened 50-ml centrifuge tube and pass the fecal suspension through it by inverting the tube. Collect the filtrate and discard the cell strainer.

    This step removes larger particles of debris.

Washing and removal of lipids

4. Centrifuge the filtrate for 5 min at 500 × *g*, 4°C, with the brakes off (to reduce disturbance of the pellet).

5. Using a 5-ml pipette, carefully discard the supernatant in an autoclavable receptacle, leaving a loose pellet in 5 ml at the bottom.

    All waste should be discarded into this receptacle and autoclaved once the experiment is completed.

6. Add 15 ml of 0.5% Tween 80 to make a final volume of 20 ml and resuspend the pellet by vortexing at maximum speed.

7. Centrifuge the suspension for 5 min at 500 × *g*, 4°C.

8. Repeat steps 5-7 two more times, for a total of three washes in 0.5% Tween 80.

    These steps dissolve any free lipids in the samples.

9. Discard the supernatant, leaving 5 ml at the bottom.

10. Add 15 ml of dH_2O for a total volume of 20 ml and resuspend pellet by vortexing at maximum speed.

11. Centrifuge the suspension for 5 min at 500 × *g*, 4°C.

12. Repeat steps 9–11 two more times, for a total of three washes in dH_2O.

    These steps remove any remaining Tween 80 from the suspension.

Perform 1.5 M sucrose flotation

13. Discard the supernatant leaving 5 ml at the bottom, and then add 15 ml of refrigerated dH_2O and resuspend the pellet.

14. Using a 5-ml pipette, carefully layer an equal volume (20 ml) of refrigerated 1.5 M sucrose solution under the fecal suspension, taking care to minimize disturbance at the fluid interface.

    This is the first flotation step to separate the components of the fecal suspension

172

15. Centrifuge the flotation for 10 min at 1300 × g, 4°C, with the brake off.

   After centrifugation, a cloudy layer should be visible between the two liquids, this is referred to as the mediated phase. The mediated phase will contain the cysts in it, and the heavier debris will settle to the bottom. Carefully collect 15 ml of the mediated phase and transfer into a new 50-ml centrifuge tube, taking care not to disturb the interface between the dH$_2$O and the sucrose excessively.

16. Add an equal volume (15 ml) of refrigerated dH$_2$O to the collected mediated phase and mix thoroughly by vortexing at maximum speed.

17. Centrifuge the suspension for 5 min at 500 × g, 4°C.

18. Discard the supernatant leaving a pellet in 5 ml at the bottom.

19. Add 15 ml dH$_2$O to make a total volume of 20 ml and resuspend the pellet by vortexing at maximum speed.

20. Centrifuge the suspension for 5 min at 500 × g, 4°C.

21. Repeat steps 20–22 one more time.

   These steps wash the sucrose from the suspension.

Perform 0.85 M sucrose flotation

22. Discard the supernatant leaving 5 ml at the bottom, then add 15 ml of dH$_2$O and resuspend the pellet.

23. Using a 5-ml pipette, carefully layer an equal volume (20 ml) of refrigerated 0.85 M sucrose solution under the fecal suspension, taking care to minimize disturbance at the fluid interface.

24. Centrifuge the flotation for 5 min at 1600 × g, 4°C, with the brake off.

   This is the second flotation step.

   After centrifugation the cysts will settle to the bottom and the lighter debris will be in suspension within the mediated phase.

25. Using a 5-ml pipette, carefully discard most of the contents of the flotation, leaving 5 ml at the bottom.

26. Add 15 ml dH$_2$O to make a total volume of 20 ml and resuspend the pellet by vortexing at maximum speed.

27. Centrifuge the suspension for 5 min at 500 × g, 4°C.

28. Discard the supernatant leaving 5 ml at the bottom.

29. Add 15 ml dH$_2$O to make a total volume of 20 ml and resuspend pellet by vortex.

30. Centrifuge the suspension for 5 min at 500 × g, 4°C.

31. Repeat steps 29–31 once more.

   These steps wash the sucrose from the suspension.

Filtration of remaining large particles

32. Discard the supernatant leaving 5 ml at the bottom, then add 5 ml of PBS (making a total of 10 ml) and resuspend the pellet.

33. Add 0.2 g of 0.1 mm zirconia/silica beads to the suspension and vortex thoroughly for 1 min.

173
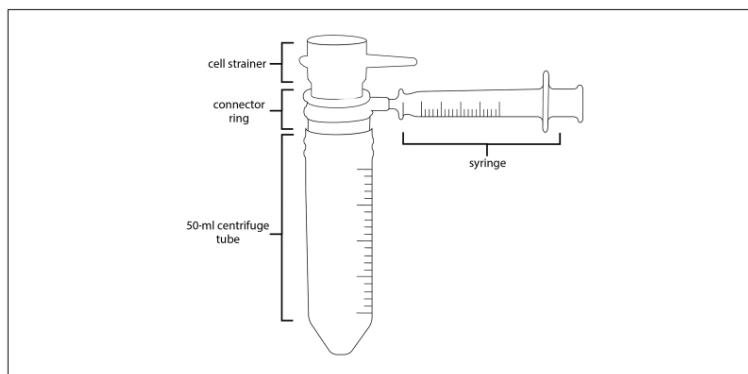
Figure 1 Schematic of filtration setup. The cell strainer is attached to a connector ring, which is, in turn, attached to the top of a 50-ml centrifuge tube. A syringe attached to the connector ring is used to create a region of low pressure that assists in and speeds up the filtration process through the cell strainer.

> *This step separates any coagulated particles in the suspension making subsequent filtration more efficient.*

34. Immediately after vortexing, pass the suspension through a 20-μm cell strainer with a connector ring attached to a new 50-ml centrifuge tube. Attach a 50-ml syringe to the connector ring and gradually pull the piston back to create an area of low pressure that assists filtration (Fig. 1).

> *Larger particles and the zirconia/silica beads will be collected on the cell strainer.*

35. Collect the filtrate and discard the cell strainer.

Filtration of smaller particles

36. Vortex the filtrate thoroughly at maximum speed, and then filter through a 5-μm cell strainer with a connector ring attached in a new 50-ml centrifuge tube, as was done in step 35.

> *The cysts will be collected on the cell strainer and smaller particles will pass through.*

37. Carefully place the 5-μm cell strainer turned upside down over a new 50-ml centrifuge tube (Fig. 2).

38. Wash the cysts off the cell strainer using 15–20 ml refrigerated dH_2O. To do this, attach a connector ring to an empty 50-ml centrifuge tube and place this over the cell strainer. Connect a fully extended syringe to the connector ring and push air in to create an area of high pressure that assists washing off the cysts when the cell strainer cup is filled with dH_2O (Fig. 2).

> *When adding more dH_2O to the system simply detach the connector ring with the syringe and empty 50-ml tube from the cell strainer, add more water then reattach them to the cell strainer. The water should be added with a 5-ml pipette to prevent spillage. Without the increased pressure the collected cysts do not readily wash of the cell strainer. This is a very important step.*

39. Collect the 50-ml centrifuge tube with the washed cysts and centrifuge for 8 min at $2000 \times g$, 4°C.

40. Discard the supernatant leaving 1 ml and then resuspend the pellet. Using a transfer pipette, move the solution into a new 1.5-ml microcentrifuge tube.

174
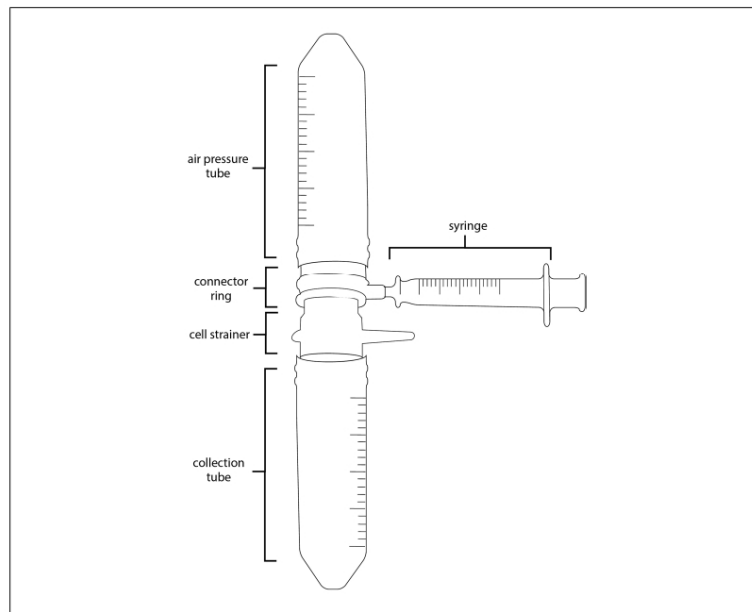
Figure 2 High pressure-assisted washing. After filtration, the 5-μm cell strainer is flipped over onto a fresh 50-ml centrifuge tube to wash off the collected cysts. An empty centrifuge tube with a connector ring and syringe attached to it is used to create an area of high pressure to support the washing off of the cysts with H_2O.

41. Spin down the cyst solution for 5 min at 500 × *g*, 4°C. Discard 900 μl of the supernatant into the waste receptacle.

42. Wash once in refrigerated PBS by adding 900 μl PBS to the tube, centrifuging for 5 min at 500 × *g*, 4°C, and then discarding 900 μl of the supernatant.

43. Add 900 μl refrigerated PBS to the sample and resuspend by vortexing at maximum speed.

44. Add 10 μl of the antibiotic solution to the cyst suspension.

Counting and storage
45. Clean a hemocytometer and coverslip with 70% ethanol, then dry them and fix the coverslip in position

46. Take 10 μl of the cyst suspension and add to the hemocytometer.

47. Place the hemocytometer on an inverted microscope. Using the 20× objective count all the cysts in each of the 4 × 4 corner squares highlighted in Figure 3.

48. Calculate the number of cysts/ml in your sample using this formula: $\frac{\text{Total}}{4} \times 10{,}000$

49. Store cysts suspensions at 4°C.

   The cysts remain viable for up to 3 months (Bingham, Jarroll, Meyer, & Radulescu, 1979).

175
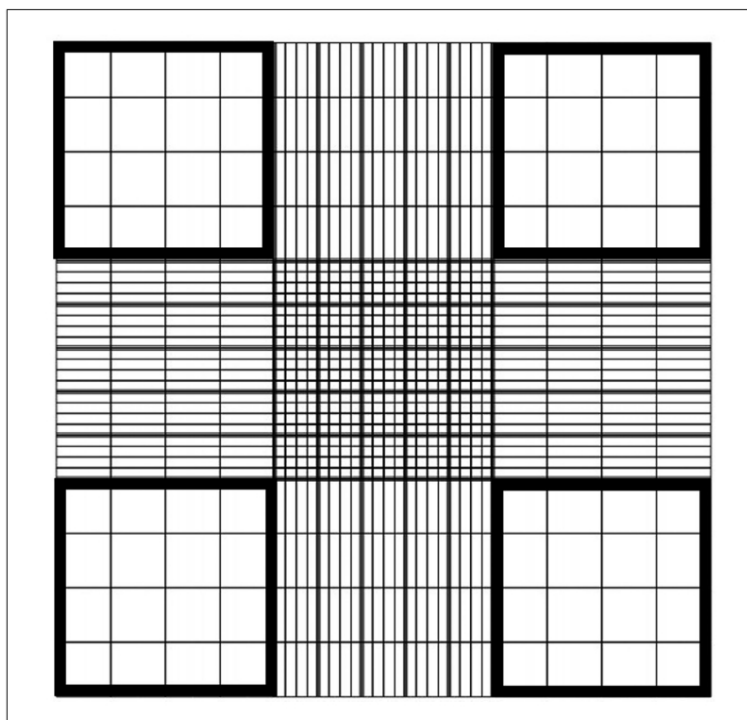
Figure 3 Diagram of the viewing pane of a typical hemocytometer. The highlighted areas show the boxes where cysts should be enumerated. The number of cysts in all four boxes should be counted and totaled.

CYST VIABILITY TEST

Giardia cysts come in two forms: type 1 and type 2 cysts (Gillin, Boucher, Rossi, & Reiner, 1989). Type 1 cysts are considered to be viable and better for excystation, while type 2 cysts are considered non-viable. However, both type 1 and type 2 cysts are suitable for studies involving DNA and RNA extraction. Here we detail the procedure to carry out a dye-exclusion test to determine the viability of purified *Giardia* cysts.

Additional Materials *(also see the Basic Protocol)*

> Purified cyst suspension (see the Basic Protocol)
> Trypan Blue Solution, 0.4% (ThermoFisher, cat. no. 15250061)
> 0.1 M Phosphate-buffered saline, pH 7.3 (PBS)
> Hemocytometer

Dilution of cyst suspension in viability dye

1. In a 1.5-ml microcentrifuge tube mix the purified cyst suspension 1:1 with 0.4% trypan blue dye by vortexing thoroughly at maximum speed.

   If the number of purified cysts is high, mix 200 μl of the cyst suspension with 200 μl trypan blue. If numbers are low, mix 500 μl of the cyst suspension with 500 μl trypan blue.

2. Clean a hemocytometer and coverslip with 70% ethanol, then dry them and fix the coverslip in position.
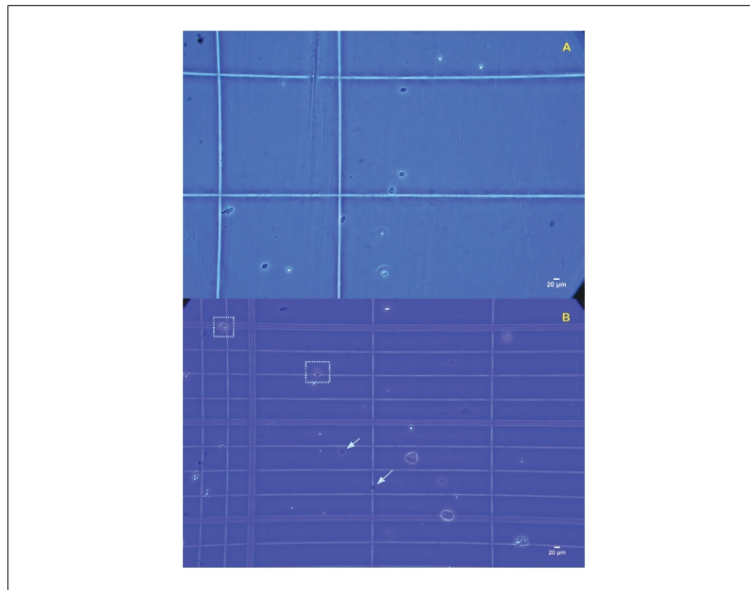
Figure 4 Brightfield microscopy images of purified *Giardia intestinalis* cysts. (**A**) 20× magnification of purified cysts showing little debris. (**B**) trypan blue staining of purified cysts to show presence of viable cysts. Type 1 cysts (highlighted by dashed boxes) appear as bright ovals and the cell body is uniformly distributed within the cyst wall (the cyst wall is clearly visible), these are trypan-blue negative (white). In type 2 cysts (highlighted by arrows) the main body of the cell appears to be fully or partially detached from the cyst wall, these are trypan-blue positive (blue).

3. Using a pipette, take 10 µl of the cyst and dye mixture and add to the hemocytometer.

4. Place the hemocytometer on an inverted microscope. Using the 20× objective count the type 1 and type 2 cysts in each of the 4 × 4 corner squares highlighted in Figure 3.

   Type 1 cysts (highlighted by dashed boxes in Fig. 4B) appear as bright ovals and the cell body is uniformly distributed within the cyst wall (the cyst wall is clearly visible), these are trypan-blue negative (white). In type 2 cysts (highlighted by arrows in Fig. 4B) the main body of the cell appears to be fully or partially detached from the cyst wall, these are trypan-blue positive (blue) (Luján & Svärd, 2011).

5. Calculate the number of each type of cyst using this formula: $\frac{\text{Total}}{4} \times 2 \times 10,000$.

   Type 1 cysts are considered to be viable and are ideal for excystation. The ratio of Type 1 cysts to Type 2 cysts will inform you of the proportion of cysts in your sample that are viable.

REAGENTS AND SOLUTIONS

Antibiotic solution

   For 1 ml of solution:

   In a 1.5-ml microcentrifuge tube, make up 100 U/ml penicillin, 100 µg/ml streptomycin (ThermoFisher, cat. no. 15140163), and 250 ng/ml amphotericin B (ThermoFisher, cat. no. 15290018) in phosphate-buffered saline (PBS). Store up to 6 months at −20°C.

Sucrose solution, 0.85 M

For 500 ml of solution:

Dissolve 145.5 g of molecular grade sucrose (Sigma Aldrich, cat. no. 57-50-1) in dH$_2$O up to 500 ml.
Store up to 1 month at 4°C.

Sucrose solution, 1.5 M

For 500 ml of solution:

Dissolve 256.7 g of molecular grade sucrose (Sigma Aldrich, cat. no. 57-50-1) in dH$_2$O up to 500 ml.
Store up to 1 month at 4°C.

COMMENTARY

Background Information

Giardia begins its life cycle as a cyst. The cyst is the environmental/transmissible stage of this parasite and it possesses a hardy wall made up of a mesh of cyst wall proteins complexed to a singular sugar polymer of (β1–3)-linked *N*-acetylgalactosamine (GalNAc) (Samuelson & Robbins, 2011), which makes it tremendously stable in cool and moist conditions, resistant to most solvents, and able to survive in the environment for several months in cold conditions (∼10°C) (Bingham et al., 1979; Jarroll & Hoff, 1988; Rovid Spickler, 2005). The infective dose required to initiate infection in humans is approximately 10–-100 cysts (Leggett, Cornwallis, & West, 2012). Acute infections normally build up over 3 weeks, with a peak at 8 days post-infection (Cernikova, Faso, & Hehl, 2018). According to current understanding, *Giardia* cysts come in two varieties: type 1 cysts which, when viewed with light microscopy, appear as bright ovals with the cell body uniformly distributed within a clearly visible cell wall; type 2 cysts, which are ovoid in shape but appear darker when using a light microscope, and the cell body appears fully or partially detached from the clearly visible cell wall. Type 1 cysts are considered to be more viable and better for in-vitro excystation than type 2 cysts (Gillin et al., 1989; Luján & Svärd, 2011). Current methods for the purification of *Giardia* cysts from feces include sucrose- and Percoll-based flotation (Afshin et al., 2011; Alvarado & Wasserman, 2006; Sauch, 1984; Walderich, Müller, Bracha, Knobloch, & Burchard, 1997). Here, we modified the two-phase sucrose flotation method described by Afshin et al. (2011), and employed the use of cell strainers and pressure-assisted filtration to obtain a 10-fold increase in the yield of cysts. In addition, the application of zirconia/silica beads func-

tioned to reduce the amount of contaminants in the sample by separating any debris that coagulated in the suspension, thereby making the subsequent filtration steps more efficient (Fig. 4A). On average, *Giardia* cysts have a length of 11-14 μm and a width of 7-10 μm (Luján & Svärd, 2011). By using the cell strainers to filter matter larger and smaller than the dimensions of the cysts, we were able to significantly reduce the amount of debris and contaminants in the sample. The standard method for the identification of *Giardia* cysts in feces has been the ether sedimentation technique, which relied on the ability of the diagnostician to differentiate cysts from fecal matter and other microbes present in the sample (Hooshyar, Rostamkhani, Arbabi, & Delavari, 2019). This method requires two or more examinations of feces from the same individual for reliable diagnosis. As a result of this, and the fact that cyst shedding varies during the course of an infection, it has been difficult to estimate the average load of cysts in infected individuals. The procedure outlined in this protocol could provide a method for acquiring accurate estimates of the cyst load in infected individuals. Furthermore, our method does not require any stains or fluorescent microscopes to identify *Giardia* cysts. The lack of debris in the sample makes it reliable, easy and efficient to identify cysts based on their morphological characteristics with little training or prior experience needed. The purification of large numbers of cysts from fecal samples could give us a better understanding of host-parasite interactions and advance the manipulation and understanding of the molecular biology of this parasite.

Critical Parameters

The fecal samples required for this experiment should be verified as positive for *Giardia*

178

infection prior to application of the protocol. Possible methods for the diagnosis of fecal samples for *Giardia* include light microscopy of stool samples, fluorescent antibody tests, ELISA, and the amplification of DNA fragments by PCR.

Temperature can affect the specific gravity of liquids, so it is important that all the sucrose solutions are stored at 4°C until the flotation steps to preserve their integrity. Purification will not produce a high yield if the mediated phase is disturbed during the 1.5 M sucrose flotation, this is the most delicate step of the entire protocol. Take care to only extract the cloudy layer at the interphase and some of the layer of dH_2O that is close to the layer, but do not puncture down into the sucrose layer as this will trap the cysts in the sucrose. Improper collection of the mediated phase during this step will result in low yields.

During the filtration of small particles in the Basic Protocol it is especially important that an area of high pressure is created when washing the cysts off the 5-μm cell strainer. For this reason, it is important that a syringe of 50 ml volume or larger is used, and that the plunger is full extended before attachment to the connector ring. This will increase the amount and force of air that can be forced into the system.

When using the cysts, consider that their infectivity reduces with time; however, they can maintain infectivity for up to 3 months after purification if stored at 4°C.

Troubleshooting

In a fecal sample with a high load of cysts, more cysts can be purified if the 1.5 M sucrose flotation is collected after extraction of the mediated phase, diluted with dH_2O, washed three times, and put through the rest of the procedure. The resulting cyst suspension might have more contamination in it as a result, but if you have a limited supply of *Giardia*-positive fecal samples this is a great way to maximize output from particularly cyst-heavy fecal samples.

The method outlined in this protocol is efficient at removing most of the contaminating matter from the cyst suspensions. If there is a significant amount of debris in the final cyst suspension, transfer it back into a clean 50-ml centrifuge tube and repeat the process from the 1.5 M sucrose flotation onwards. This should effectively clean the sample. In addition, the addition of antibiotics and antimycotics to the cyst suspension at the final steps should kill any remaining contaminating microbes in the solution. A possible source of contamination that should be considered is the sucrose solutions. If they are stored for more than 1 month, they are vulnerable to fungal contaminants, and these can be passed on to the cyst suspension. To forestall this, prepare the sucrose solutions close to the day you intend to make use of them.

If you have a low yield, it could be because of improper collection of the mediated phase during the 1.5 M sucrose flotation, or it could be because there was initially a low number of cysts in the fecal sample used for the experiment.

Understanding Results

The original method achieves a recovery rate of 1.5×10^4 cysts from 2 g of feces (Afshin et al., 2011), and a previous one, by Walderich et al. (1997), achieves a return rate of approximately 5×10^4 cysts from 2 g of feces. With our method, we were able to reliably recover between $1-1.5 \times 10^5$ cysts from 1 g of feces, and the dye exclusion test shows that approximately 50% of the cysts in the suspension were viable (Fig. 4B).

NOTE: Factors such as the source of the fecal samples, the conditions in which they were transported and stored, and the length of time they have been kept in storage can affect the viability of the cysts recovered. Viability is important if you plan on running subsequent culturing or excystation assays but do not affect experiments involving DNA or RNA purification.

Time Considerations

Preparation of the sucrose solutions can take 1 hr or more due to the time it takes for the large mass of sucrose to dissolve, so it is advised that the solutions are prepared prior to the day of experimentation. This time can be reduced by using a heated magnetic stirrer. The entire procedure can be carried out in 1 day and can take up to 6 hr when four samples are purified simultaneously by one person.

Author Contributions

Paul Ogbuigwe: Conceptualization; investigation; methodology; writing-original draft. **Anthony B. Pita:** Funding

acquisition; methodology; supervision; validation. **Matthew A. Knox:** Methodology; supervision; validation; writing-review & editing.**Niluka Velathanthiri:** Validation. **David T. S. Hayman:** Project administration; resources; supervision; writing-review & editing.

Literature Cited

Adeyemo, F. E., Singh, G., Reddy, P., & Stenström, T. A. (2018). Methods for the detection of Cryptosporidium and Giardia: From microscopy to nucleic acid based tools in clinical and environmental regimes. *Acta Tropica*, *184*, 15–28. doi: 10.1016/J.ACTATROPICA.2018.01.011.

Afshin, B., Jafar, M., Esmaeel, F., & Reza, G. (2011). Introducing a simple and economical method to purify Giardia lamblia cysts. *African Journal of Biotechnology*, *10*(42), 8498–8501. doi: 10.5897/AJB11.391.

Alvarado, M. E., & Wasserman, M. (2006). Quick and efficient purification of Giardia intestinalis cysts from fecal samples. *Parasitology Research*, *99*(3), 300–302. doi: 10.1007/s00436-006-0143-x.

Bingham, A. K., Jarroll, E. L., Meyer, E. A., & Radulescu, S. (1979). Giardia sp.: Physical factors of excystation in vitro, and excystation vs eosin exclusion as determinants of viability. *Experimental Parasitology*, *47*(2), 284–291. doi: 10.1016/0014-4894(79)90080-8.

Burnett, L. C., Lunn, G., & Coico, R. (2009). Biosafety: Guidelines for working with pathogenic and infectious microorganisms. *Current Protocols in Microbiology*, *13*(1), 1A.1.1–1A.1.14. doi: 10.1002/9780471729259.mc01a01s13.

Cernikova, L., Faso, C., & Hehl, A. B. (2018). Five facts about Giardia lamblia. *PLoS Pathogens*, *14*(9), e1007250. doi: 10.1371/journal.ppat.1007250.

Dib, H. H., Lu, S. Q., & Wen, S. F. (2008). Prevalence of Giardia lamblia with or without diarrhea in South East, South East Asia and the Far East. *Parasitology Research*, *103*, 239–251. doi: 10.1007/s00436-008-0968-6.

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

Gillin, F. D., Boucher, S. E., Rossi, S. S., & Reiner, D. S. (1989). Giardia lamblia: The roles of bile, lactic acid, and pH in the completion of the life cycle in vitro. *Experimental Parasitology*, *69*(2), 164–174. doi: 10.1016/0014-4894(89)90185-9.

Hooshyar, H., Rostamkhani, P., Arbabi, M., & Delavari, M. (2019). Giardia lamblia infection: Review of current diagnostic strategies. *Gastroenterology and Hepatology from Bed to Bench*, *12*, 3–12. doi: 10.22037/ghfbb.v0i0.1414.

Jarroll, E. L., & Hoff, J. C. (1988). Effect of disinfectants on giardia cysts. *Critical Reviews in Environmental Control*, *18*(1), 1–28. doi: 10.1080/10643388809388341.

Kim, K., Hong, W., & Lee, K. (2001). Disinfection characteristics of waterborne pathogenic protozoa Giardia lamblia. *Biotechnology and Bioprocess Engineering*, *6*(2), 95–99. doi: 10.1007/BF02931953.

Leggett, H. C., Cornwallis, C. K., & West, S. A. (2012). Mechanisms of pathogenesis, infective dose and virulence in human parasites. *PLoS Pathogens*, *8*(2), e1002512. doi: 10.1371/journal.ppat.1002512.

Luján, H. D., & Svärd, S. G. (2011). Giardia: A Model Organism. In *Springer Link* (Vol. 91). doi: 10.1017/CBO9781107415324.004.

Rovid Spickler, A. (2005). *Giardiasis*. Retrieved from www.cfsph.iastate.edu.

Ryan, U., Hijjawi, N., Feng, Y., & Xiao, L. (2019). Giardia: An under-reported foodborne parasite. *International Journal for Parasitology*, *49*(1), 1–11. doi: 10.1016/J.IJPARA.2018.07.003.

Saaed, F. M. A., & Ongerth, J. E. (2019). Giardia and Cryptosporidium in children with diarrhea, Kufra, Libya, a North African migration route city. *International Journal of Hygiene and Environmental Health*, *222*(5), 840–846. doi: 10.1016/j.ijheh.2019.04.006.

Samuelson, J., & Robbins, P. (2011, January 1). A simple fibril and lectin model for cyst walls of Entamoeba and perhaps Giardia. *Trends in Parasitology*, *27*, 17–22. doi: 10.1016/j.pt.2010.09.002.

Sauch, J. F. (1984). Purification of *Giardia muris* Cysts by Velocity Sedimentation. In *Applied and Environmental Microbiology* (Vol. *48*). Retrieved from http://aem.asm.org/.

Walderich, B., Müller, L., Bracha, R., Knobloch, J., & Burchard, G. D. (1997). A new method for isolation and differentiation of native Entamoeba histolytica and E. dispar cysts from fecal samples. *Parasitology Research*, *83*(7), 719–721. doi: 10.1007/s004360050326.

Winiecka-Krusnell, J., & Linder, E. (1998). Cysticidal effect of chlorine dioxide on Giardia intestinalis cysts. *Acta Tropica*, *70*(3), 369–372. doi: 10.1016/S0001-706X(98)00036-9.

180

9.2 Appendix B

Supplementary Material to Chapter 3

Source code for sequence processing and analysis source code in R.

dada2 analysis - alternate database and trimmed reads

Patrick J Biggs, 2020

28/05/2020

The dada2 side

Getting set up

Load the required packages:

```
library(dada2)

packageVersion("dada2")

library(ShortRead)

packageVersion("ShortRead")

library(ggplot2)
packageVersion("ggplot2")
```

The path

Define a path variable to check it is all OK for the work we are going to do:

```
# is our path OK?
path <- ("C:/users/pobig/Documents/Massey/MGS_Data/MGS00195_1_Paul_Ogbuigw
e_Delivery/MGS00195_1_Paul_Ogbuigwe_Delivery/MGS00195_1/fQsequences/")
#path <- ("/home/pbiggs/extraDrive2/students/PaulO/MGS00142_1/fQsequences/
")
path

fns <- list.files(path)
fns
```

Collecting our data

```r
# extract out our fastq sequences
fastqs <- fns[grepl(".fastq$", fns)]
fastqs

# sort them to ensure reads are in the same order
fastqs <- sort(fastqs)

# make sub-lists for the forward and reverse reads
fnFs <- fastqs[grepl("_R1", fastqs)] # Just the forward read files
fnRs <- fastqs[grepl("_R2", fastqs)] # Just the reverse read files

# get the sample.names
sample.names <- sapply(strsplit(fnFs, "_"), `[`, 1)
#sample.names <- sapply(strsplit(fnFs, "_"), function(x){paste(x[[1]], x[[2]], sep="_")})
sample.names

# specify the full path to the fnFs and fnRs
fnFs <- file.path(path, fnFs)
fnRs <- file.path(path, fnRs)
fnFs

fnRs
```

Examine the quality profiles of forward and reverse reads

It is important to look at your data. We start by visualizing the quality profiles along the sequencing reads

```r
# Visualize the quality profile of the forward reads
plotQualityProfile(fnFs[1:2])

#plotQualityProfile(fnFs[[12]])
plotQualityProfile(fnRs[1:2])

#plotQualityProfile(fnRs[[12]])
```

Perform filtering and trimming

```r
# make directory and filenames for the filtered fastqs
filt_path <- file.path(path, "filtered")
if(!file_test("-d", filt_path)) dir.create(filt_path)
filt_path

# make list of filtered names for later
filtFs <- file.path(filt_path, paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(filt_path, paste0(sample.names, "_R_filt.fastq.gz"))

# Perform the trimming and filtering
```

```
#out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(240,160), max
N=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE, compress=TRUE, multithread=TRUE
)
#out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, maxN=0, truncLen=230, ma
xEE=c(2,2), truncQ=2, rm.phix=TRUE, compress=TRUE, multithread=TRUE)
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, maxN=0, truncLen = 230, t
rimLeft = 20, trimRight = 19, maxEE=c(2,2), truncQ=2, rm.phix=TRUE, compre
ss=TRUE, multithread=TRUE)

## 27-May-2020 - not sure why I have to trim only 18bp off the Read 2 end

out
```

Dereplication

```
# dereplicate the forward and reverse reads separately
derepFs <- derepFastq(filtFs, verbose=TRUE)

derepRs <- derepFastq(filtRs, verbose=TRUE)

# rename the derep-class objects by the sample names
names(derepFs) <- sample.names
names(derepRs) <- sample.names
names(derepFs)
```

Error rates

```
# forward reads first and then look at the output
start_time <- Sys.time()
errF <- learnErrors(derepFs, multithread=TRUE)

## 118574190 total bases in 564639 reads from 5 samples will be used for l
earning the error rates.

dadaFs.lrn <- dada(derepFs, err=errF, multithread=TRUE)

## Sample 1 - 55235 reads in 9708 unique sequences.
## Sample 2 - 113042 reads in 14843 unique sequences.
## Sample 3 - 136387 reads in 18298 unique sequences.
## Sample 4 - 118718 reads in 16385 unique sequences.
## Sample 5 - 141257 reads in 17792 unique sequences.
## Sample 6 - 95812 reads in 15391 unique sequences.
## Sample 7 - 95836 reads in 22012 unique sequences.
## Sample 8 - 106343 reads in 16210 unique sequences.
## Sample 9 - 8184 reads in 2792 unique sequences.
## Sample 10 - 144483 reads in 26256 unique sequences.
## Sample 11 - 116354 reads in 22439 unique sequences.
## Sample 12 - 121820 reads in 14773 unique sequences.
## Sample 13 - 20678 reads in 5902 unique sequences.
## Sample 14 - 103784 reads in 18313 unique sequences.
## Sample 15 - 75624 reads in 11425 unique sequences.
```

```
## Sample 16 - 112267 reads in 12583 unique sequences.
## Sample 17 - 87460 reads in 11150 unique sequences.

end_time <- Sys.time()
end_time - start_time

start_time1 <- Sys.time()
errR <- learnErrors(derepRs, multithread=TRUE)

## 118574190 total bases in 564639 reads from 5 samples will be used for l
earning the error rates.

dadaRs.lrn <- dada(derepRs, err=errR, multithread=TRUE)

## Sample 1 - 55235 reads in 15404 unique sequences.
## Sample 2 - 113042 reads in 21061 unique sequences.
## Sample 3 - 136387 reads in 25493 unique sequences.
## Sample 4 - 118718 reads in 23734 unique sequences.
## Sample 5 - 141257 reads in 24507 unique sequences.
## Sample 6 - 95812 reads in 19814 unique sequences.
## Sample 7 - 95836 reads in 28269 unique sequences.
## Sample 8 - 106343 reads in 24744 unique sequences.
## Sample 9 - 8184 reads in 3121 unique sequences.
## Sample 10 - 144483 reads in 39869 unique sequences.
## Sample 11 - 116354 reads in 33518 unique sequences.
## Sample 12 - 121820 reads in 21446 unique sequences.
## Sample 13 - 20678 reads in 6331 unique sequences.
## Sample 14 - 103784 reads in 22832 unique sequences.
## Sample 15 - 75624 reads in 15918 unique sequences.
## Sample 16 - 112267 reads in 19785 unique sequences.
## Sample 17 - 87460 reads in 15655 unique sequences.

end_time1 <- Sys.time()
end_time1 - start_time1

# plot the errors as a trellis plot
plotErrors(errF, nominalQ=TRUE)
```

Sample inference and merging paired reads

```
# Infer the sequence variants in each sample
#dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
#dadaRs <- dada(derepRs, err=errR, multithread=TRUE)

dadaFs <- dadaFs.lrn
dadaRs <- dadaRs.lrn

# inspect the results in more detail
dadaFs[[1]]

# Merge the denoised forward and reverse reads
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
```

```
# Inspect the merger data.frame from the first sample
head(mergers[[1]])
```

Constructing the sequence table and removing chimaeras

```
# Construct sequence table
seqtab <- makeSequenceTable(mergers)
dim(seqtab)

#head(seqtab)

# look at the top 2 x 2 only
seqtab[1:2, 1:2]

# Inspect the distribution of sequence lengths
table(nchar(getSequences(seqtab)))

# Remove chimeric sequences
seqtab.nochim <- removeBimeraDenovo(seqtab, verbose=TRUE)

sum(seqtab.nochim)/sum(seqtab)
```

Checking our progress

```
# Remove chimeric sequences with some complicated code
getN <- function(x) sum(getUniques(x)) # our first function
track <- cbind(out, sapply(dadaFs, getN), sapply(dadaRs, getN), sapply(mer
gers, getN), rowSums(seqtab.nochim))
colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR", "merge
d",
"nonchim")
rownames(track) <- sample.names
track
```

modify seqtab with only sequences that have hits for *Giardia* in BLAST search

```
modtab <- seqtab
modtab.df <- as.data.frame(t(modtab))
modtab.df$seq <- row.names(modtab.df)
row.names(modtab.df) <- paste0("seq", seq(1:nrow(modtab.df)))
no_hits <- read.table("no_hits.txt", header=TRUE)
ditch <- no_hits$SeqID
'%nin%' <- Negate('%in%')
modtab_nohit <- modtab.df[!row.names(modtab.df)%nin%ditch,]
hits <- read.table("hits.txt", header=TRUE)
punt <- hits$SeqID
modtab_hit <- modtab.df[!row.names(modtab.df)%nin%punt,]
modtab_hitn <- modtab_hit
row.names(modtab_hitn) <- (modtab_hit$seq)
modtab_hitn$seq <- NULL
```

```
modtab_hitz <- data.matrix(modtab_hitn)
modtab_hitz <- t(modtab_hitz)
```

what taxonomy can we do with the database?

We need to modify the previous fasta file to chnage the headers to add in the strain names, thus:

">XXX" becomes
">Eukaryota;Diplomonadida;Hexamitidae;*Giardia*;*Giardia_intestinalis*;XXX;" when using the file '*gdh*_DB_20200130mod.fa.'

However, for the file '*gdh*_DB_20200130modNames.fa', ">XXX" becomes
">Eukaryota;Fornicata;Hexamitidae;*Giardia*;*Giardia_intestinalis*;" as we are looking at the assmblage being the genus and the pru # being the species. so…

```
# use a reference data set to assign taxonomy to the reads
#training <- ("/home/pbiggs/extraDrive2/students/PaulO/dada2_analysis/giar
dia/allFastaGoodMod.fa")
#training <- ("/home/pbiggs/extraDrive2/students/PaulO/dada2_analysis/giar
dia/gdh_DB_20200130modNames.fa")
#training <- ("/home/pbiggs/extraDrive2/students/PaulO/dada2_analysis/giar
dia/gdh_DB_20200130mod.fa")
#training <- ("/home/pbiggs/extraDrive2/students/PaulO/MGS00195_1/referenc
e/gdh_DB_20200130.fa.gz")
#training <- ("/home/pbiggs/Dropbox/PhDstudents/current/Paul/GiardiaWork/n
ew393combinedFull2.fasta")
training <- ("C:/Users/pobig/Documents/Bioinformatics/final_Giardia/new393
combinedFull2.fasta")
training

taxa <- assignTaxonomy(modtab_hitz, training, tryRC=TRUE, taxLevels = c("K
ingdom", "Phylum", "Class",
  "Order", "Family", "Assemblage", "Subtype"), multithread=FALSE)

# Removing sequence rownames for display only
taxa.print <- taxa
rownames(taxa.print) <- NULL
head(taxa.print)

library("readxl")        # necessary to import the data from Excel file
library("dplyr")         # filter and reformat data frames

library("tibble")        # Needed for converting column to row names

taxa <- read_excel("C:/Users/pobig/Documents/Bioinformatics/final_Giardia/
taxa.xlsx", sheet = "taxa")

taxa <- taxa %>%
  tibble::column_to_rownames("Seqs")
taxa <- as.matrix(taxa)

taxa.print <- taxa
```

```
rownames(taxa.print) <- NULL
head(taxa.print)
```

On to phyloseq

Let's load our packages

```
# load our required packages
library(phyloseq)

packageVersion("phyloseq")

# starting to make a dataframe for the samples by getting their names
samples.out <- rownames(modtab_hitz)
samples.out
```

let's load in the smaple data

```
# load in sample metadata
#sampleIn <- read.table("C:/Users/pobig/Documents/Massey/Metabarcoding/dad
a2_test/sample2.txt", header = TRUE)
#sampleD <- as.data.frame(sampleIn)
#class(sampleIn)
library ("readxl")
library ("tibble")
sampleIn <- read_excel("C:/Users/pobig/Documents/Bioinformatics/final_Giar
dia/sample2_Copy.xlsx", sheet = "sampleMeta")
sampleIn <- sampleIn %>%
  tibble::column_to_rownames("Sample")

samples = sample_data(sampleIn)
```

now to make a phyloseq object

```
ps <- phyloseq(otu_table(modtab_hitz, taxa_are_rows=FALSE), sample_data(sa
mpleIn), tax_table(taxa))
ps
```

let's do some basic plotting

```
# add a colour palette for people with deuteranopia ;)
library("ggplot2")

cblind <- (c("#1D91C0", "#624B27", "#CB181D", "#F46D43", "#FAE093", "#A6CE
E3", "#74C476", "#EF3B2C", "#000000", "#004949", "#009292", "#F9E211", "#B
A2F00", "#425266", "#3D4928", "#008CEC"))

# plotting diversity by day using the Shannon and Simpson measures
plot_richness(ps, x="LibraryID", measures=c("Shannon", "Simpson"), color="
Outbreak_Location") + scale_colour_manual(values = cblind)

# a quick look at the data via an ordination method
ord.nmds.bray <- ordinate(ps, method="NMDS", distance="bray")
```

```
plot_ordination(ps, ord.nmds.bray, color="Outbreak_Location", title="Bray
NMDS of MGS00195") + scale_colour_manual(values = cblind)
```

change the plotting here for the top 50, but this time, can plot the data based on Genus -
which for us is assemblage.

```
# what about the top 50 taxa?
top50 <- names(sort(taxa_sums(ps), decreasing=TRUE))[1:50]
head(top50)

ps.top50 <- transform_sample_counts(ps, function(OTU) OTU/sum(OTU))
ps.top50 <- prune_taxa(top50, ps.top50)

plot_bar(ps.top50, x="Sanger_type", fill="Assemblage") + facet_wrap(~Outbr
eak_Location) + scale_colour_manual(values = cblind)

# exploratory bar plots
par(mar = c(10, 4, 4, 2) + 0.1) # make more room on the bottom margin
N <- 50
barplot(sort(taxa_sums(ps), TRUE)[1:N]/nsamples(ps), las=2)

# plot a basic heatmap
pH <- plot_heatmap(ps.top50, method = "NMDS", distance = "bray", sample.la
bel="LibraryID", sample.order = "LibraryID", taxa.label = "Assemblage", ta
xa.order = "Assemblage")
pH

#Hawke's Bay 2010

out1 <- subset_samples(ps, Outbreak_Location == "Hawke's Bay_2010")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top50 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top50)

out1.top50 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top50 <- prune_taxa(top50, out1.top50)

out1H <- plot_heatmap(out1.top50, method = "NMDS", distance = "bray", samp
le.label="LibraryID", sample.order = "LibraryID", taxa.label = "Assemblage
", taxa.order = "Assemblage")
out1H

#Gisborne 2014

out1 <- subset_samples(ps, Outbreak_Location == "Gisborne_2014")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top50 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top50)
```

```
out1.top50 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top50 <- prune_taxa(top50, out1.top50)

out1H <- plot_heatmap(out1.top50, method = "NMDS", distance = "bray", samp
le.label="LibraryID", sample.order = "LibraryID", taxa.label = "Assemblage
", taxa.order = "Assemblage")
out1H

#Hawke's Bay 2015

out1 <- subset_samples(ps, Outbreak_Location == "Hawke's Bay_2015")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top50 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top50)

out1.top50 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top50 <- prune_taxa(top50, out1.top50)

out1H <- plot_heatmap(out1.top50, method = "NMDS", distance = "bray", samp
le.label="LibraryID", sample.order = "LibraryID", taxa.label = "Assemblage
", taxa.order = "Assemblage")
out1H

#Auckland 2017

out1 <- subset_samples(ps, Outbreak_Location == "Auckland_2017")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top50 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top50)

out1.top50 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top50 <- prune_taxa(top50, out1.top50)

out1H <- plot_heatmap(out1.top50, method = "NMDS", distance = "bray", samp
le.label="LibraryID", sample.order = "LibraryID", taxa.label = "Assemblage
", taxa.order = "Assemblage")
out1H
```

Massey ID	LibraryID	Number	Organism	Outbreak	Location	Year	Sanger_type	Source	Description	Outbreak_Location
1997	1997_S1	1	Giardia	TRUE	HawkesBay	2010	BIV	Human	1997_S1	Hawke's Bay_2010
1998	1998_S2	2	Giardia	TRUE	HawkesBay	2010	BIV	Human	1998_S2	Hawke's Bay_2010

1999	1999_S3	3	*Giardia*	TRUE	HawkesBay	2010	BIV	Human	1999_S3	Hawke's Bay_2010
10015	10015_S4	4	*Giardia*	TRUE	Gisborne	2014	All	Human	10015_S4	Gisborne_2014
10046	10046_S5	5	*Giardia*	TRUE	Gisborne	2014	BIV	Human	10046_S5	Gisborne_2014
10047	10047_S6	6	*Giardia*	TRUE	Gisborne	2014	BIV	Human	10047_S6	Gisborne_2014
10048	10048_S7	7	*Giardia*	TRUE	Gisborne	2014	BIV	Human	10048_S7	Gisborne_2014
10049	10049_S8	8	*Giardia*	TRUE	Gisborne	2014	BIV	Human	10049_S8	Gisborne_2014
10936	10936_S9	9	*Giardia*	TRUE	HawkesBay	2015	BIV	Human	10936_S9	Hawke's Bay_2015
10937	10937_S10	10	*Giardia*	TRUE	HawkesBay	2015	BIV	Human	10937_S10	Hawke's Bay_2015
10938	10938_S11	11	*Giardia*	TRUE	HawkesBay	2015	All	Human	10938_S11	Hawke's Bay_2015
10939	10939_S12	12	*Giardia*	TRUE	HawkesBay	2015	BIV	Human	10939_S12	Hawke's Bay_2015
10940	10940_S13	13	*Giardia*	TRUE	HawkesBay	2015	BIII	Human	10940_S13	Hawke's Bay_2015
11359	11359_S16	14	*Giardia*	FALSE	Christchurch	2016	unknown	Human	11359_S16	Routine Surveillance_2016
13273	13273_S14	15	*Giardia*	TRUE	Auckland	2017	BIV	Human	13273_S14	Auckland_2017
13805	13805_S23	16	*Giardia*	FALSE	PalmerstonNorth	2017	unknown	Canine	13805_S23	Routine Surveillance_2017
14201	14201_S15	17	*Giardia*	FALSE	Otago	2017	unknown	Human	14201_S15	Routine Surveillance_2018

Table B.1. Sample data.

Figure B.1. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Blenheim in 2017. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.

9.3 Appendix C

Supplementary Material to Chapter 4

Sequence analysis source code made in R using phyloseq R Library.

test_phyloseq_crypto

Paul Ogbuigwe

25/04/2021

R Markdown

Load required

```r
library("phyloseq")
library("ggplot2")        # graphics
library("readxl")         # necessary to import the data from Excel file
library("dplyr")          # filter and reformat data frames

library("tibble")         # Needed for converting column to row names
library(Manu)
library(RColorBrewer)
library(viridis)
```

Sequence Data

Load sequence data

```r
otu_mat<- read_excel("./CRYPMETdata_V2.xlsx", sheet = "otuAbundance")
tax_mat<- read_excel("./CRYPMETdata_V2.xlsx", sheet = "otuTaxonomy")
samples_df <- read_excel("./CRYPMETdata_V2.xlsx", sheet = "sampleMeta")

# define row names

otu_mat <- otu_mat %>%
  tibble::column_to_rownames("otu")

tax_mat <- tax_mat %>%
  tibble::column_to_rownames("otu")

samples_df <- samples_df %>%
  tibble::column_to_rownames("Sample")
```

```
# transform into matrices and tax tables

otu_mat <- as.matrix(otu_mat)
tax_mat <- as.matrix(tax_mat)
```

Transform data into phyloseq objects

```
OTU = otu_table(otu_mat, taxa_are_rows = TRUE)
TAX = tax_table(tax_mat)
samples = sample_data(samples_df)

crypto <- phyloseq(OTU, TAX, samples)
crypto

# Visualise data

sample_names(crypto)

rank_names(crypto)

sample_variables(crypto)
```

Time for some plotting!

```
# add a colour palette for people with deuteranopia ;)

cblind <- (c("#1D91C0", "#624B27", "#CB181D", "#F46D43", "#FAE093", "#A6CE
E3", "#74C476", "#EF3B2C", "#000000", "#004949", "#009292", "#F9E211", "#B
A2F00", "#425266", "#3D4928", "#008CEC"))

crypto

#omit sequences with zero reads

cryptoNoZero <- prune_taxa(taxa_sums(crypto)>=1, crypto)
cryptoNoZero

# plotting diversity by outbreak using the Shannon and Simpson measures

plot_richness(cryptoNoZero, x = "LibraryID", measures=c("Shannon", "Simpso
n"), color = "Outbreak") + scale_colour_manual(values = cblind) #+ theme_b
w()

# a quick look at the data via an ordination method

ord.nmds.bray <- ordinate(cryptoNoZero, method="NMDS", distance="bray")

plot_ordination(cryptoNoZero, ord.nmds.bray, color="Outbreak", title="Bray
NMDS of MGS00142") + scale_colour_manual(values = cblind) #+ theme_bw()
```

change the plotting here for the top 50, but this time, can plot the data based on Subtype
(or gene family)

```
# what about the top 50 taxa?

top50 <- names(sort(taxa_sums(cryptoNoZero), decreasing=TRUE))[1:50]
head(top50)

crypto.top50 <- transform_sample_counts(cryptoNoZero, function(OTU) OTU/sum(OTU))
crypto.top50 <- prune_taxa(top50, crypto.top50)

crypto.top50AG <- tax_glom(crypto.top50, "Genotype")

cblind7 <- (c("#74C476", "#009292", "#F9E211", "#BA2F00", "#425266", "#3D4928", "#008CEC"))

plot_bar(crypto.top50AG, x="Sanger_type", fill="Genotype") + facet_wrap(~Outbreak, ncol = 3) + scale_fill_manual(values = cblind) #+ theme_bw()

#plot_bar(crypto.top50, x="Sanger_type", fill="Genotype") + facet_wrap(~Outbreak) + scale_fill_manual(values = cblind7) + theme_bw()


# exploratory bar plots

par(mar = c(10, 4, 4, 2) + 0.1) # make more room on the bottom margin
N <- 50
barplot(sort(taxa_sums(cryptoNoZero), TRUE)[1:N]/nsamples(cryptoNoZero), las=2)

# plot a basic heatmap

pH <- plot_heatmap(crypto.top50, method = "NMDS", distance = "bray", sample.label = "LibraryID", sample.order = "LibraryID", taxa.label = "Genotype", taxa.order = "Genotype")
#pH <- plot_heatmap(crypto.top50AG, method = "NMDS", distance = "bray", sample.label = "Year", taxa.label = "Genotype")
pH
```

#separate heatmap by outbreak

```
#Auckland 2010

out1 <- subset_samples(crypto, Outbreak == "Auckland_2010")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

out1H <- plot_heatmap(out2, method = "NMDS", distance = "bray", sample.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Christchurch_2010")

#remove sequences with no reads in this outbreak
```

```r
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sample.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Hawke's Bay_2013")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:100]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sample.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Waikato_2013")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:100]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sample.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Wellington_2013")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2
```

```
#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Taranaki_2013")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Auckland_2015")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Auckland_2017")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2
```

```r
#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Blenheim_2017")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

out1 <- subset_samples(crypto, Outbreak == "Wellington_2018")

#remove sequences with no reads in this outbreak
out2 <- prune_taxa(taxa_sums(out1)>=1, out1)
out2

#actually 50
top100 <- names(sort(taxa_sums(out2), decreasing=TRUE))[1:50]
head(top100)

out1.top100 <- transform_sample_counts(out2, function(OTU) OTU/sum(OTU))
out1.top100 <- prune_taxa(top100, out1.top100)

out1H <- plot_heatmap(out1.top100, method = "NMDS", distance = "bray", sam
ple.label = "MasseyID", sample.order = "LibraryID", taxa.label = "Genotype
", taxa.order = "Genotype")
out1H

#library("tinytex")
#tinytex::install_tinytex()
```

Table C.1. The top 12 sequences according to the number of samples in which they were present. The first column shows sequence identifier, the second column shows the subtype of that sequence, and the third column shows the percentage of samples in which at least one copy of that sequence was present.

Sequence	Subtype	Percentage
seq1	IbA10G2	77.1
seq2	IgA17	48.6
seq4	IIaA18G3R1	38.1
seq22	IgA16	27.6
seq3	IgA20	21.9
seq17	IgA11	20.0
seq45	IgA18	18.1
seq7	IfA12G1R5	15.2
seq354	IbA10G2	13.3
seq13	IIdA17G1	10.5
seq9	IIaA19G4R1	9.5
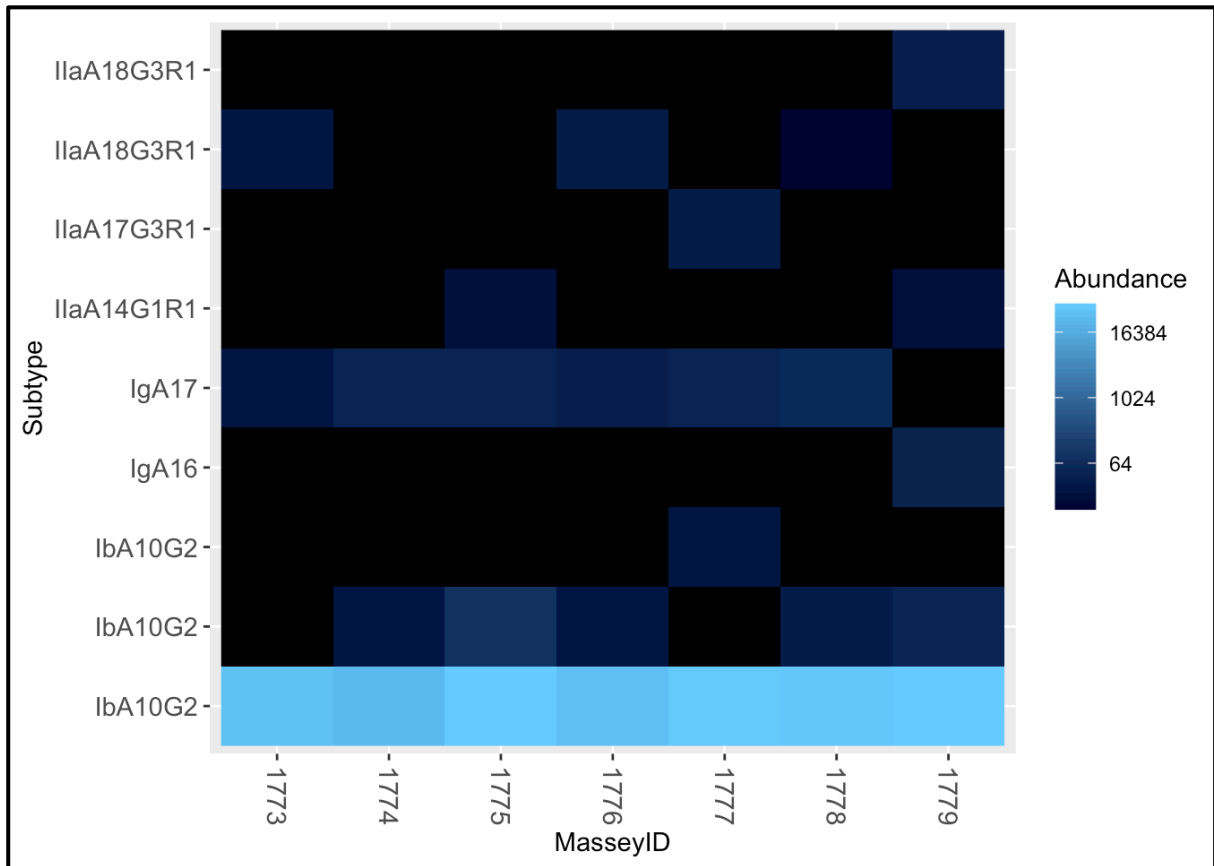
Figure C.1. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Auckland in 2010. The multiple variants (subtypes) of each genotype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.
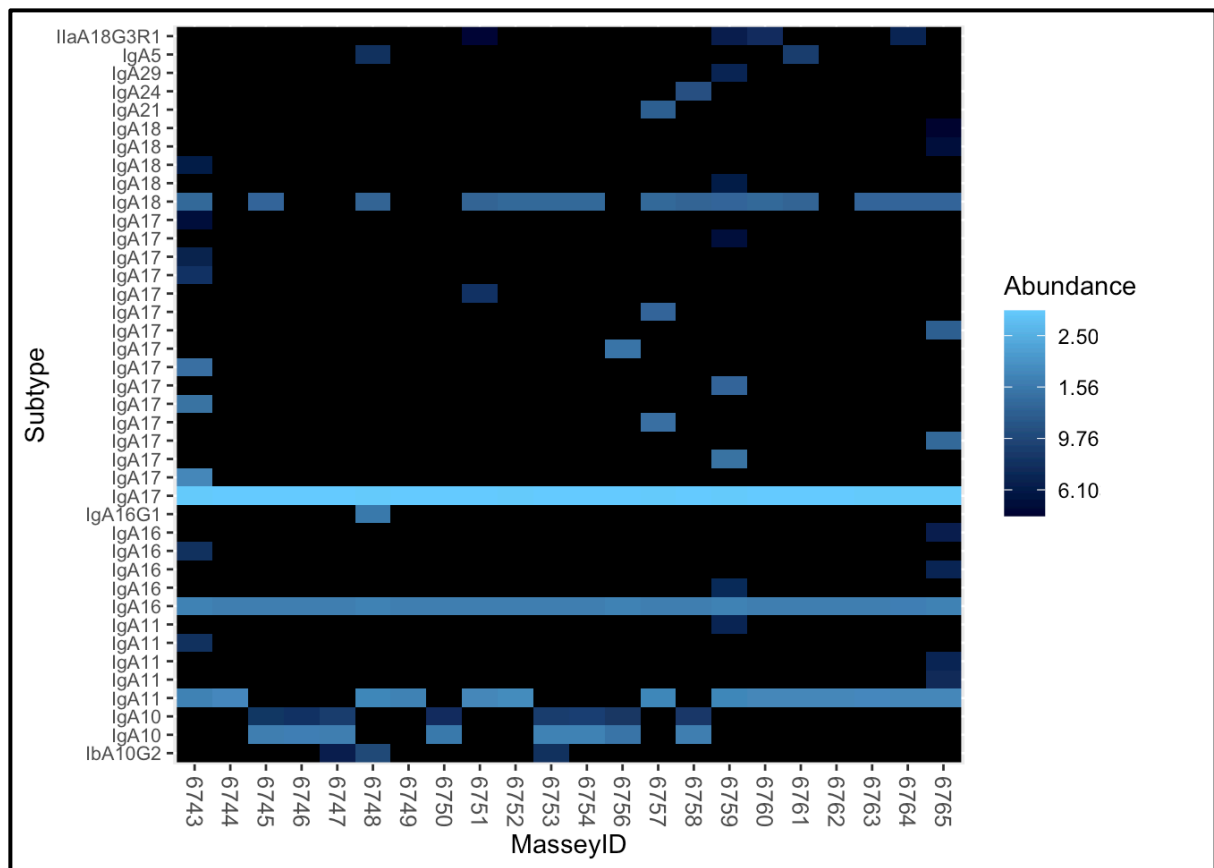
Figure C.2. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Hawke's Bay in 2013. The multiple variants (subtypes) of each genotype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.

Figure C.3. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Waikato in 2013. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.
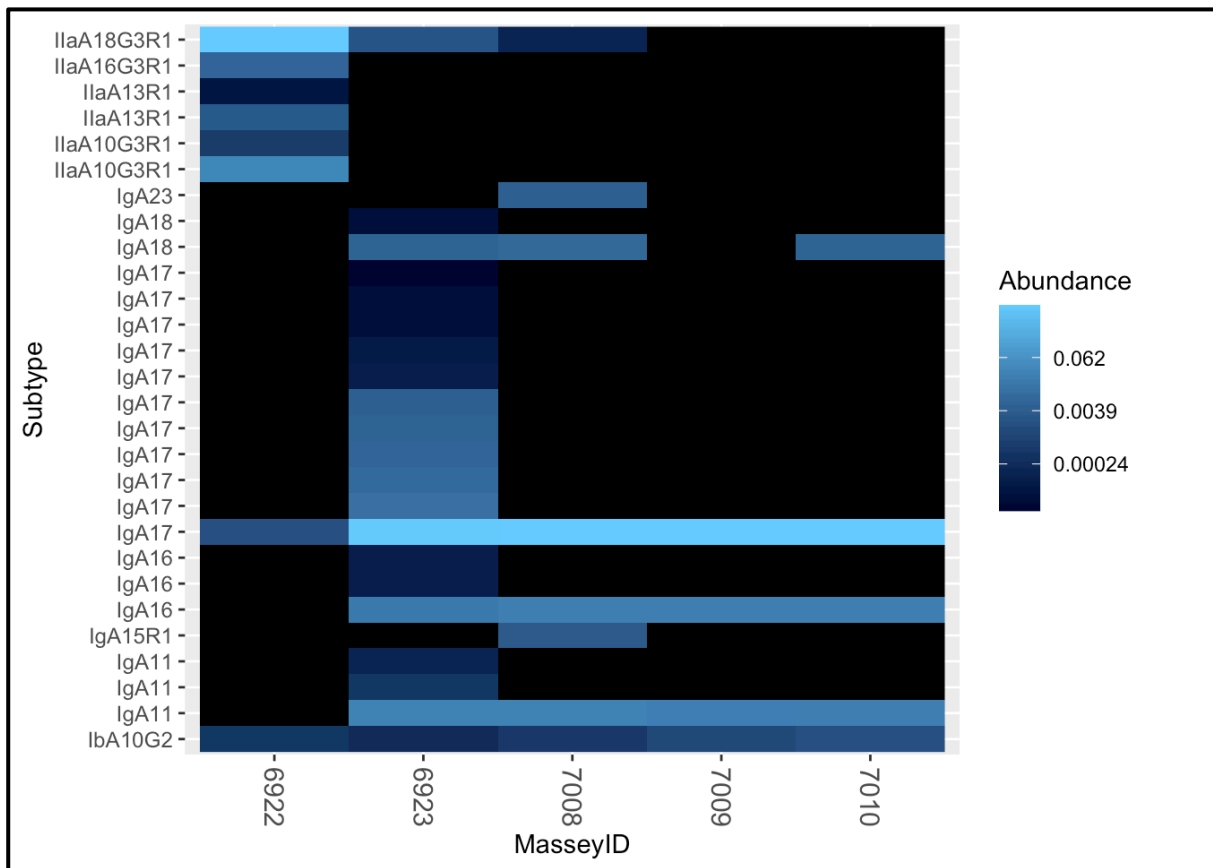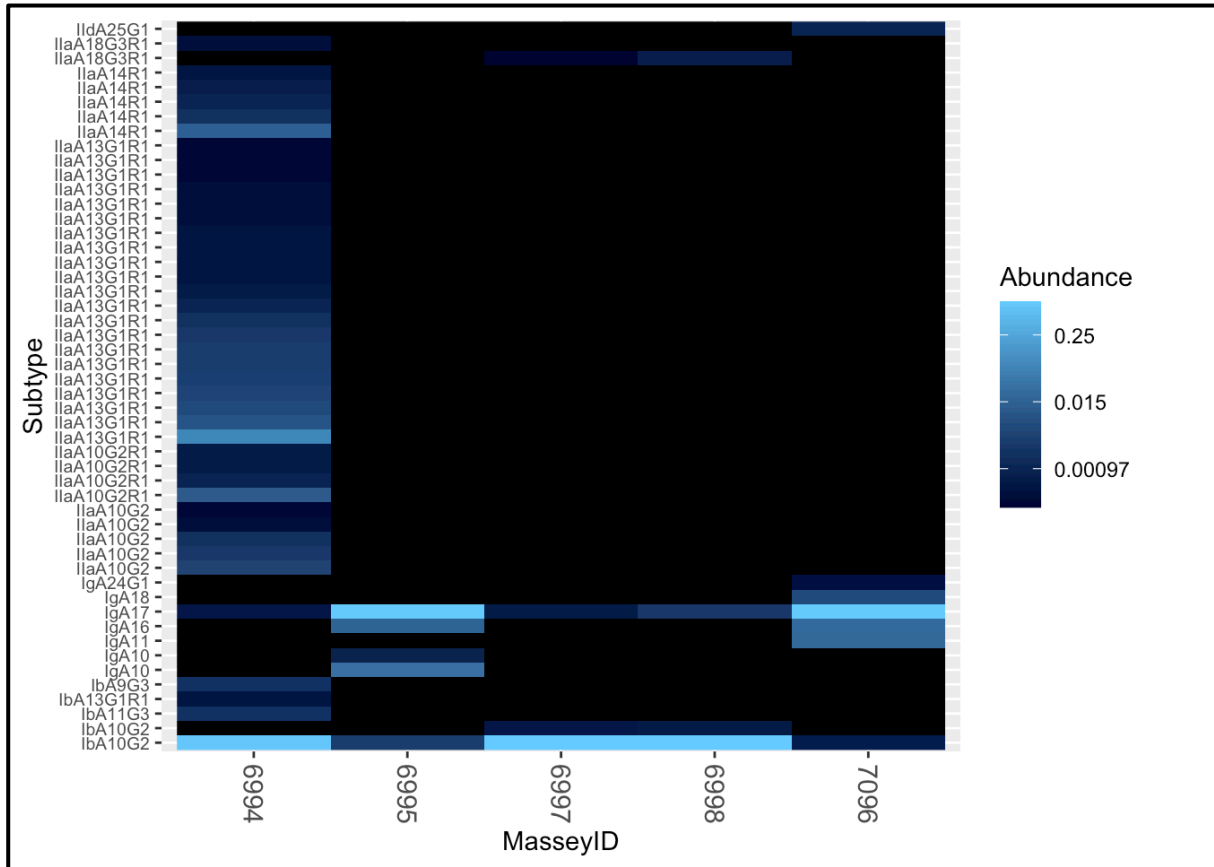
Figure C.4. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Wellington in 2013. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.
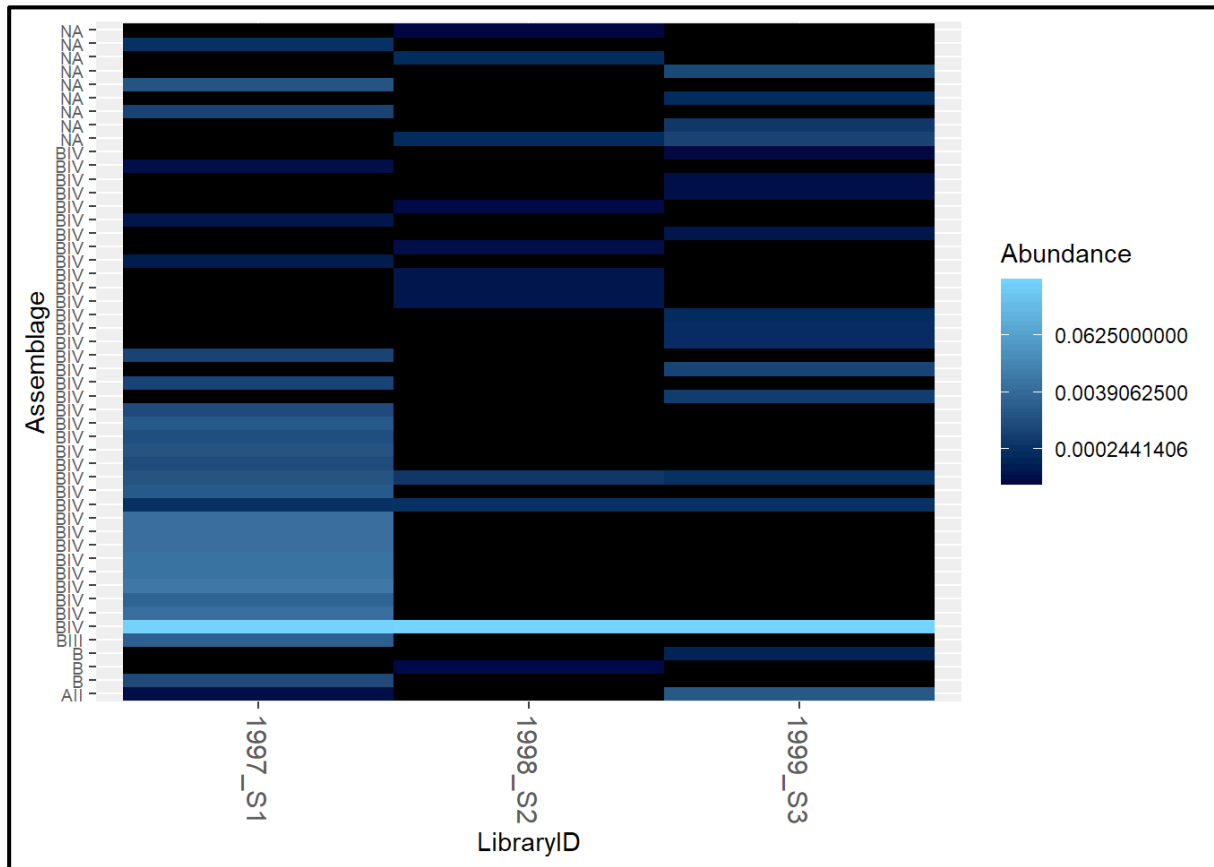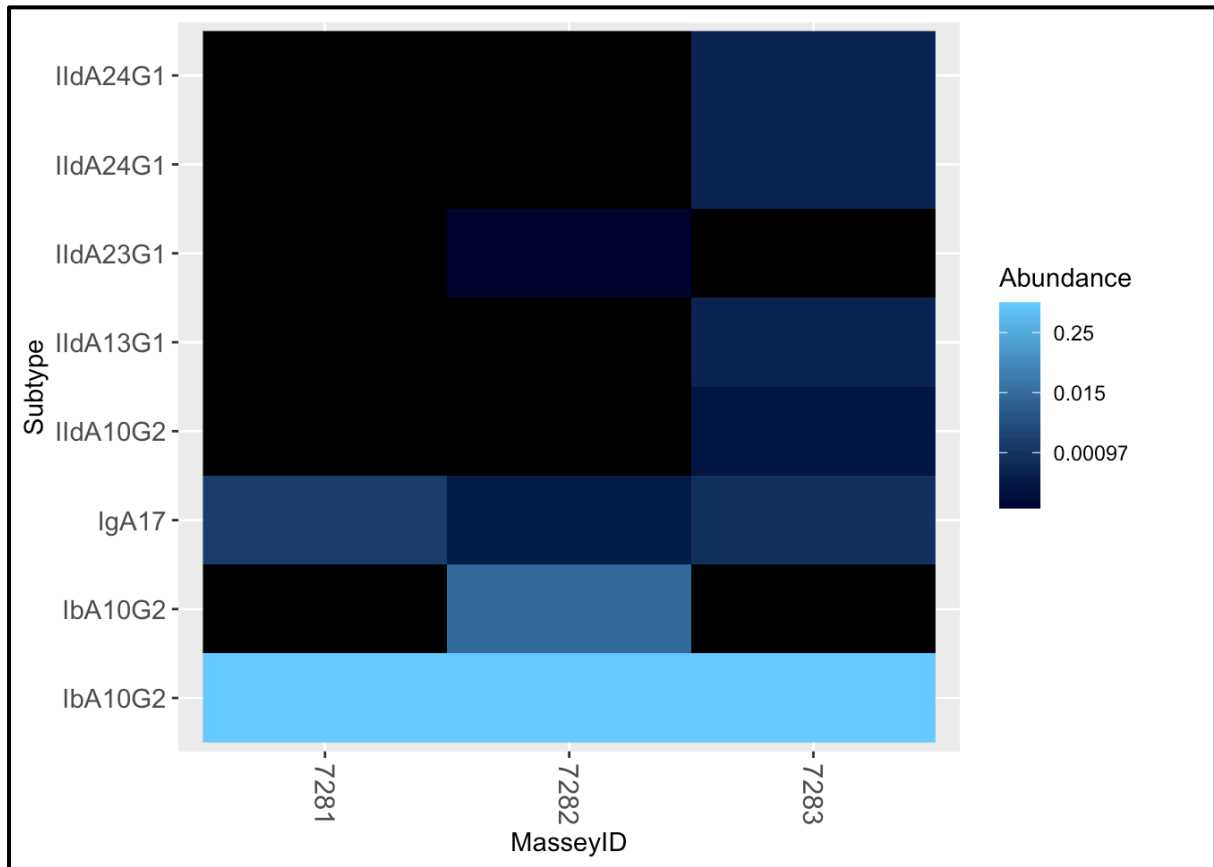
Figure C.5. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Taranaki in 2013. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.
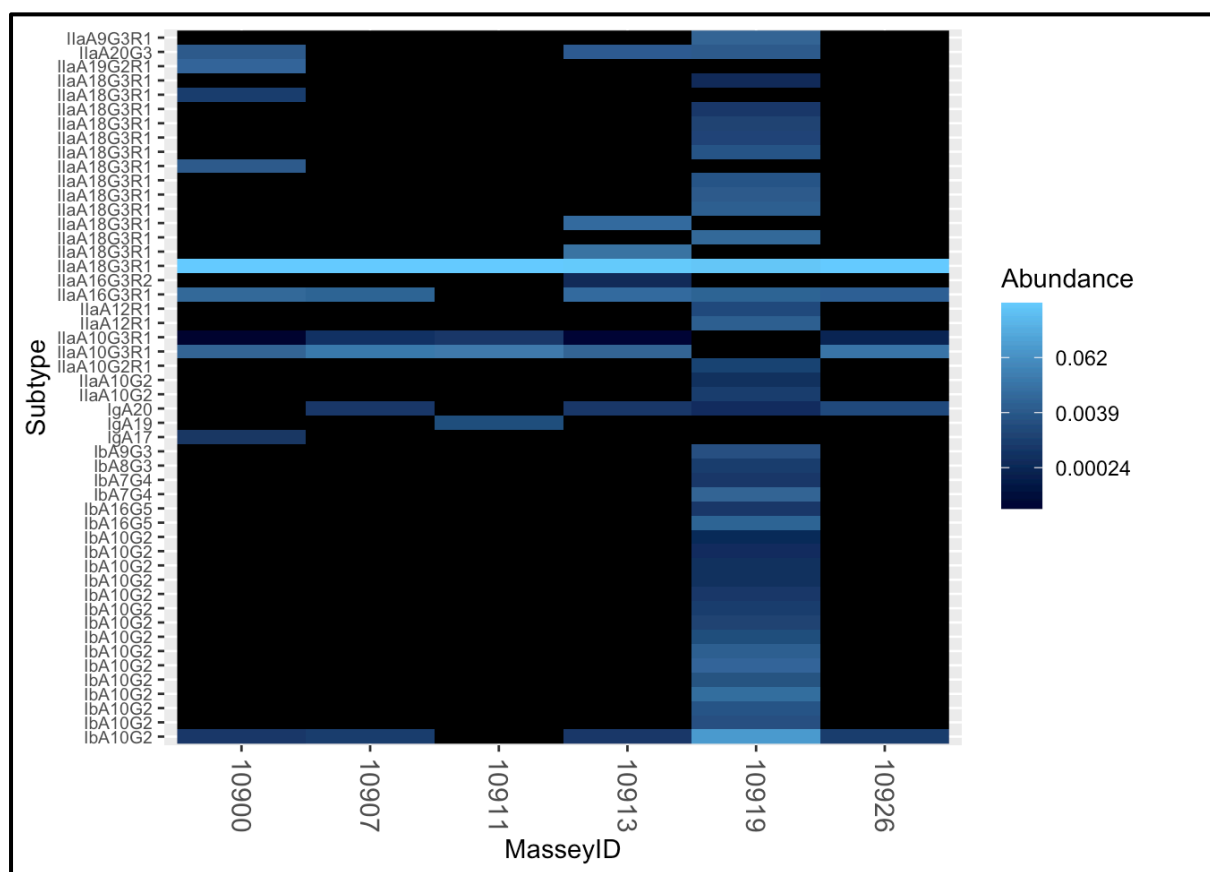
Figure C.6. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Auckland in 2015. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.

Figure C.7. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Auckland in 2017. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.
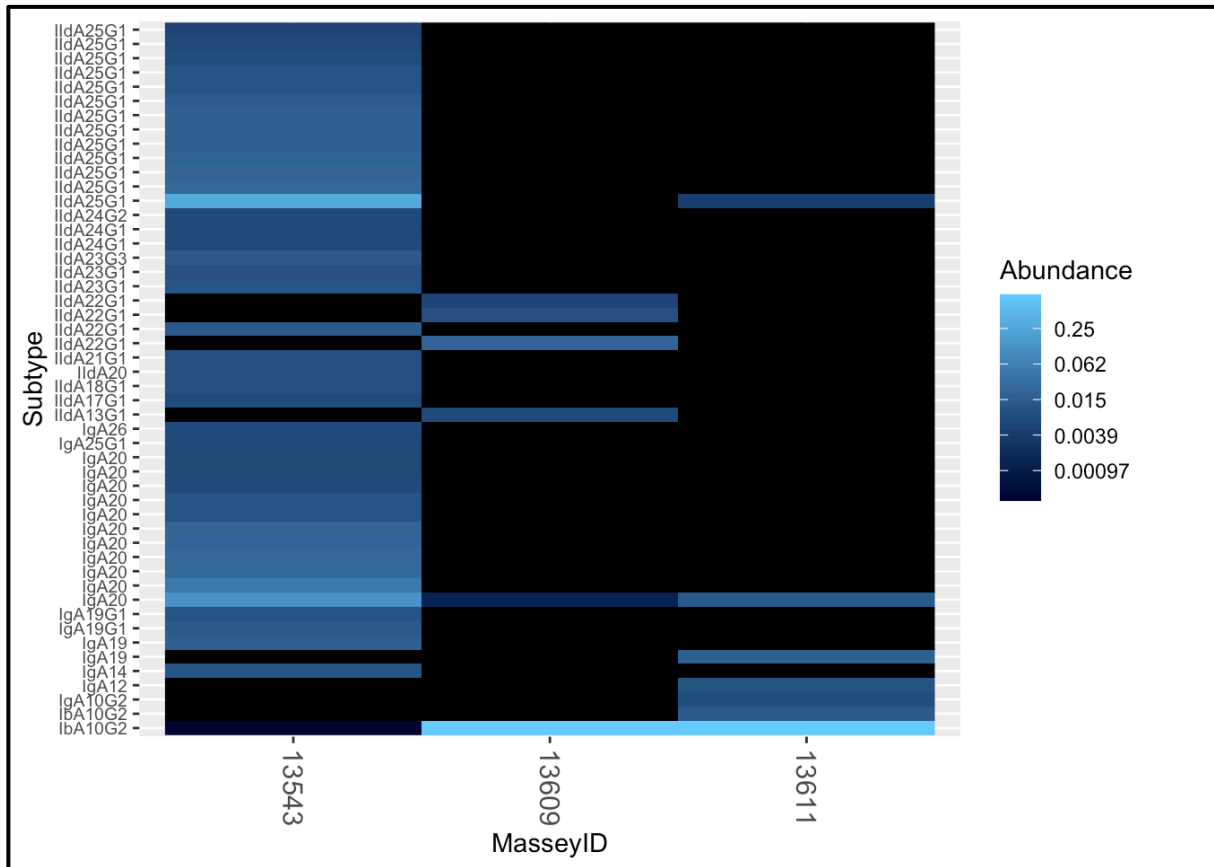
Figure C.8. Heatmap showing the relative abundance of the top 50 sequences present in samples from the outbreak of cryptosporidiosis that occurred in Blenheim in 2017. The multiple variants (subtypes) of each subtype present in each sample are displayed on the Y axis; the subtype of the most abundant variants and the hyper-transmissible IIaA15G2R1 variant are displayed in full on the same axis.

9.4 Appendix D

Supplementary material to Chapter 5.
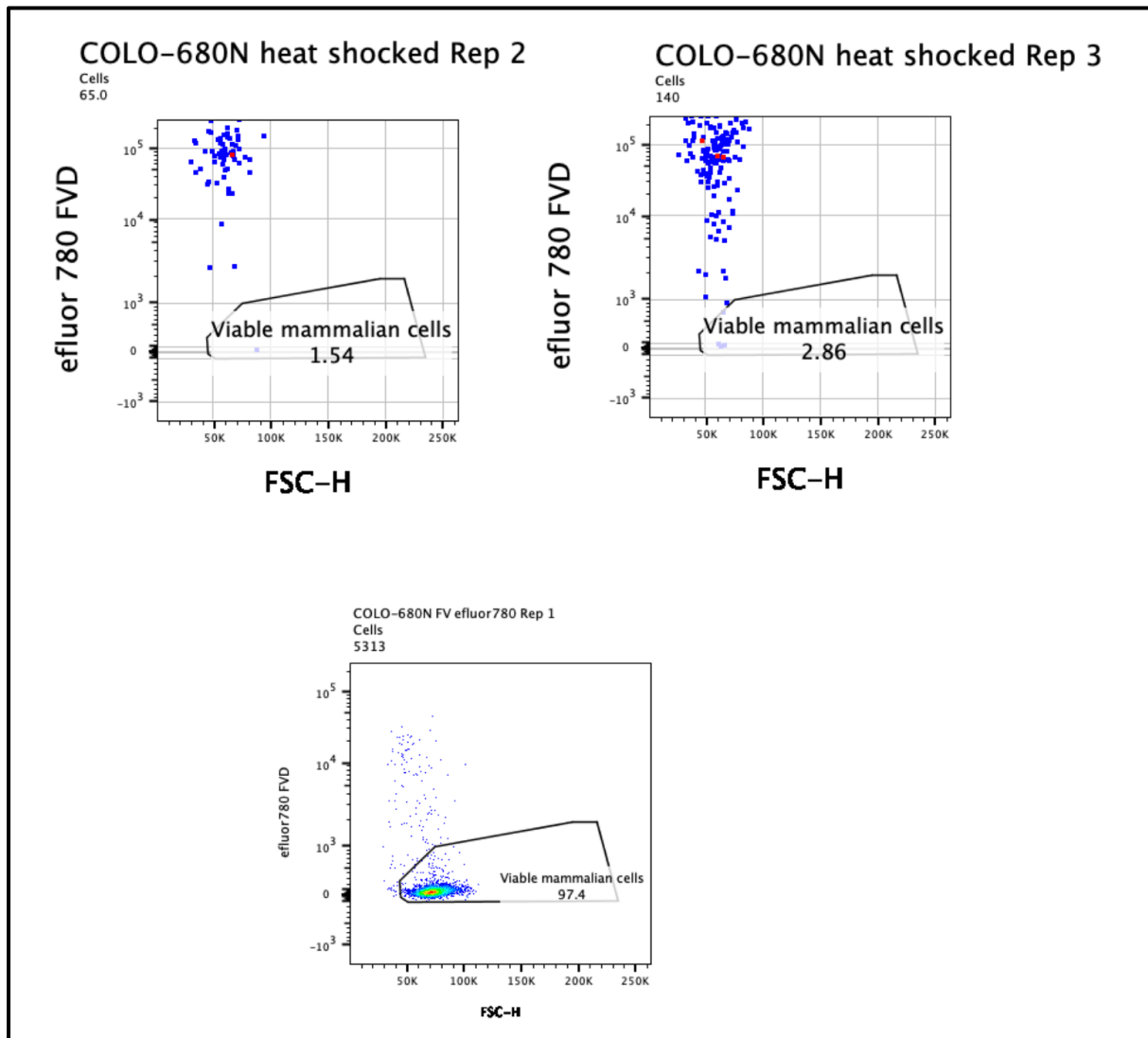
Figure D.1. Validating that eflour780 fixable viability dye (FVD)-positive cells are detected in region of FSC-H. FVD was used to distinguish viable cells from dead cells. This figure shows that the dead cells are detected withing the FSC-H region used for all samples in this study. Significantly fewer viable cells in heat-shocked samples (top panels) compared to control (bottom panel).
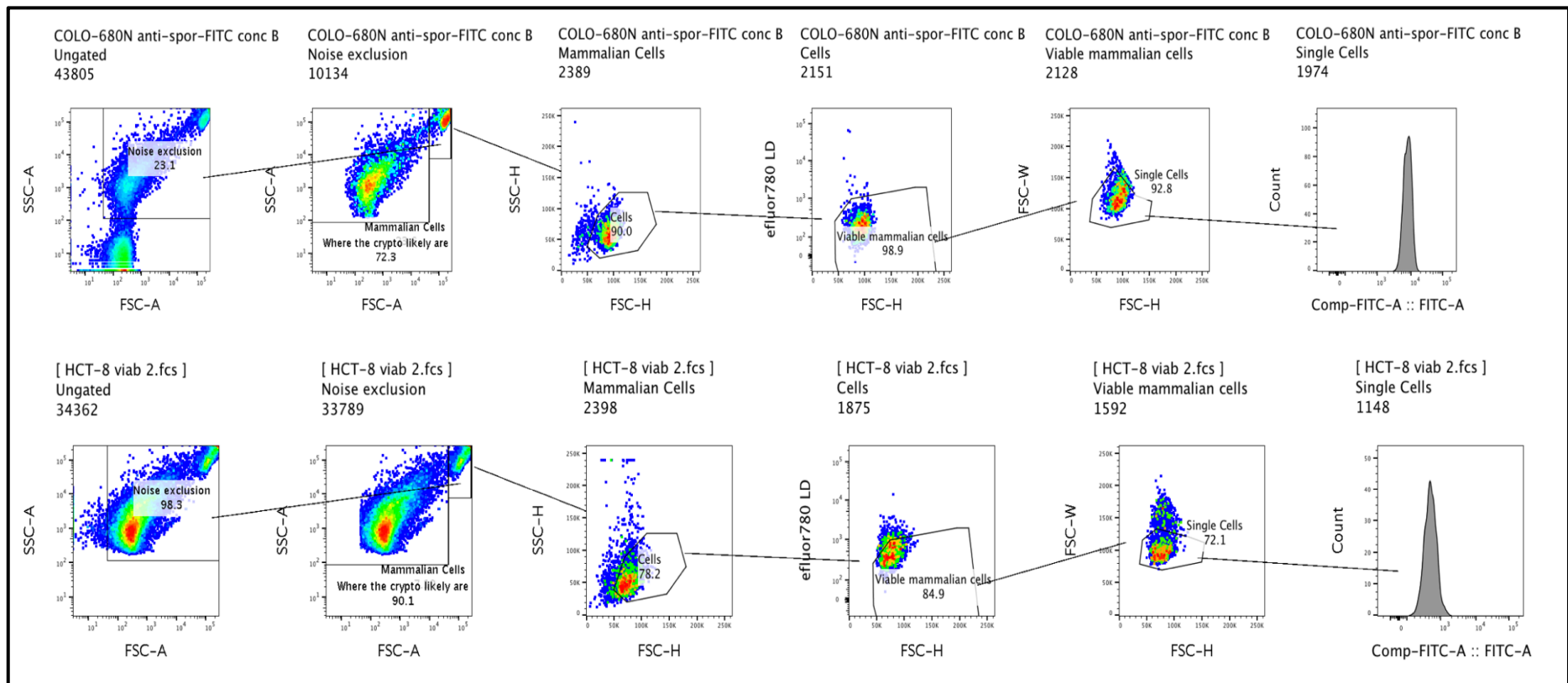
Figure D.2. Demonstrating gating hierarchy for mammalian cells. COLO-680N on top row, HCT-8 on bottom row.
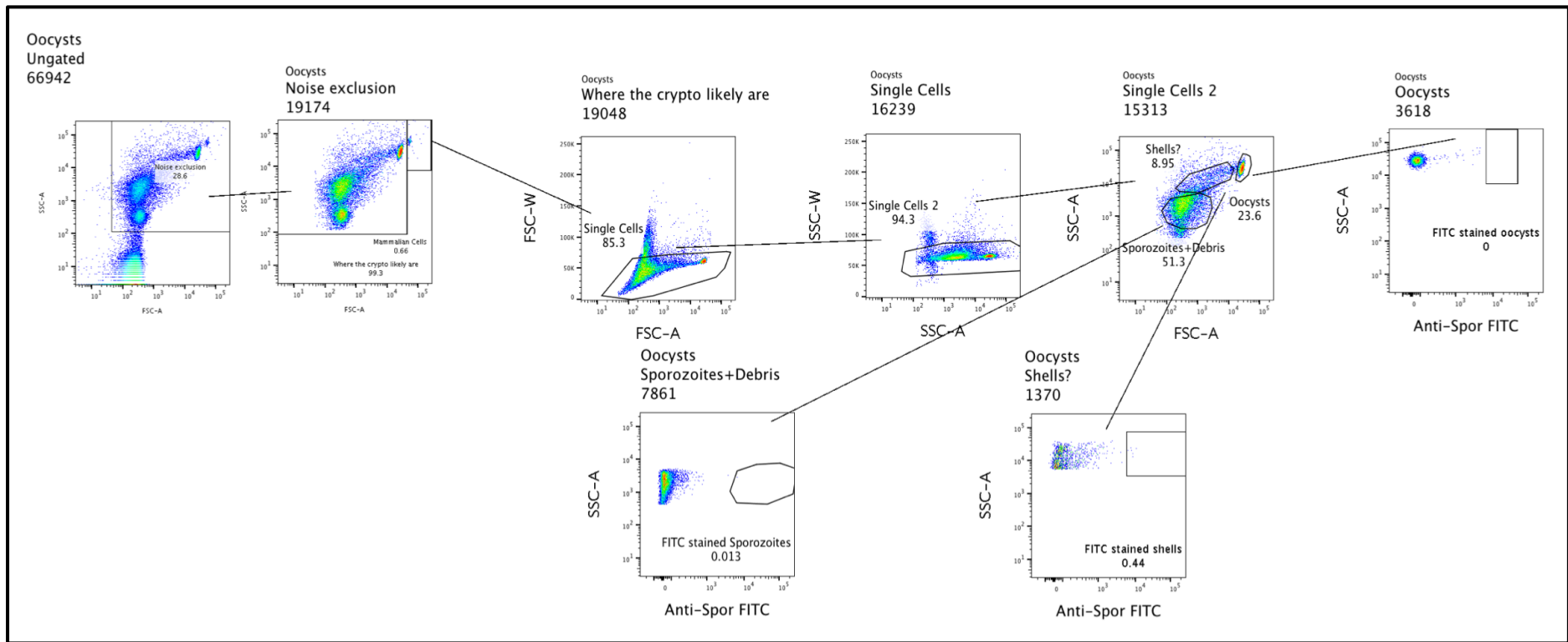
Figure D.3. Demonstrating gating hierarchy for oocysts.
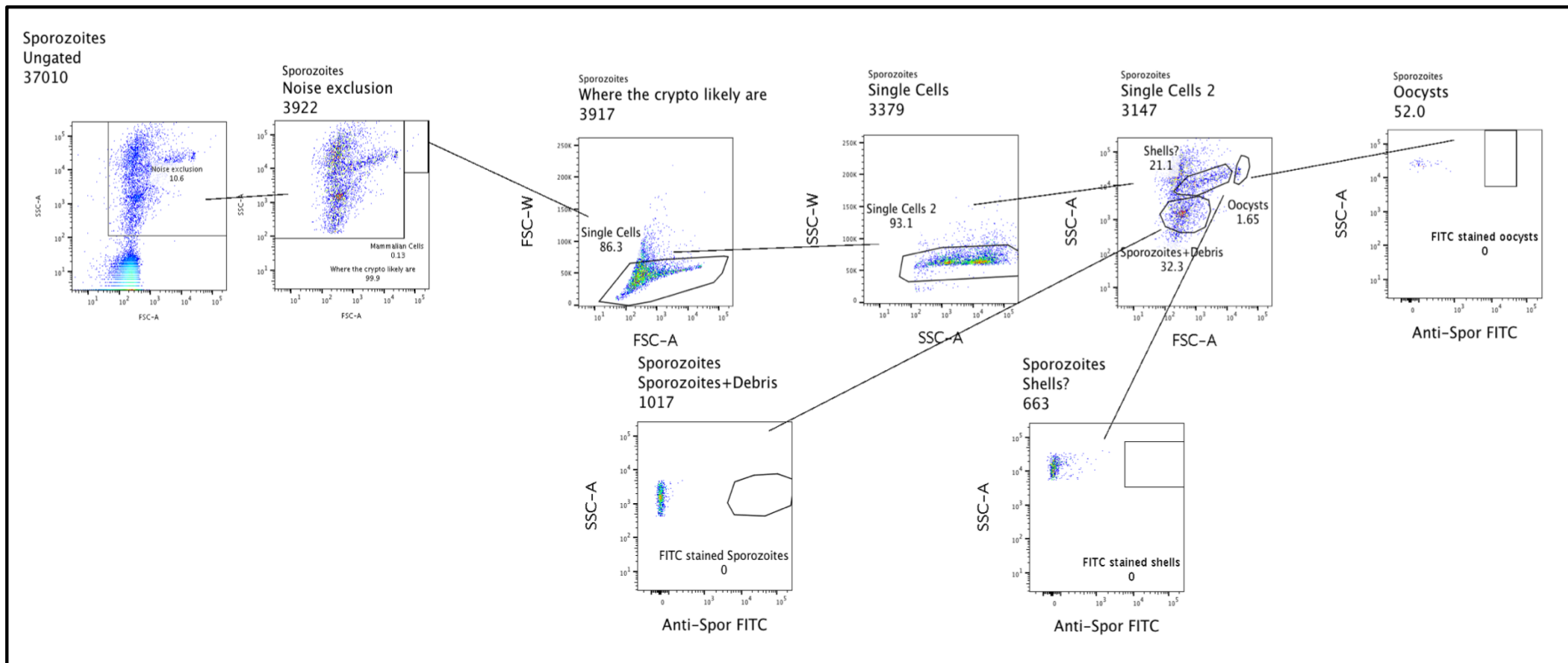
Figure D.4. Demonstrating gating hierarchy for unstained sporozoites.
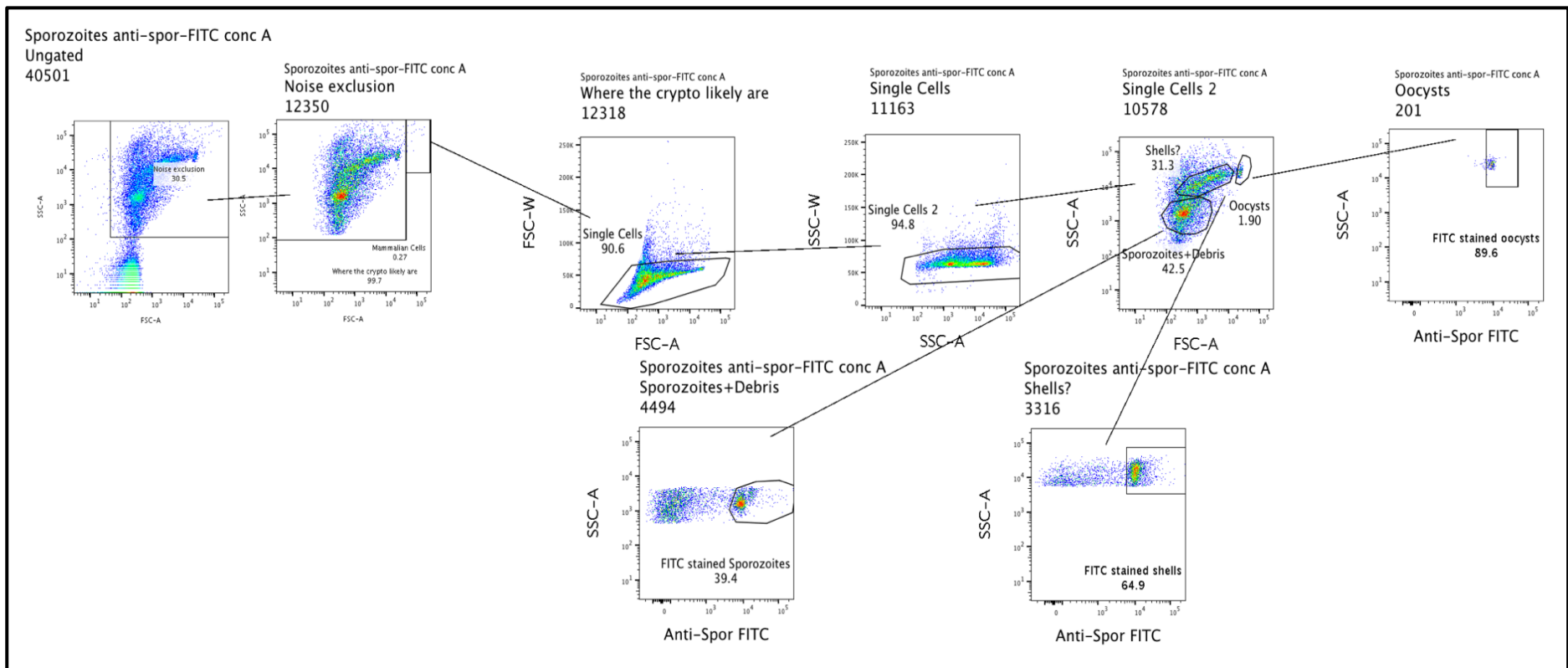
Figure D.5. Demonstrating gating hierarchy for anti-spor FITC-stained sporozoites.
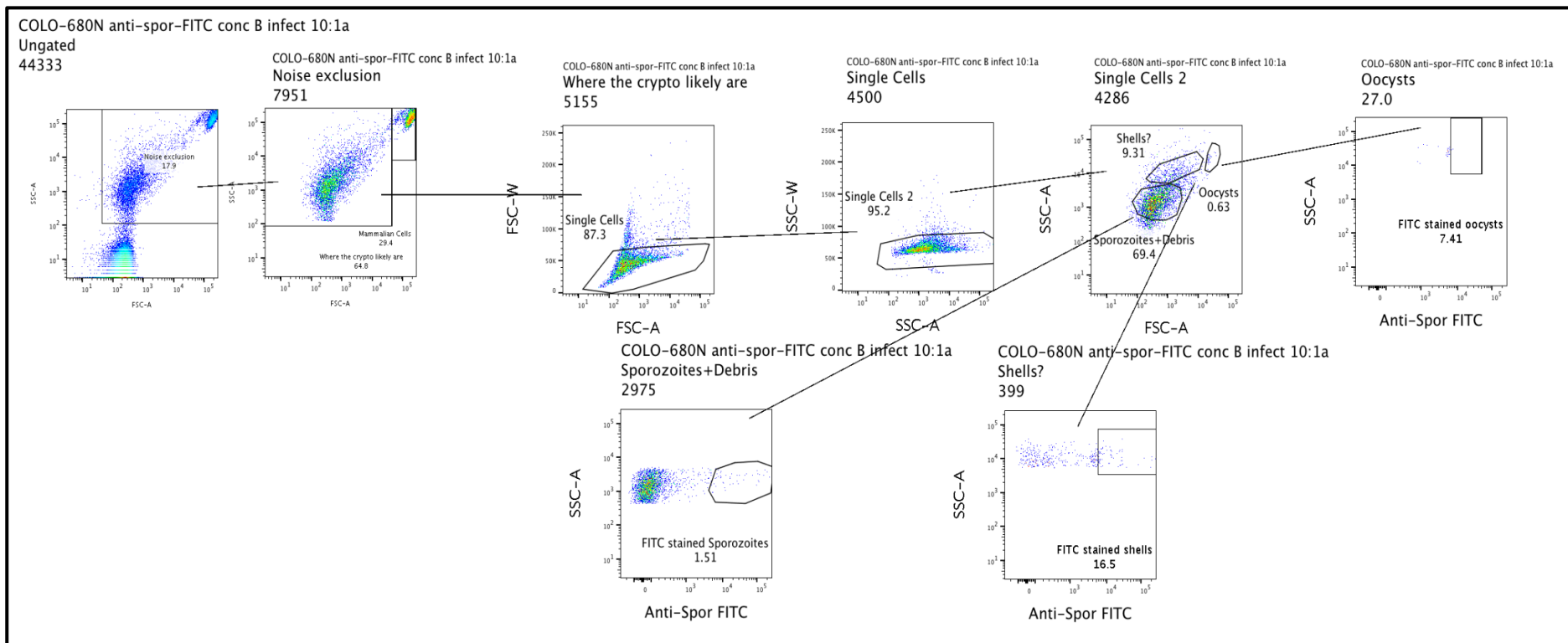
Figure D.6. Demonstrating gating hierarchy for infected cells.

Repeat experiments to verify sigM
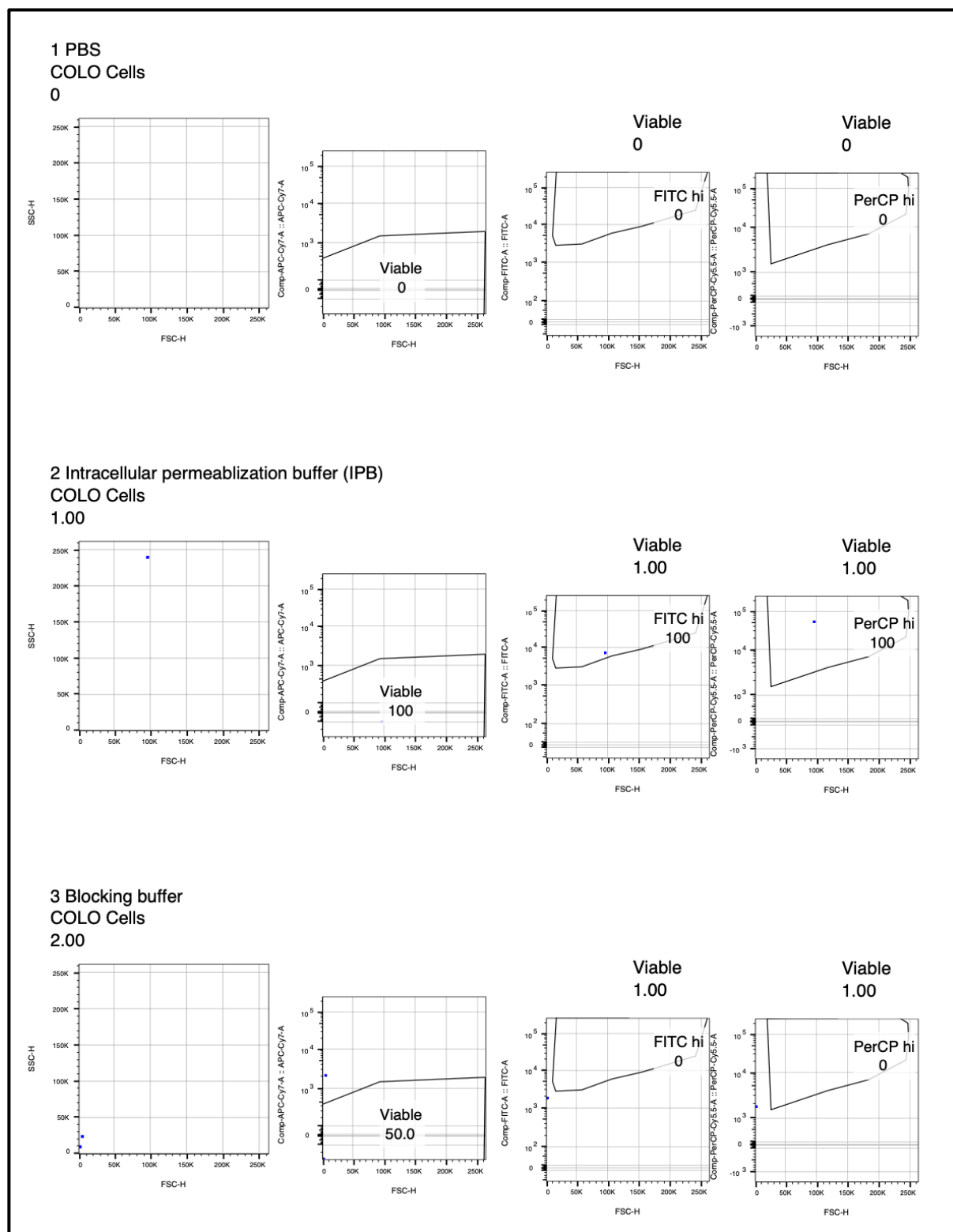
Figure D.7. Reagents show no significant populations expressing anti-spor FITC (FITC) or sigM (PerCP).
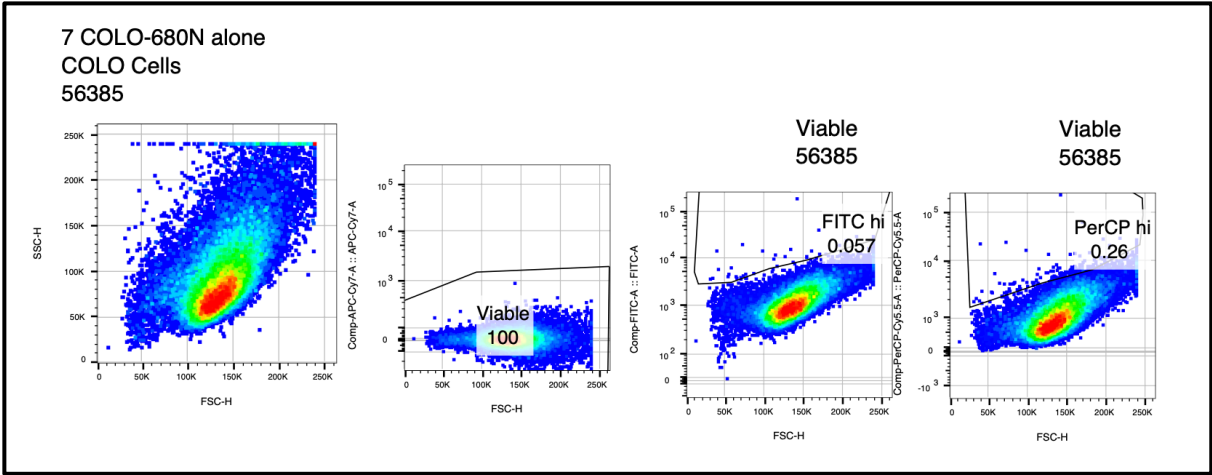
Figure D.8. Untreated cells do not express anti-spor FITC (FITC) or sigM (PerCP).
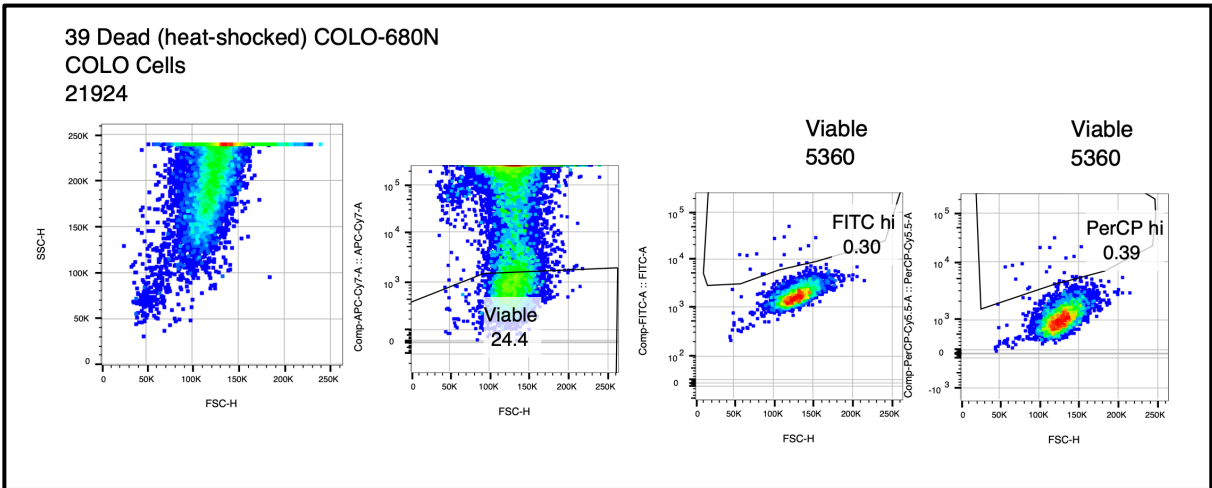
Figure D.9. Dead (heat-shocked) cells do not express anti-spor FITC (FITC) or sigM (PerCP).
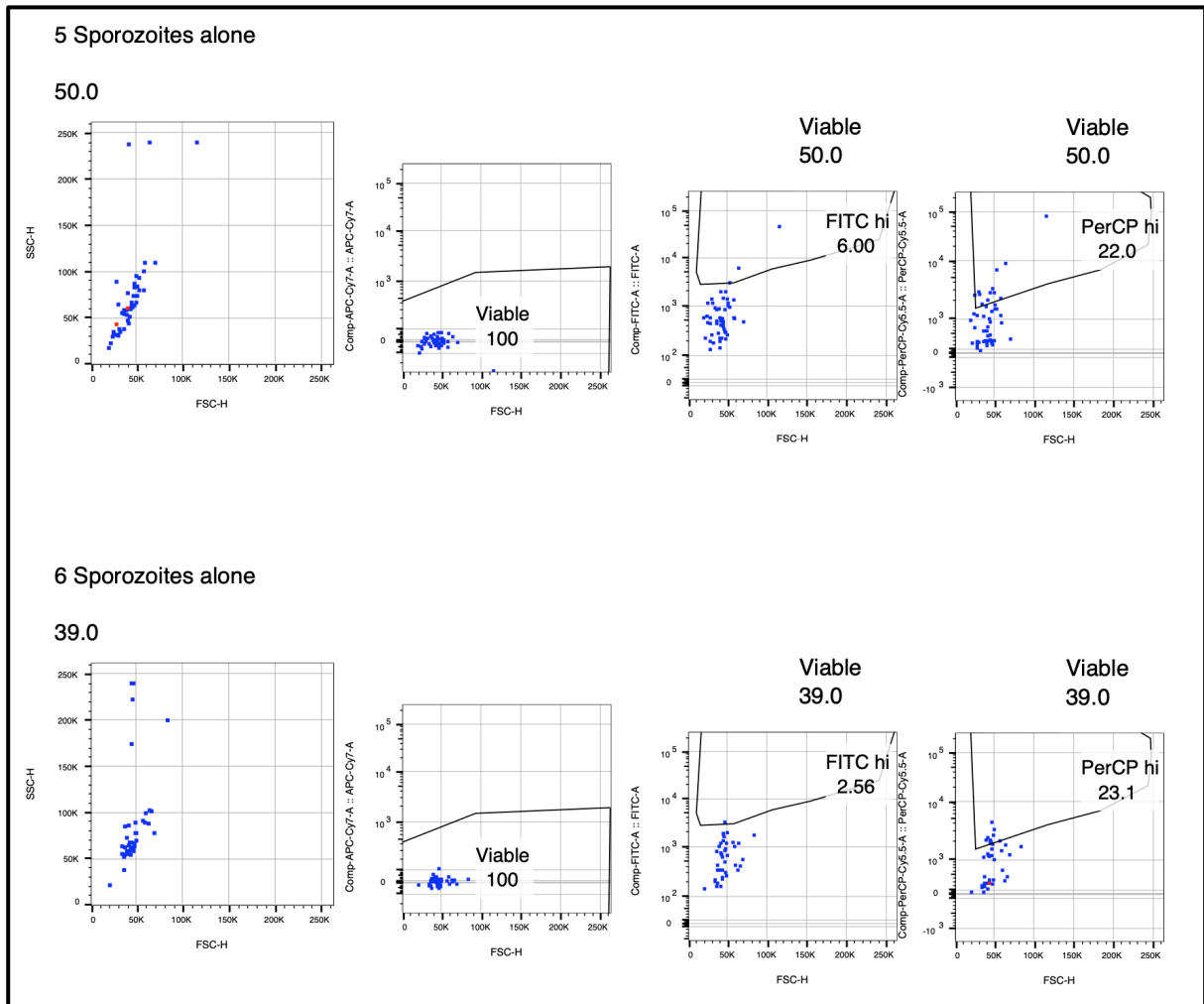
Figure D.10. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in sporozoite only samples. Sporozoites show higher expression of sigM than FITC.
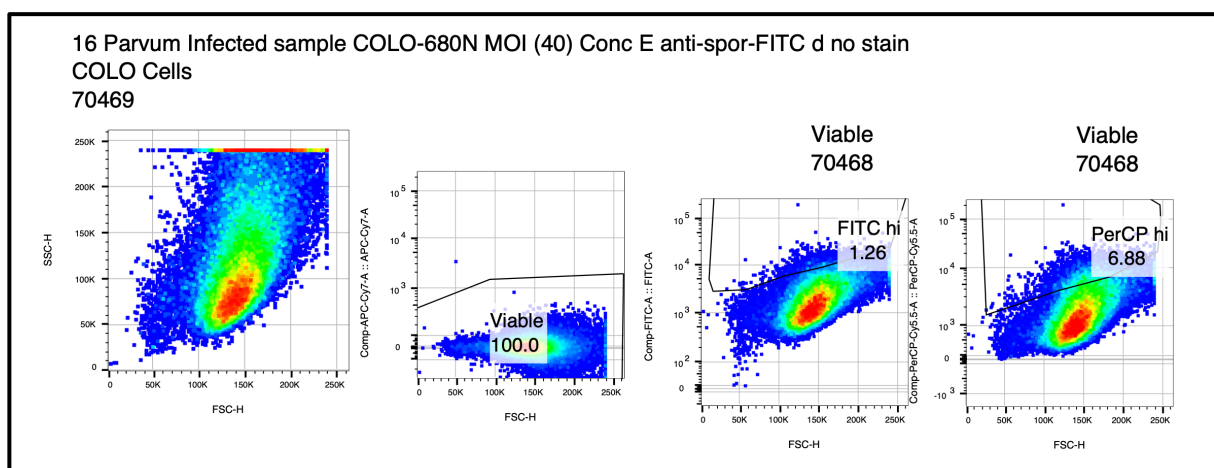
Figure D.11. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in unstained infected cells. Cells were infected with *C. parvum* at an MOI of 40.
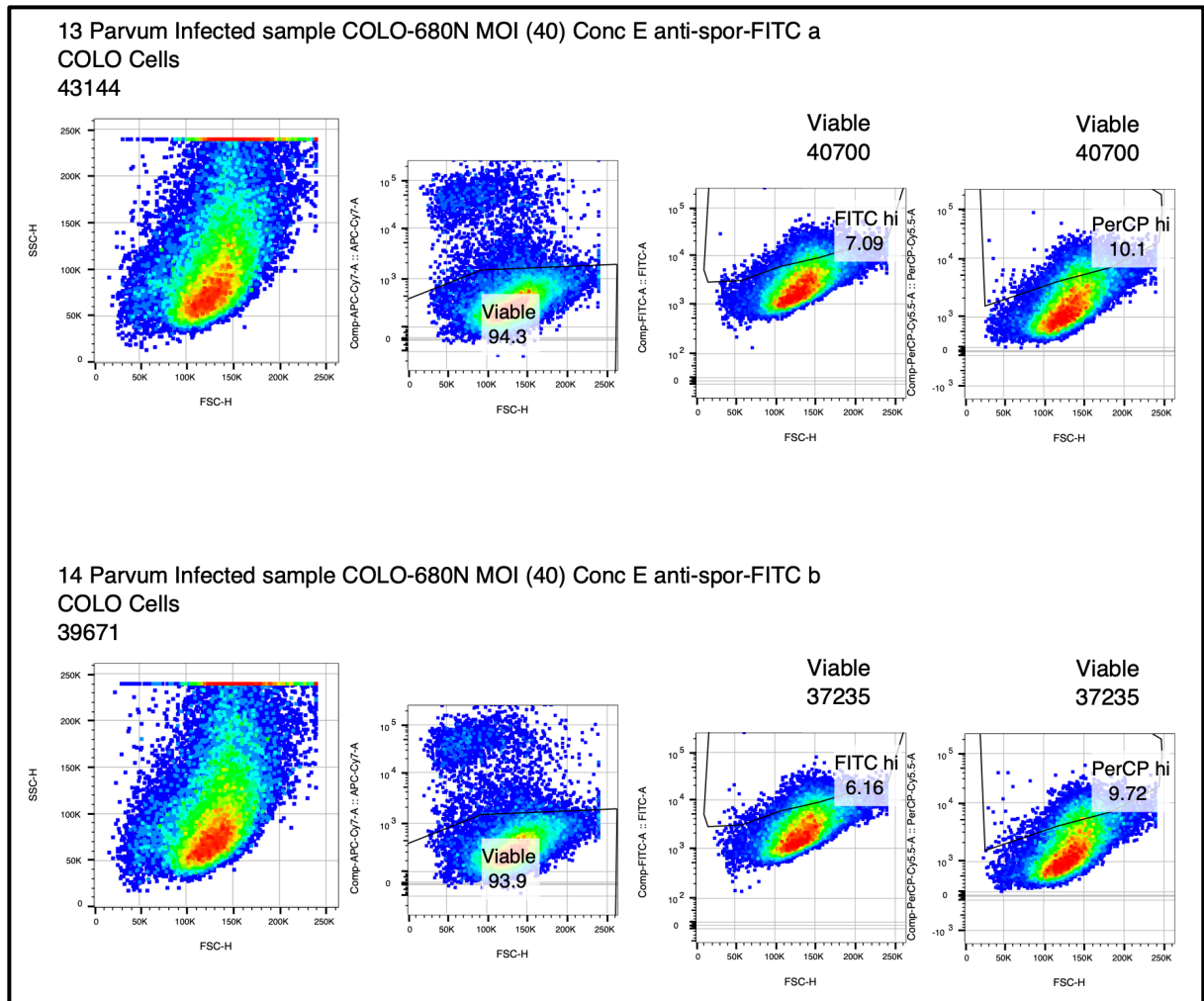
Figure D.12. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in infected cells. Cells were infected with *C. parvum* at an MOI of 40.
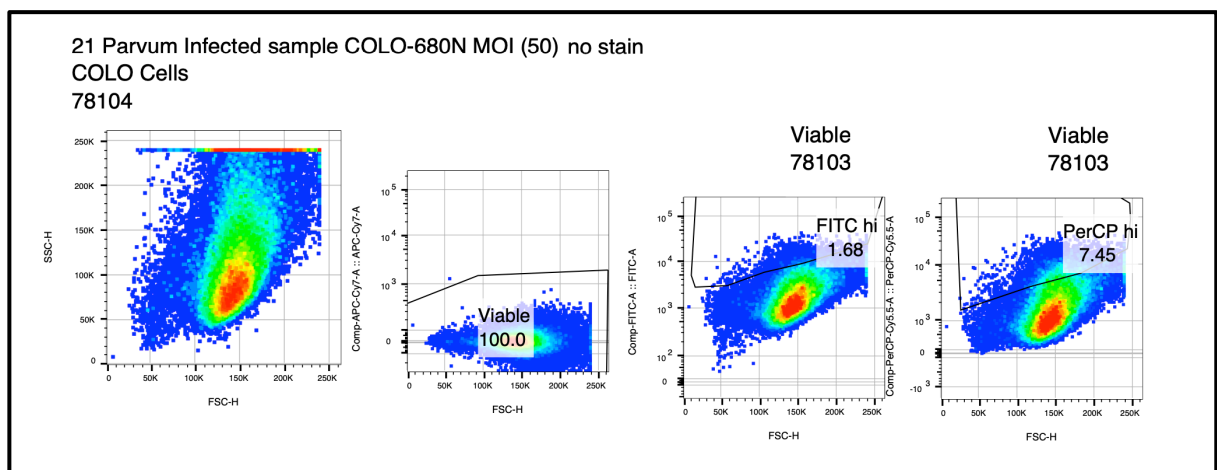
Figure D.13. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in unstained infected cells. Cells were infected with *C. parvum* at an MOI of 50.
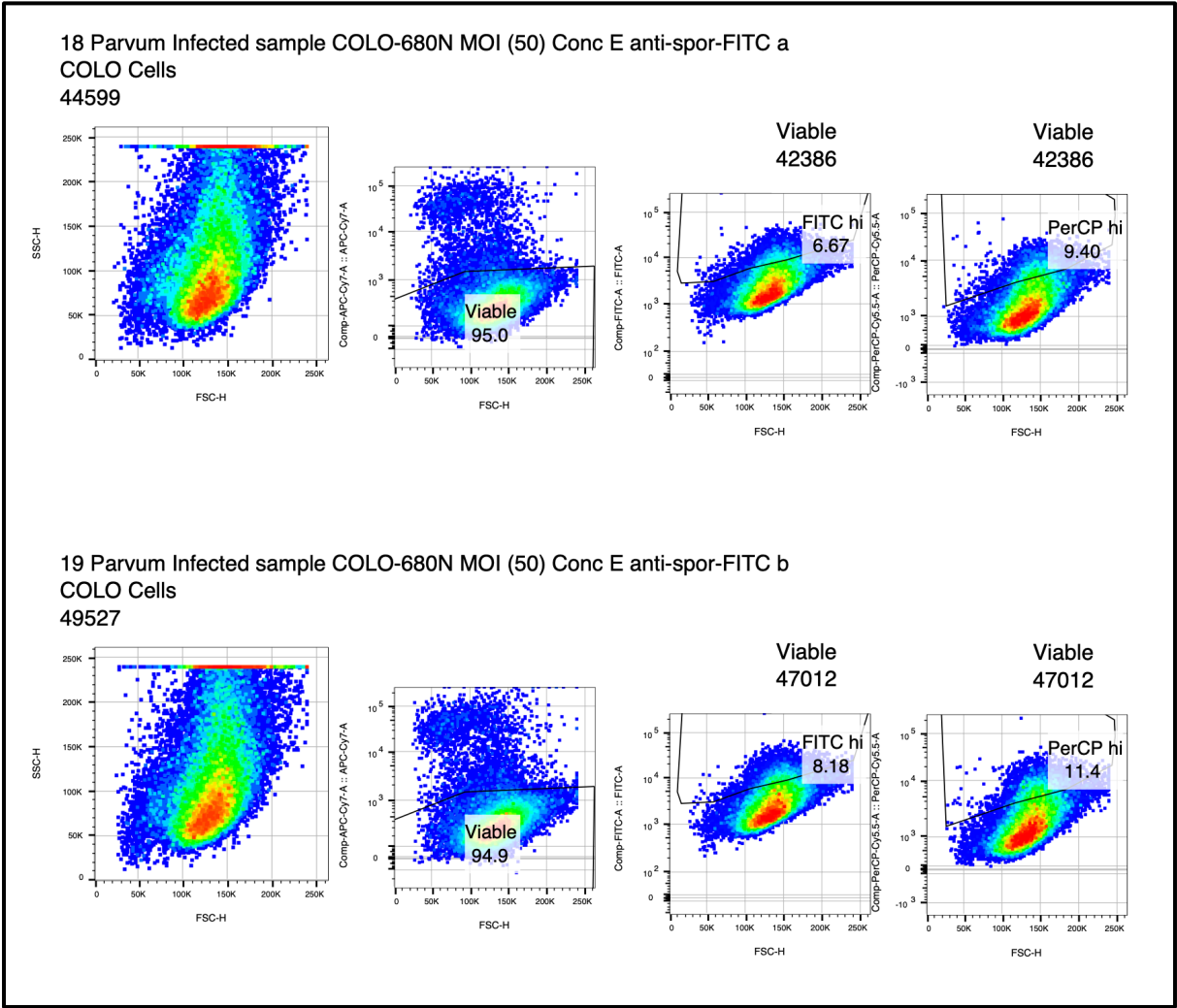
Figure D.14. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in infected cells. Cells were infected with *C. parvum* at an MOI of 50.
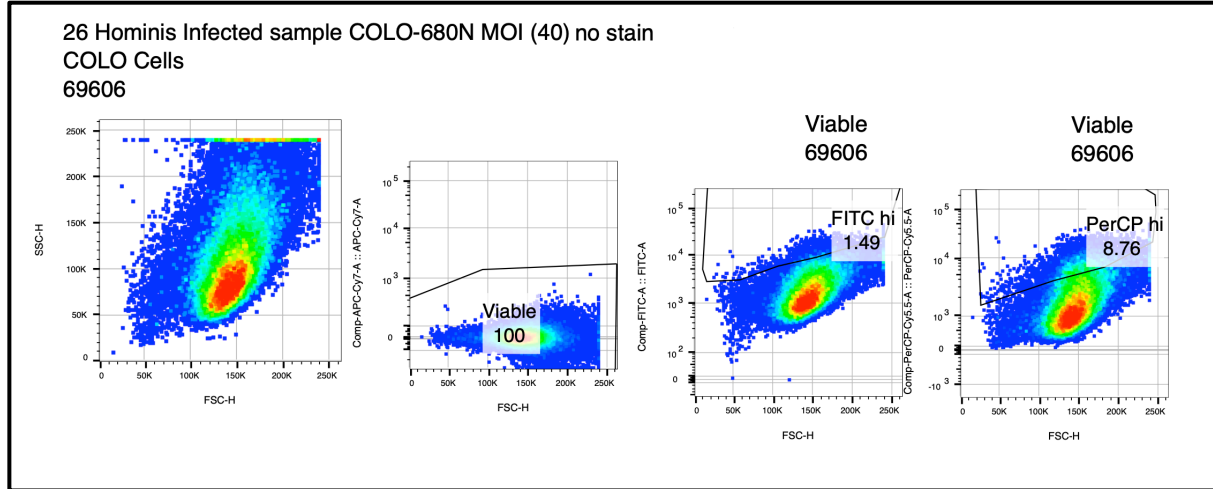
Figure D.15. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in unstained infected cells. Cells were infected with *C. hominis* at MOI 40.
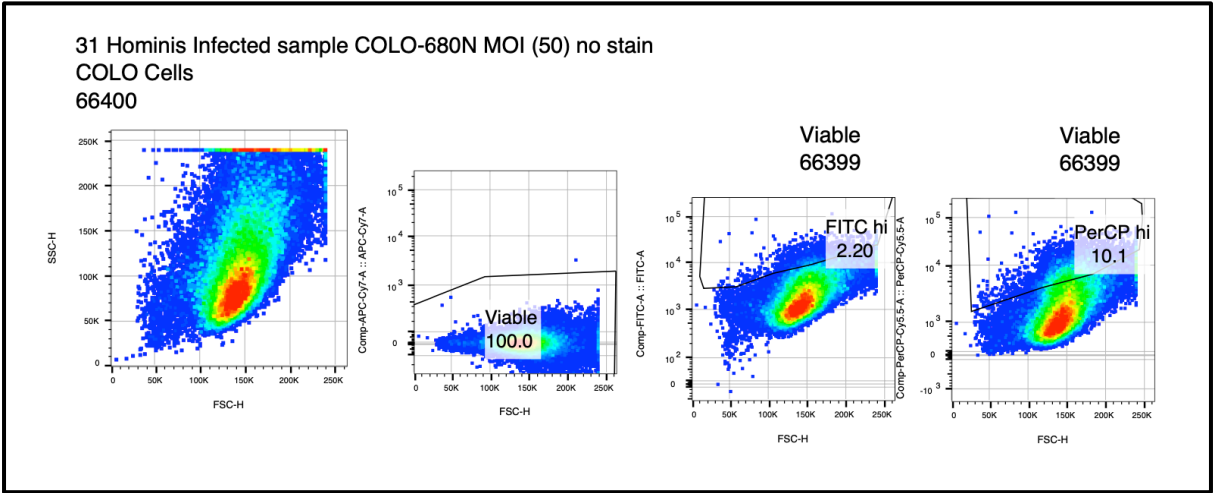
Figure D.16. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in unstained infected cells. Cells were infected with *C. hominis* at an MOI of 50.
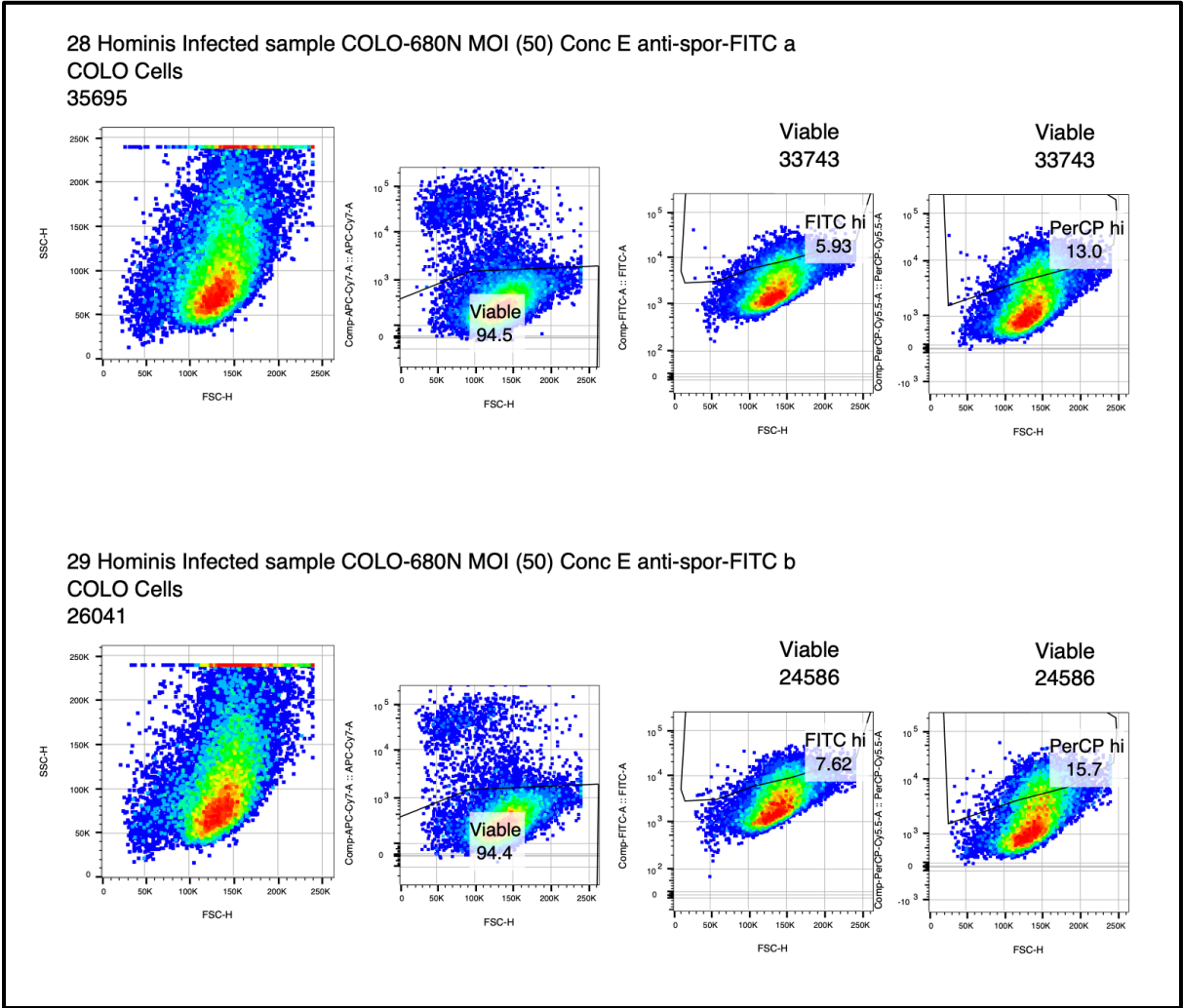
Figure D.17. Comparison of anti-spor FITC (FITC) and sigM (PerCP) in infected cells. Cells were infected with *C. hominis* at an MOI of 50.
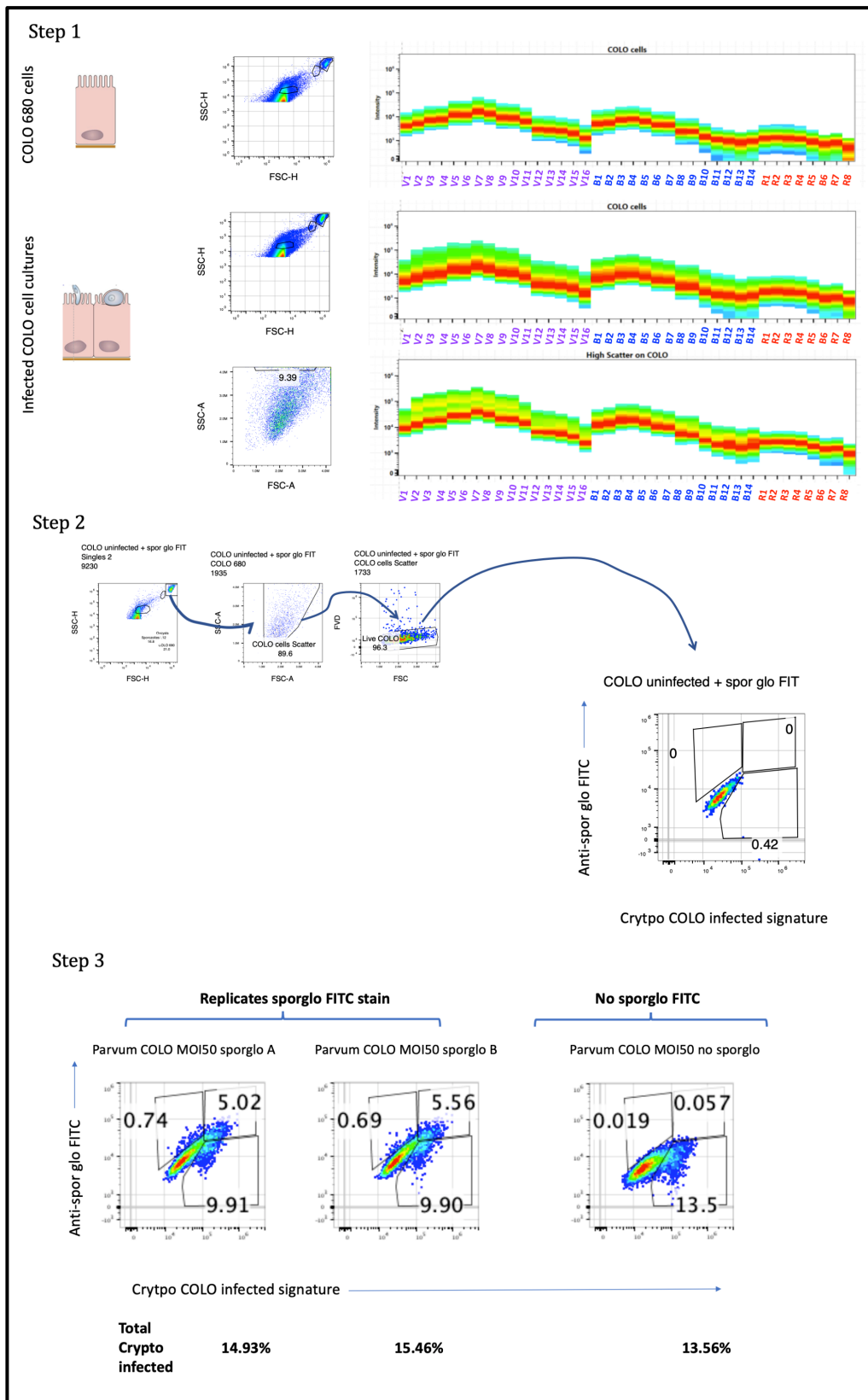
Figure D.18. Gating strategy for the detection of infected cell on Aurora flow cytometer.
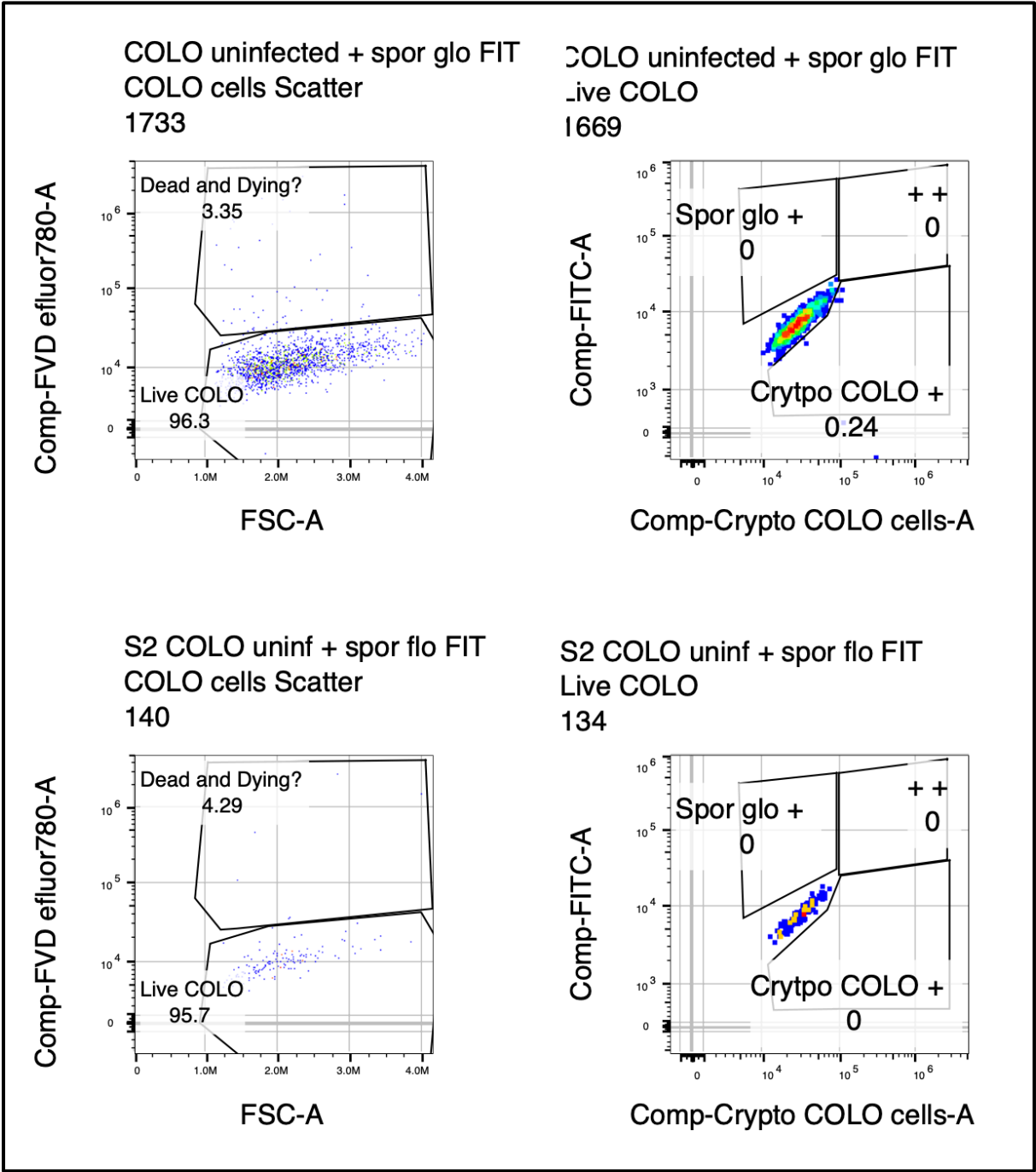
Figure D.19. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in uninfected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Viability displayed on the left panels, fluorophore expression on the right panels.
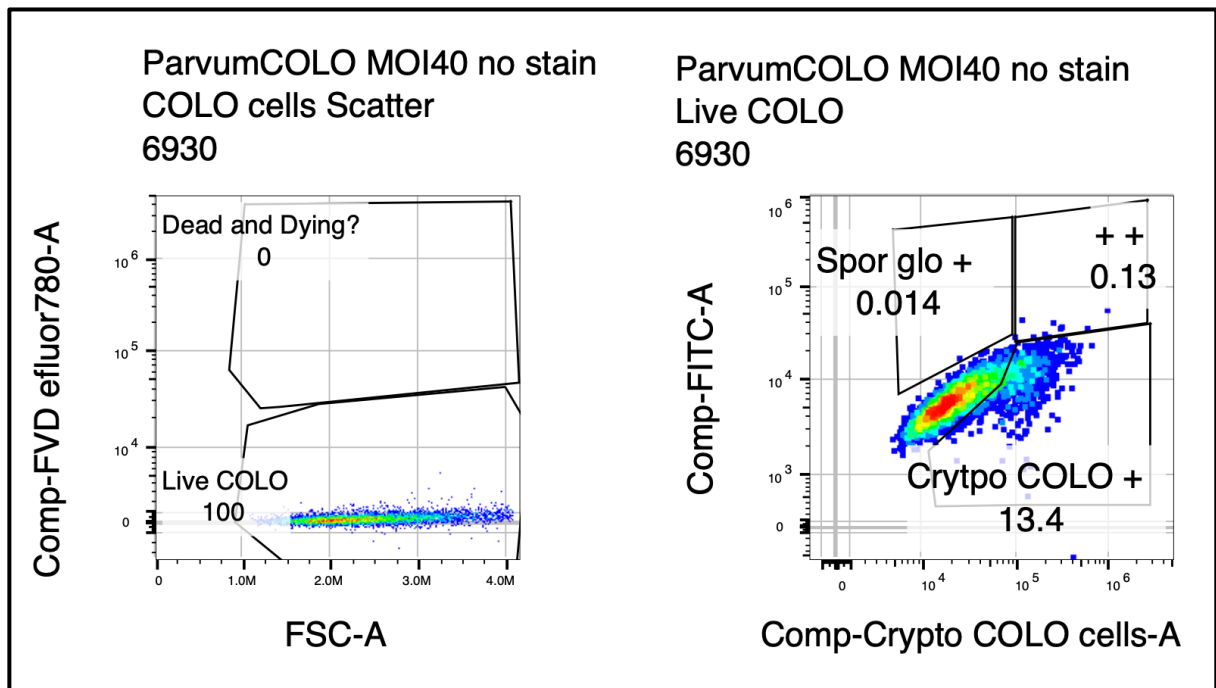
Figure D.20. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in unstained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. parvum* at an MOI of 40. Viability displayed on the left panels, fluorophore expression on the right panels.
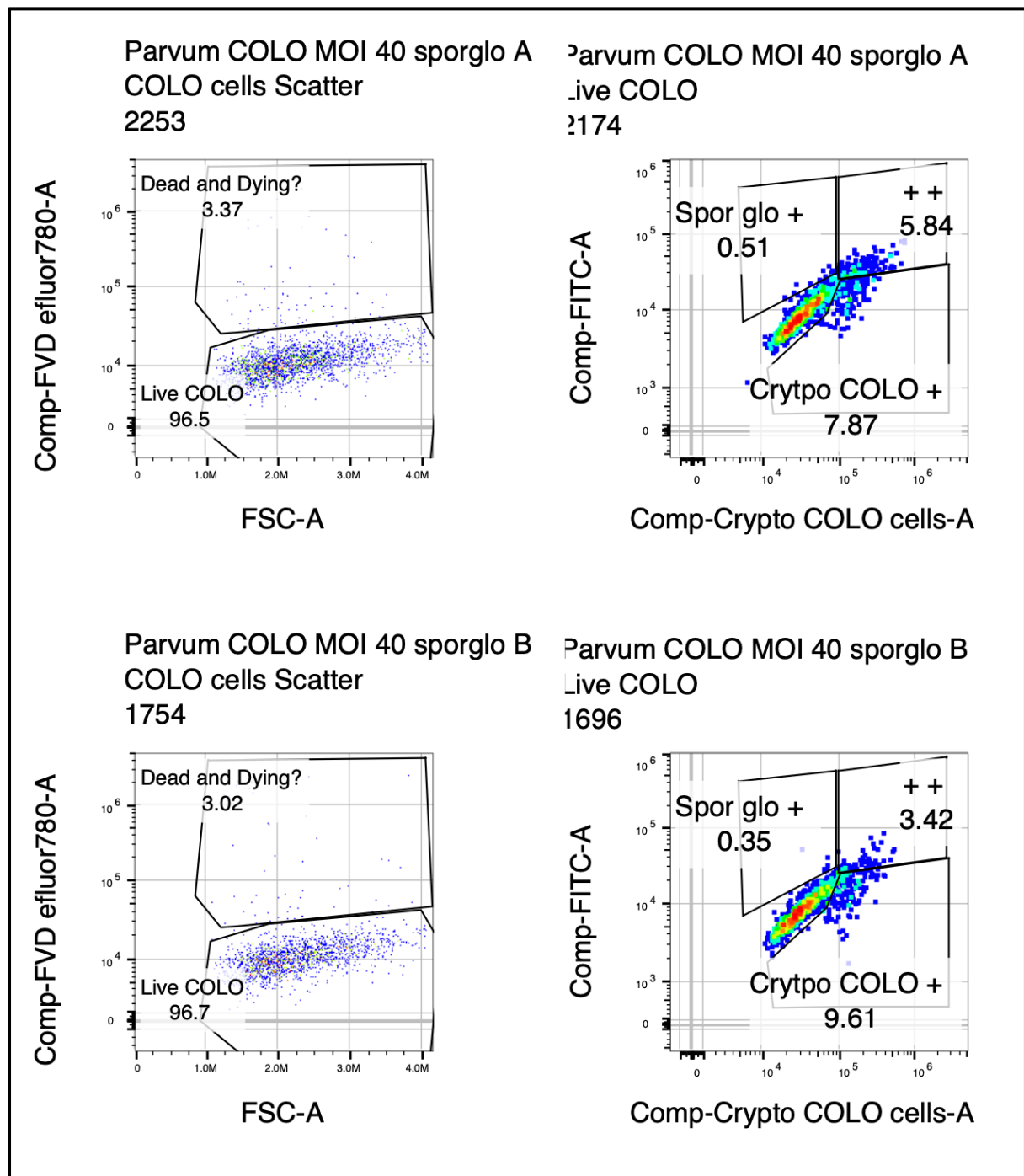
Figure D.21. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in stained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. parvum* at an MOI of 40. Viability displayed on the left panels, fluorophore expression on the right panels.
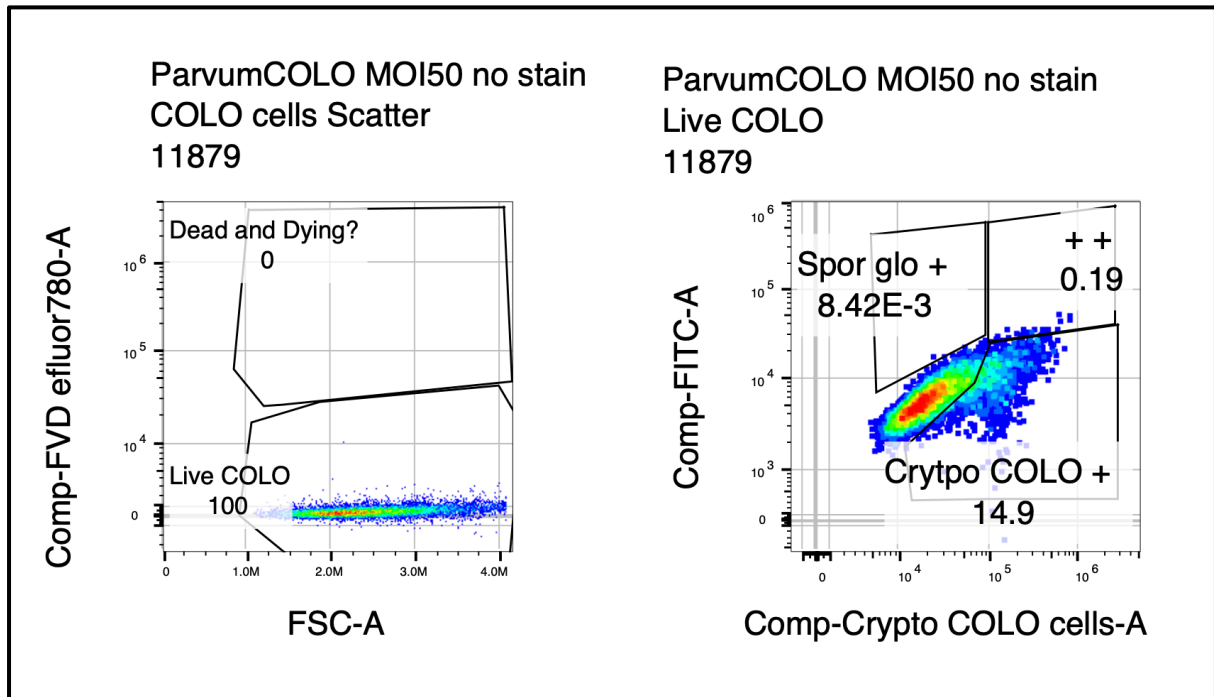
Figure D.22. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in unstained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. parvum* at an MOI of 50. Viability displayed on the left panels, fluorophore expression on the right panels.
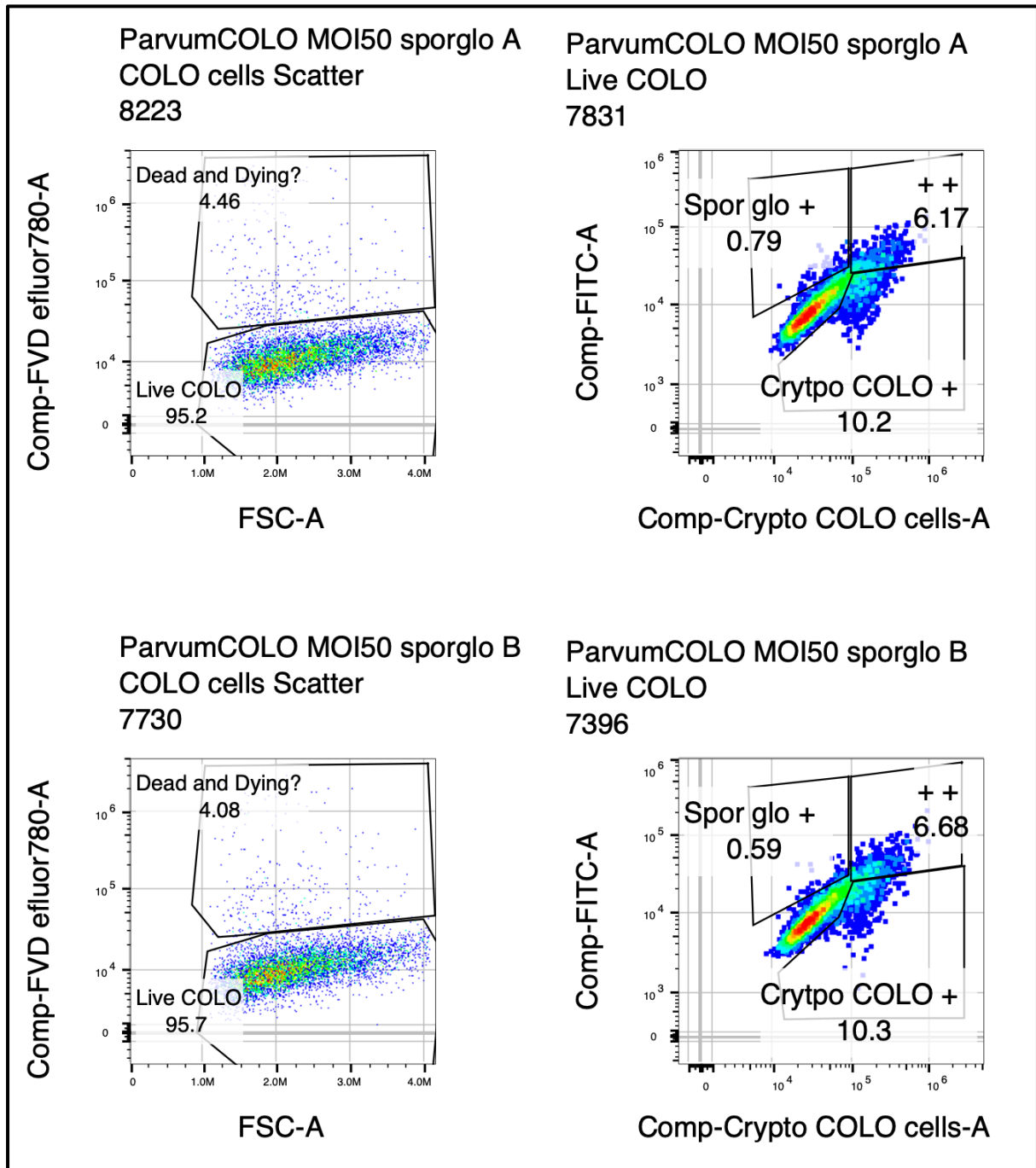
Figure D.23. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in stained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. parvum* at an MOI of 50. Viability displayed on the left panels, fluorophore expression on the right panels.
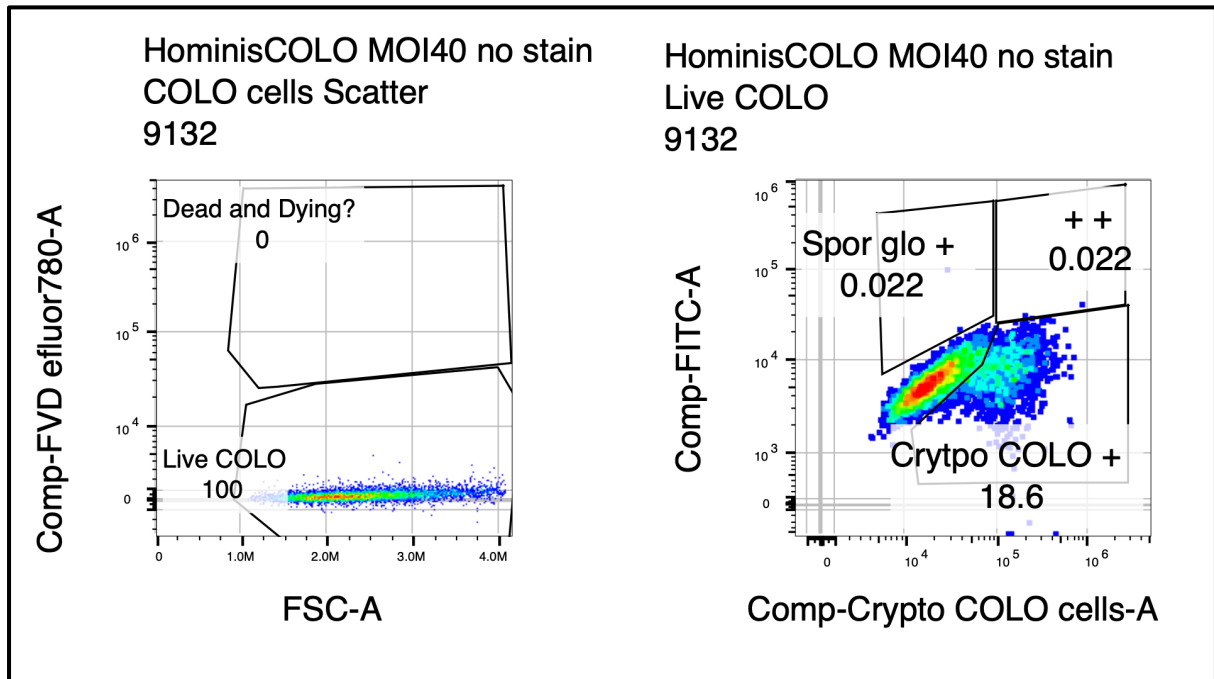
Figure D.24. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in unstained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. hominis* at an MOI of 40. Viability displayed on the left panels, fluorophore expression on the right panels.
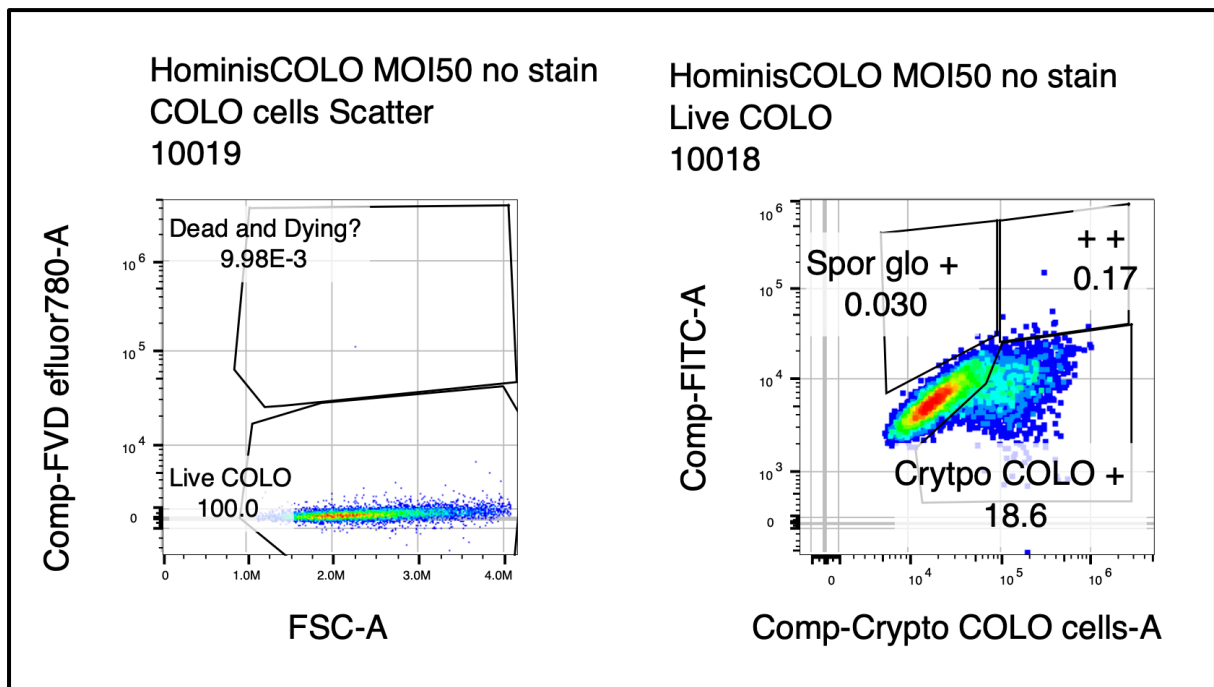
Figure D.25. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in unstained infected cells using the Aurora flow cytometer. (++) signifies cells expressing

both anti-spor FITC and sigM. Cells infected with *C. hominis* at an MOI of 50. Viability displayed on the left panels, fluorophore expression on the right panels.
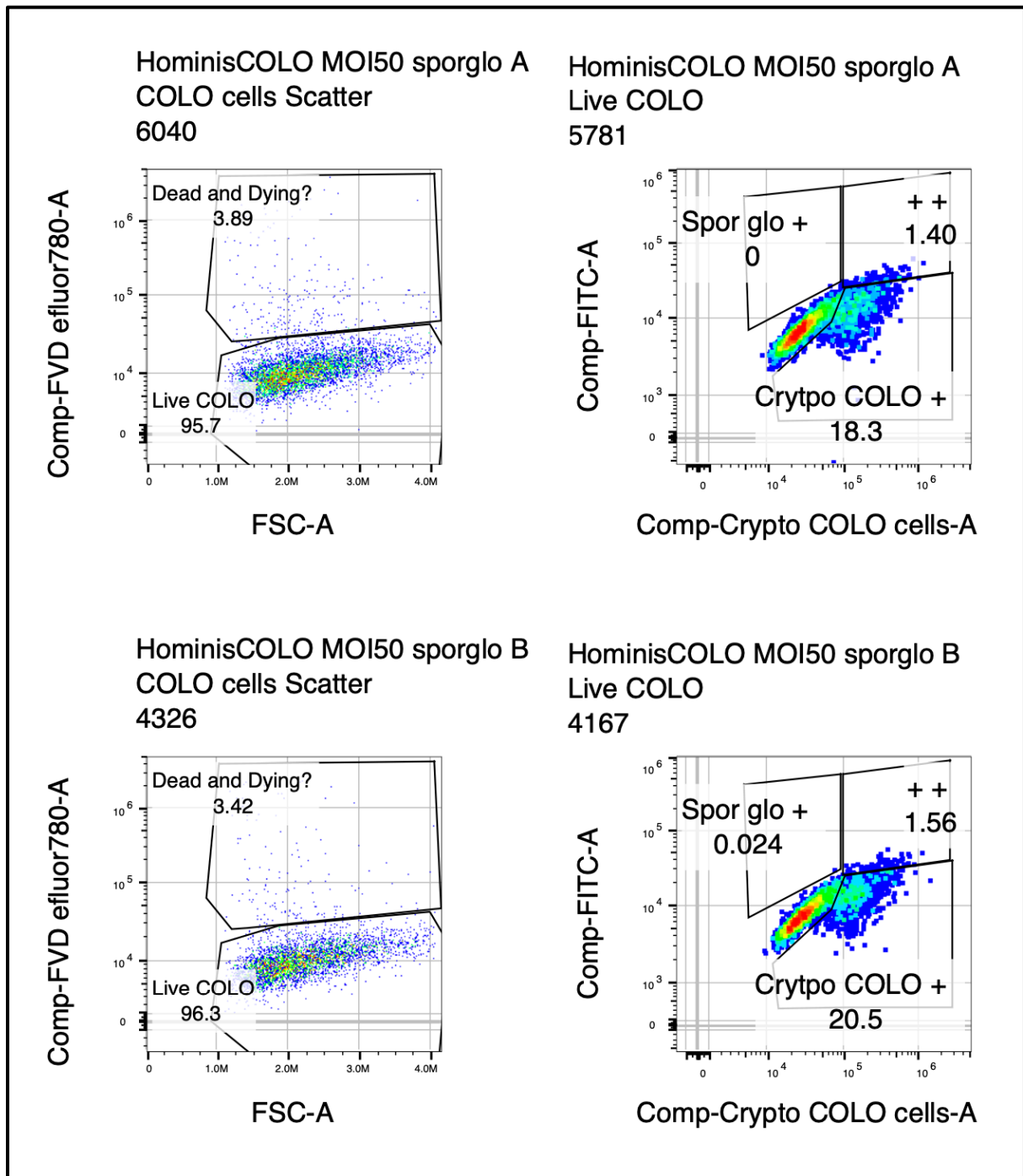
Figure D.26. Analysing anti-spor FITC (Spor glo +) and sigM (Crypto COLO +) expression in stained infected cells using the Aurora flow cytometer. (++) signifies cells expressing both anti-spor FITC and sigM. Cells infected with *C. hominis* at an MOI of 50. Viability displayed on the left panels, fluorophore expression on the right panels.
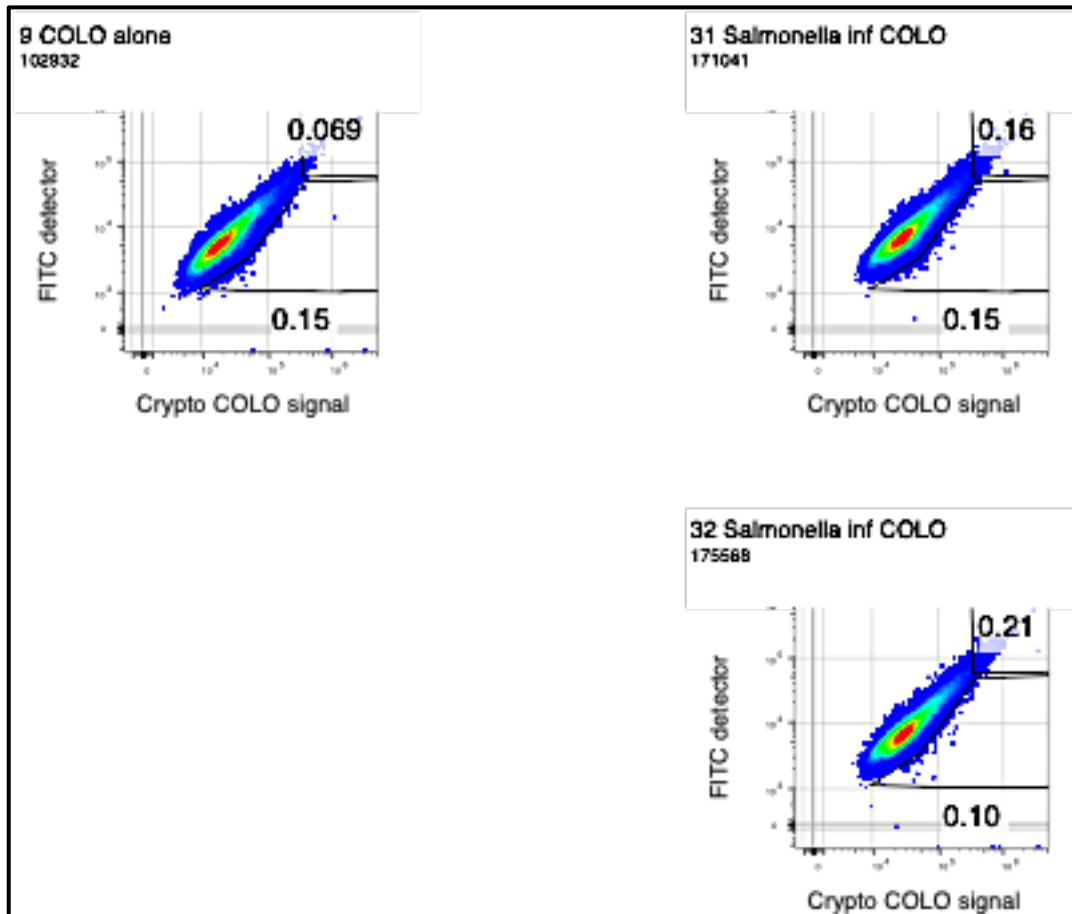
Figure D.27. Analysing anti-spor FITC (FITC) and sigM (Crypto COLO) expression in *S. typhimurium* infected cells using the Aurora flow cytometer. Control on the left panel, infected cells on the right panels. Infected cells show no significant expression of anti-spor FITC or sigM.

9.5 Appendix E

Supplementary material to Chapter 6.

Table E.1. List of target genes used in this study.

Gene ID	Accession	Flags	Species
cgd2_3270	XM_626517.1	Housekeeping Gene	*Cryptosporidium parvum*
cgd5_3160	XM_001388245.1	Housekeeping Gene	*Cryptosporidium parvum*
cgd6_2090	XM_627569.1	Housekeeping Gene	*Cryptosporidium parvum*
cgd6_4270	XM_627759.1	Housekeeping Gene	*Cryptosporidium parvum*
cgd_2_140	XM_625383.1		*Cryptosporidium parvum*
cgd1_1660	XM_628006.1		*Cryptosporidium parvum*
cgd1_2270	XM_628059.1		*Cryptosporidium parvum*
cgd1_2400	XM_628069.1		*Cryptosporidium parvum*
cgd1_2880	XM_628107.1		*Cryptosporidium parvum*
cgd1_300	XM_627890.1		*Cryptosporidium parvum*
cgd1_3810	XM_628192.1		*Cryptosporidium parvum*
cgd1_640	XM_627916.1		*Cryptosporidium parvum*
cgd1_880	XM_001388111.1		*Cryptosporidium parvum*
cgd1_910	XM_627937.1		*Cryptosporidium parvum*
cgd2_1070	XM_626335.1		*Cryptosporidium parvum*
cgd2_20	XM_625373.1		*Cryptosporidium parvum*
cgd2_200	XM_625389.1		*Cryptosporidium parvum*
cgd2_2200	XM_626425.1		*Cryptosporidium parvum*
cgd2_2540	XM_626456.1		*Cryptosporidium parvum*
cgd2_3000	XM_626494.1		*Cryptosporidium parvum*
cgd2_3070	XM_626501.1		*Cryptosporidium parvum*
cgd2_350	XM_625403.1		*Cryptosporidium parvum*
cgd2_3730	XM_626558.1		*Cryptosporidium parvum*
cgd2_430	XM_625410.1		*Cryptosporidium parvum*
cgd2_4320	XM_626611.1		*Cryptosporidium parvum*
cgd2_820	XM_626315.1		*Cryptosporidium parvum*
cgd3_1300	XM_626705.1		*Cryptosporidium parvum*
cgd3_1400	XM_626715.1		*Cryptosporidium parvum*
cgd3_1570	XM_001388065.1		*Cryptosporidium parvum*
cgd3_1940	XM_626765.1		*Cryptosporidium parvum*
cgd3_2250	XM_626792.1		*Cryptosporidium parvum*
cgd3_2940	XM_626851.1		*Cryptosporidium parvum*
cgd3_330	XM_001388055.1		*Cryptosporidium parvum*

Gene ID	Accession	Flags	Species
cgd3_3770	XM_626924.1		*Cryptosporidium parvum*
cgd3_3790	XM_626926.1		*Cryptosporidium parvum*
cgd3_3930	XM_626938.1		*Cryptosporidium parvum*
cgd3_4150	XM_001388094.1		*Cryptosporidium parvum*
cgd3_4260	XM_626969.1		*Cryptosporidium parvum*
cgd3_510	XM_626636.1		*Cryptosporidium parvum*
cgd3_760	XM_626659.1		*Cryptosporidium parvum*
cgd4_1940	XM_625788.1		*Cryptosporidium parvum*
cgd4_2260	XM_625816.1		*Cryptosporidium parvum*
cgd4_3080	XM_625892.1		*Cryptosporidium parvum*
cgd4_3160	XM_625900.1		*Cryptosporidium parvum*
cgd4_3550	XM_625931.1		*Cryptosporidium parvum*
cgd4_3620	XM_625938.1		*Cryptosporidium parvum*
cgd4_4310	XM_625446.1		*Cryptosporidium parvum*
cgd4_4460	XM_625460.1		*Cryptosporidium parvum*
cgd5_10	XM_625966.1		*Cryptosporidium parvum*
cgd5_1470	XM_626094.1		*Cryptosporidium parvum*
cgd5_2370	XM_626174.1		*Cryptosporidium parvum*
cgd5_2800	XM_626216.1		*Cryptosporidium parvum*
cgd5_3040	XM_626236.1		*Cryptosporidium parvum*
cgd5_4590	XM_625307.1		*Cryptosporidium parvum*
cgd6_10	XM_625341.1		*Cryptosporidium parvum*
cgd6_1070	XM_627479.1		*Cryptosporidium parvum*
cgd6_1080	XM_627480.1		*Cryptosporidium parvum*
cgd6_2170	XM_627575.1		*Cryptosporidium parvum*
cgd6_2330	XM_627590.1		*Cryptosporidium parvum*
cgd6_2600	XM_627613.1		*Cryptosporidium parvum*
cgd6_3010	XM_001388297.1		*Cryptosporidium parvum*
cgd6_3800	XM_627719.1		*Cryptosporidium parvum*
cgd6_3850	XM_627723.1		*Cryptosporidium parvum*
cgd6_3920	XM_627729.1		*Cryptosporidium parvum*
cgd6_3990	XM_001388307.1		*Cryptosporidium parvum*
cgd6_40	XM_625344.1		*Cryptosporidium parvum*
cgd6_4190	XM_627752.1		*Cryptosporidium parvum*
cgd6_4620	XM_627791.1		*Cryptosporidium parvum*
cgd6_4630	XM_627792.1		*Cryptosporidium parvum*
cgd6_4860	XM_627813.1		*Cryptosporidium parvum*
cgd6_4910	XM_627816.1		*Cryptosporidium parvum*
cgd6_5400	XM_627858.1		*Cryptosporidium parvum*
cgd6_5410	XM_627859.1		*Cryptosporidium parvum*
cgd7_120	XM_001388325.1		*Cryptosporidium parvum*
cgd7_130	XM_628206.1		*Cryptosporidium parvum*
cgd7_1460	XM_628328.1		*Cryptosporidium parvum*
cgd7_1830	XM_628358.1		*Cryptosporidium parvum*

Gene ID	Accession	Flags	Species
cgd7_2110	XM_628385.1		*Cryptosporidium parvum*
cgd7_2250	XM_628397.1		*Cryptosporidium parvum*
cgd7_2280	XM_628400.1		*Cryptosporidium parvum*
cgd7_2300	XM_628401.1		*Cryptosporidium parvum*
cgd7_2430	XM_628414.1		*Cryptosporidium parvum*
cgd7_2540	XM_628424.1		*Cryptosporidium parvum*
cgd7_3010	XM_628466.1		*Cryptosporidium parvum*
cgd7_3020	XM_628467.1		*Cryptosporidium parvum*
cgd7_3240	XM_001388357.1		*Cryptosporidium parvum*
cgd7_3250	XM_628487.1		*Cryptosporidium parvum*
cgd7_4050	XM_628560.1		*Cryptosporidium parvum*
cgd7_480	XM_628236.1		*Cryptosporidium parvum*
cgd7_700	XM_001388332.1		*Cryptosporidium parvum*
cgd7_790	XM_628266.1		*Cryptosporidium parvum*
cgd8_1840	XM_001388389.1		*Cryptosporidium parvum*
cgd8_2340	XM_627137.1		*Cryptosporidium parvum*
cgd8_3230	XM_627219.1		*Cryptosporidium parvum*
cgd8_3390	XM_627234.1		*Cryptosporidium parvum*
cgd8_3460	XM_627239.1		*Cryptosporidium parvum*
cgd8_3480	XM_627241.1		*Cryptosporidium parvum*
cgd8_3520	XM_627245.1		*Cryptosporidium parvum*
cgd8_3770	XM_627269.1		*Cryptosporidium parvum*
cgd8_4050	XM_627296.1		*Cryptosporidium parvum*
cgd8_4220	XM_627312.1		*Cryptosporidium parvum*
cgd8_4320	XM_627320.1		*Cryptosporidium parvum*
cgd8_4500	XM_627336.1		*Cryptosporidium parvum*
cgd8_60	XM_625470.1		*Cryptosporidium parvum*
ARG1	NM_000045.3		*Homo sapiens*
ARG2	NM_001172.3		*Homo sapiens*
BCL2	NM_000657.2		*Homo sapiens*
CAMP	NM_004345.3		*Homo sapiens*
CASP1	NM_001223.3		*Homo sapiens*
CASP3	NM_004346.3		*Homo sapiens*
CASP4	NM_001225.3		*Homo sapiens*
CASP5	NM_004347.1		*Homo sapiens*
CCL5	NM_002985.2		*Homo sapiens*
CD36	NM_000072.3		*Homo sapiens*
CISH	NM_145071.2		*Homo sapiens*
CX3CL1	NM_002996.3		*Homo sapiens*
CXCL8	NM_000584.2		*Homo sapiens*
DEFA1	NM_004084.2		*Homo sapiens*
DEFB1	NM_005218.3		*Homo sapiens*
DEFB4B	NM_001205266.1		*Homo sapiens*
FAS	NM_000043.4		*Homo sapiens*

Gene ID	Accession	Flags	Species
FASLG	NM_000639.1		*Homo sapiens*
GUSB	NM_000181.3		*Homo sapiens*
HPRT1	NM_000194.1		*Homo sapiens*
ICAM1	NM_000201.2		*Homo sapiens*
IFNG	NM_000619.2		*Homo sapiens*
IL18	NM_001562.3		*Homo sapiens*
IL1B	NM_000576.2		*Homo sapiens*
IL1R1	NM_001320984.1		*Homo sapiens*
IL33	NM_033439.2		*Homo sapiens*
IL6	NM_000600.3		*Homo sapiens*
NFKB1	NM_003998.2		*Homo sapiens*
PTGS2	NM_000963.1		*Homo sapiens*
PYCARD	NM_013258.3		*Homo sapiens*
RELA	NM_021975.2		*Homo sapiens*
RPL19	NM_000981.3		*Homo sapiens*
SDHA	NM_004168.1		*Homo sapiens*
SIRT1	NM_012238.4		*Homo sapiens*
SOCS1	NM_003745.1		*Homo sapiens*
SOCS2	NM_003877.3		*Homo sapiens*
SOCS3	NM_003955.3		*Homo sapiens*
SOCS4	NM_199421.1		*Homo sapiens*
STAT6	NM_003153.3		*Homo sapiens*
TRAF2	NM_021138.3		*Homo sapiens*

Table E.2. Fifty genes most highly expressed in oocysts/sporozoites from Matos et al., (2019). (*) represents the number of orthologues within eupathDB genomes. (**) represents the number of OrthoMCL orthologous sequences.

Gene ID	Product Description	Ortholog Group	Ortholog count Eupath DB*	Ortholog count total**
cgd2_140	Stress-associated_endoplasmic_reticulum_protein	OG5_130196	5	58
cgd1_2400	Uncharacterized_secreted_protein	OG5_222471	8	2
cgd1_3810	Uncharacterized_protein	OG5_194982	9	3
cgd1_640	Signal_peptide_region_containing_protein	OG5_194890	9	3
cgd1_880	Eukaryotic_initiation_factor_4A	OG5_126984	15	154
cgd1_910	AN1-type_and_A20-type_Zinc_finger	OG5_127624	13	112
cgd2_200	Uncharacterized_protein	OG5_130876	12	50
cgd2_2540	Glutaredoxin	OG5_126864	17	172
cgd2_3070	High_mobility_group_box_domain_containing_protein	OG5_126740	39	202
cgd2_3730	Glutathione_S-transferase_C-terminal_domain_containing_protein	OG5_180098	9	4
cgd2_430	Signal_peptide_containing_protein	OG5_222485	6	2
cgd2_4320	Thioredoxin/glutathione_reductase_selenoprotein	OG5_126785	23	189
cgd2_820	Translation_initiation_factor_SUI1	OG5_127384	12	124
cgd3_1570	Profilin	OG5_141746	11	14
cgd3_1940	Translationally_controlled_tumour_protein_associated_protein	OG5_127617	11	112
cgd3_330	unspecified product	OG5_129031	11	77
cgd3_3770	_Hsp90	OG5_126623	17	314
cgd3_4150	RNA_recognition_motif_domain_containing_protein	OG5_129225	10	73
cgd3_510	GDP-fucose_transporter	OG5_188054	8	3
cgd3_760	Ribosomal_protein_L7Ae/L30e/S12e/Gadd45	OG5_128156	12	97

Gene ID	Product Description	Ortholog Group	Ortholog count EupathDB*	Ortholog count total**
cgd4_3550	Secreted_Kazal_domain-containing_protein	OG5_129460	24	69
cgd4_3620	Immunodominant_antigen_23393226	OG5_222616	7	2
cgd4_4310	Ring_finger_domain_containing_protein	OG5_191746	7	3
cgd5_10	Uncharacterized_Secreted_Protein	OG5_222658	7	2
cgd5_1470	Nucleoside_diphosphate_kinase	OG5_126708	22	214
cgd5_2800	Actin_depolymerizing_factor	OG5_127118	12	143
cgd6_10	Uncharacterized_protein	OG5_222668	7	2
cgd6_2600	Uncharacterized_protein	OG5_195457	9	3
cgd6_3010	Polyadenylate-binding_protein	OG5_126795	15	187
cgd6_3850	50S_ribosomal_protein_L30e-like	OG5_127322	15	127
cgd6_3920	Uncharacterized_protein_with_Tetratricopeptide-like_helical	OG5_175069	9	4
cgd6_40	Uncharacterized_protein	OG5_164335	7	6
cgd6_4860	DEAD/DEAH_box_helicase_with_GUCT_domain	OG5_128505	15	89
cgd6_4910	Zinc_finger_C3H1-type_domain_containing_protein	OG5_190023	11	3
cgd6_5400	Uncharacterized_Secreted_Protein	OG5_214933	6	2
cgd6_5410	Uncharacterized_Secreted_Protein	OG5_222709	7	2
cgd7_2300	Translation_elongation_factor_IF5A	OG5_126922	13	162
cgd7_2430	Translation_initiation_factor_eIF-5_Tif5p/ZnR+W2_domain-containing_protein	OG5_127779	13	106
cgd7_3010	Uncharacterized_protein	OG5_195622	9	3
cgd7_3240	RNA_polymerase_archaeal_subunit_P/eukaryotic_subunit_RPABC4	OG5_130124	7	59
cgd7_3250	Uncharacterized_Protein	OG5_195630	9	3
cgd7_480	L-lactate/malate_dehydrogenase	OG5_126911	29	164
cgd7_790	Uncharacterized_protein	OG5_169153	6	5
cgd8_3230	AP2/ERF_domain_containing_protein	OG5_195785	7	3

Gene ID	Product Description	Ortholog Group	Ortholog count Eupath DB*	Ortholog count total**
cgd8_3480	60S_ribosomal_protein_L34	OG5_127025	12	151
cgd8_3520	Uncharacterized_Secreted_Protein	OG5_135537	9	23
cgd8_3770	DNAJ_like_chaperone	OG5_126883	17	168
cgd8_4220	High_mobility_group_box_domain_containing_protein	OG5_144330	10	12
cgd8_4320	Uncharacterized_Protein	OG5_195816	9	3
cgd8_60	Uncharacterized_Protein	OG5_222757	7	2

Table E.3. Fifty genes most highly expressed in trophozoites/meronts from Matos et al., (2019). (*) represents the number of orthologues within eupathDB genomes. (**) represents the number of OrthoMCL orthologous sequences.

Gene ID	Product Description	Ortholog Group	Ortholog count euk pathogens*	Ortholog count total**
cgd1_1660	60S_ribosomal_protein_L36	OG5_127134	12	141
cgd1_2270	40S_ribosomal_protein_S26	OG5_127138	12	141
cgd1_2880	Uncharacterized_protein	OG5_cpar|cgd1_2880	3	
cgd1_300	40S_ribosomal_protein_S21	OG5_127484	11	118
cgd1_3810	Uncharacterized_protein	OG5_194982	9	3
cgd2_1070	40S_ribosomal_protein_S25	OG5_126998	12	153
cgd2_200	Uncharacterized_protein	OG5_130876	12	50
cgd2_2200	Ribosomal_protein_L37	OG5_126918	12	163
cgd2_3000	40S_ribosomal_protein_S16	OG5_126930	12	161
cgd2_350	60S_ribosomal_protein_L39	OG5_127376	6	125
cgd2_820	Translation_initiation_factor_SUI1	OG5_127384	12	124
cgd3_1300	_60S_ribosomal_protein_L12	OG5_127022	12	151
cgd3_1570	Profilin	OG5_141746	11	14
cgd3_2250	60S_ribosomal_protein_L37A	OG5_127166	11	139
cgd3_3790	60S_ribosomal_protein_L19	OG5_126940	11	160
cgd3_3930	60S_ribosomal_protein_L27A	OG5_127016	11	152

Gene ID	Product Description	Ortholog Group	Ortholog count euk pathogens*	Ortholog count total**
cgd3_510	GDP-fucose_transporter	OG5_188054	8	3
cgd4_2260	_60S_acidic_ribosomal_protein_P0	OG5_127051	12	149
cgd4_3080	Ribosomal_protein_S27a_with_Zinc-binding_domain/ubiquitin/Zinc-binding_domain	OG5_127221	14	135
cgd4_3160	40S_ribosomal_protein_S3a	OG5_126852	12	175
cgd4_3620	Immunodominant_antigen_23393226	OG5_222616	7	2
cgd5_1470	Nucleoside_diphosphate_kinase	OG5_126708	22	214
cgd5_2370	60S_acidic_ribosomal_protein_L12/LP1-like_protein	OG5_126872	13	170
cgd5_3040	40S_ribosomal_protein_S7	OG5_127368	12	125
cgd5_3160	Actin	OG5_126595	23	421
cgd6_10	Uncharacterized_protein	OG5_222668	7	2
cgd6_1070	Uncharacterized_protein	OG5_128239	12	95
cgd6_1080	Glycoprotein_GP40	OG5_222676	7	2
cgd6_2170	60S_ribosomal_protein_L5	OG5_126688	12	222
cgd6_2330	Uncharacterized_protein	OG5_195453	9	3
cgd6_3990	Elongation_factor_1-alpha	OG5_126631	17	277
cgd6_40	Uncharacterized_protein	OG5_164335	7	6
cgd6_4190	60S_ribosomal_protein_L10/L16,_alpha/beta_hammerhead	OG5_126951	12	158
cgd6_4620	60S_ribosomal_protein_L26	OG5_126931	12	161
cgd6_4630	40S_ribosomal_protein_S8	OG5_127039	12	150
cgd6_5410	Uncharacterized_Secreted_Protein	OG5_222709	7	2
cgd7_130	Ribosomal_protein_S11	OG5_126681	12	225
cgd7_1460	40S_ribosomal_protein_S27	OG5_127008	11	153
cgd7_1873	Zinc-binding_ribosomal_protein_L44	OG5_126988	12	154
cgd7_2110	60S_ribosomal_proteins_L8/L2	OG5_126641	14	264
cgd7_2250	40S_ribosomal_protein_S3_KH2	OG5_126820	13	181
cgd7_2280	Ribosomal_protein_L40e	OG5_128257	27	95
cgd7_2300	Translation_elongation_factor_IF5A	OG5_126922	13	162
cgd7_2540	60S_ribosomal_protein_L35A	OG5_127072	12	147
cgd7_4050	Ribosomal_protein_L38	OG5_127448	11	120
cgd7_480	L-lactate/malate_dehydrogenase	OG5_126911	29	164
cgd8_2340	Cold-shock_DNA-binding_domain-containing__protein	OG5_126866	16	171
cgd8_3480	60S_ribosomal_protein_L34	OG5_127025	12	151
cgd8_4050	Ribosomal_protein_S29	OG5_127205	7	136
cdg8_1840	Ribosomal_protein_S4/S9	OG5_126983	13	154

Table E.4. Expression values of top 50 genes expressed in all *Cryptosporidium parvum* samples (including sporozoite data). The log2 values of the normalised read counts and the normalised read counts per kb (divided by transcript length) are listed for all libraries. Taken from Lippuner et al., (2018).

CryptoDB ID	Reads sporozoites (log2(x+1))	CryptoDB ID	Reads in vivo 2 days p.i. (log2(x+1))	CryptoDB ID	Reads in vivo 4 days p.i. (log2(x+1))	CryptoDB ID	Reads in vitro 2 days p.i. average (log2(x+1))	CryptoDB ID	Reads in vitro 4 days p.i. average (log2(x+1))
cgd8_3520	15.96	cgd6_1080	14.15	cgd7_4020	14.16	cgd6_1080	14.39	cgd8_3520	13.92
cgd7_480	15.80	cgd6_2090	13.01	cgd6_1080	13.81	cgd8_3520	13.51	cgd7_480	13.90
cgd1_640	15.76	cgd7_4020	12.85	cgd8_3520	13.71	cgd7_480	13.40	cgd1_640	13.59
cgd6_5410	14.66	cgd8_3520	12.84	cgd6_2090	13.63	cgd1_640	13.17	cgd6_1080	13.42
cgd3_3770	14.34	cgd6_200	12.57	cgd6_200	13.59	cgd6_5410	12.60	cgd6_5410	12.88
cgd4_3550	14.20	cgd6_3990	12.23	cgd1_640	13.14	cgd3_510	12.20	cgd3_510	12.36
cgd8_3230	14.04	cgd3_3370	12.15	cgd3_3370	13.10	cgd4_3550	12.06	cgd4_3550	12.34
cgd1_3810	13.83	cgd6_5410	12.00	cgd6_5410	12.77	cgd3_3770	11.99	cgd3_3770	12.25
cgd3_510	13.80	cgd5_3160	11.94	cgd7_480	12.68	cgd6_200	11.98	cgd1_3810	12.07
cgd6_2600	13.69	cgd6_10	11.89	cgd5_3160	12.57	cgd1_3810	11.91	cgd8_3230	12.02
cgd2_200	13.67	cgd1_3810	11.79	cgd6_3990	12.28	cgd8_3230	11.68	cgd6_5400	11.91
cgd6_3920	13.65	cgd7_4810	11.75	cgd3_510	12.11	cgd6_5400	11.59	cgd6_10	11.64
cgd6_5400	13.65	cgd7_480	11.71	cgd5_1470	12.06	cgd6_2450	11.53	cgd2_3110	11.59
cgd6_3010	13.63	cgd4_3090	11.69	cgd4_3550	11.93	cgd2_3110	11.39	cgd6_3010	11.57
cgd7_2430	13.42	cgd8_4830	11.68	cgd1_3810	11.93	cgd6_3010	11.38	cgd6_3920	11.55
cgd2_20	13.38	cgd6_710	11.66	cgd2_3110	11.75	cgd6_3460	11.36	cgd6_2600	11.49
cgd2_3110	13.33	cgd5_1470	11.65	cgd6_4460	11.73	cgd6_3990	11.32	cgd8_1770	11.41
cgd6_10	13.31	cgd2_790	11.60	cgd3_1570	11.72	cgd6_10	11.25	cgd6_2090	11.38
cgd8_1770	13.20	cgd7_4880	11.49	cgd6_3920	11.63	cgd6_3920	11.22	cgd2_20	11.36

CryptoDB ID	Reads sporozoites (log2(x+1))	CryptoDB ID	Reads in vivo 2 days p.i. (log2(x+1))	CryptoDB ID	Reads in vivo 4 days p.i. (log2(x+1))	CryptoDB ID	Reads in vitro 2 days p.i. average (log2(x+1))	CryptoDB ID	Reads in vitro 4 days p.i. average (log2(x+1))
cgd5_10	13.07	cgd6_3460	11.40	cgd3_3770	11.58	cgd2_20	11.21	cgd7_2430	11.20
cgd3_1570	13.00	cgd3_3770	11.36	cgd6_2330	11.52	cgd6_2090	11.20	cgd2_200	11.09
cgd4_3630	12.88	cgd3_510	11.34	cgd6_10	11.49	cgd7_3120	11.19	cgd4_3620	10.98
cgd2_940	12.83	cgd7_300	11.34	cgd6_3460	11.35	cgd6_2600	11.16	cgd2_940	10.97
cgd4_3620	12.76	cgd6_120	11.29	cgd7_4810	11.31	cgd8_1770	11.07	cgd5_10	10.91
cgd3_410	12.58	cgd4_3550	11.27	cgd8_2930	11.31	cgd7_4810	11.07	cgd4_3630	10.88
cgd6_4910	12.56	cgd8_2930	11.23	cgd6_710	11.26	cgd7_300	11.02	cgd6_200	10.88
cgd3_4150	12.46	cgd6_2330	11.22	cgd6_3790	11.25	cgd7_1730	10.97	cgd6_3460	10.81
cgd1_1580	12.45	cgd2_20	11.19	cgd2_790	11.23	cgd7_2430	10.85	cgd6_3990	10.72
cgd5_2060	12.37	cgd1_640	11.14	cgd2_200	11.22	cgd8_5230	10.77	cgd1_1580	10.66
cgd2_4320	12.34	cgd6_2450	11.06	cgd6_40	11.14	cgd4_3630	10.76	cgd3_4150	10.63
cgd7_360	12.32	cgd7_1730	11.05	cgd2_490	11.11	cgd2_790	10.73	cgd3_410	10.57
cgd6_4860	12.17	cgd3_1570	11.03	cgd8_440	11.09	cgd2_940	10.69	cgd6_4910	10.56
cgd1_880	12.15	cgd6_4460	11.00	cgd7_300	11.07	cgd7_4880	10.66	cgd7_3010	10.46
cgd1_2400	12.10	cgd6_3790	10.91	cgd7_3790	11.07	cgd8_1720	10.65	cgd3_1570	10.40
cgd8_5340	12.04	cgd6_5440	10.90	cgd4_1910	11.07	cgd2_2110	10.62	cgd2_4320	10.37
cgd8_3770	12.04	cgd7_360	10.85	cgd4_2260	11.04	cgd2_200	10.58	cgd7_4810	10.33
cgd6_1430	12.04	cgd5_1210	10.85	cgd5_10	10.94	cgd1_1580	10.55	cgd5_3160	10.33
cgd7_3010	11.99	cgd4_2260	10.80	cgd2_20	10.92	cgd5_10	10.55	cgd5_2060	10.32
cgd5_1470	11.93	cgd2_3110	10.78	cgd7_400	10.90	cgd7_4020	10.54	cgd3_3940	10.30
cgd2_430	11.93	cgd7_3120	10.75	cgd8_1770	10.89	cgd7_400	10.50	cgd6_2450	10.27
cgd3_3340	11.88	cgd8_1720	10.73	cgd7_4880	10.82	cgd4_3620	10.46	cgd8_5340	10.26
cgd8_2050	11.88	cgd8_1160	10.71	cgd6_4470	10.80	cgd6_4910	10.34	cgd6_1430	10.21
cgd4_3400	11.82	cgd8_440	10.69	cgd6_2450	10.77	cgd3_410	10.31	cgd1_2400	10.20

CryptoDB ID	Reads sporozoites (log2(x+1))	CryptoDB ID	Reads in vivo 2 days p.i. (log2(x+1))	CryptoDB ID	Reads in vivo 4 days p.i. (log2(x+1))	CryptoDB ID	Reads in vitro 2 days p.i. average (log2(x+1))	CryptoDB ID	Reads in vitro 4 days p.i. average (log2(x+1))
cgd6_4470	11.80	cgd5_1220	10.62	cgd6_3010	10.74	cgd4_680	10.30	cgd7_360	10.18
cgd5_40	11.78	cgd7_1280	10.59	cgd7_1730	10.72	cgd5_3160	10.28	cgd2_790	10.17
cgd7_4790	11.77	cgd7_3790	10.54	cgd5_1210	10.61	cgd3_4150	10.28	cgd6_4860	10.13
cgd5_3160	11.72	cgd5_10	10.49	cgd7_360	10.55	cgd1_590	10.27	cgd7_300	10.12
cgd7_940	11.71	cgd4_1910	10.49	cgd8_1720	10.53	cgd4_3220	10.27	cgd7_3120	10.09
cgd6_780	11.70	cgd6_3920	10.35	cgd6_120	10.52	cgd4_2720	10.23	cgd7_1730	10.09
cgd8_4620	11.68	cgd5_1510	10.34	cgd5_1960	10.50	cgd3_1570	10.21	cgd6_780	10.06

Table E.5. Raw mRNA reads from 48 h experiment.

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
ARG1	*Homo sapiens*	24.17	16	29	19.48	23	29	18	16	21	29	28	25	29	28	19	25
ARG2	*Homo sapiens*	200.25	12	343	56.01	301	196	19	12	208	200	122	206	343	149	336	311
BCL2	*Homo sapiens*	22.75	15	31	19.94	24	15	24	23	20	17	26	22	19	28	31	24
CAMP	*Homo sapiens*	24.42	13	31	20.02	13	24	26	27	30	31	20	24	25	20	27	26
CASP1	*Homo sapiens*	316.42	12	572	61.9	548	313	12	15	311	301	172	321	572	193	501	538
CASP3	*Homo sapiens*	467.25	22	884	65.17	749	427	22	26	410	429	235	462	858	282	823	884
CASP4	*Homo sapiens*	1025.83	26	1752	57.67	1752	993	26	33	1108	991	630	1187	1603	732	1596	1659
CASP5	*Homo sapiens*	70.08	12	126	53.35	101	45	12	19	76	71	39	81	126	53	120	98
CCL5	*Homo sapiens*	99.5	14	242	63.31	242	133	16	14	77	88	57	95	131	63	135	143
CD36	*Homo sapiens*	30.17	10	45	38.78	31	21	10	14	29	36	22	37	45	29	44	44
CISH	*Homo sapiens*	30.5	14	48	30.84	48	26	23	14	34	23	26	29	34	29	44	36
CX3CL1	*Homo sapiens*	48.17	13	84	42.61	84	49	23	13	51	60	30	44	59	33	66	66
CXCL8	*Homo sapiens*	3092.58	18	8336	78.35	8336	5825	21	18	2092	2283	1568	2677	4833	1619	3979	3860
DEFA1	*Homo sapiens*	7.5	3	18	65.2	7	5	16	4	7	4	4	3	4	18	9	9
DEFB1	*Homo sapiens*	294.17	15	534	57.23	482	286	15	21	359	329	149	312	433	205	405	534
DEFB4B	*Homo sapiens*	18.33	12	30	29.72	21	12	17	15	15	30	23	12	23	15	15	22
FAS	*Homo sapiens*	73.25	11	139	54.78	139	76	23	11	65	53	40	73	124	57	110	108

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
FASLG	*Homo sapiens*	14.08	9	21	29.89	13	11	9	14	13	19	9	9	21	17	19	15
GUSB	*Homo sapiens*	709	27	1169	56.4	1141	662	27	38	793	716	441	813	1128	503	1169	1077
HPRT1	*Homo sapiens*	2124.42	21	3819	63.75	3312	1969	24	21	2060	1939	1138	2270	3819	1388	3747	3806
ICAM1	*Homo sapiens*	1484.08	20	3366	70.02	3366	1934	20	23	1230	1161	720	1446	2597	733	2298	2281
IFNG	*Homo sapiens*	16.5	7	27	35.2	14	7	11	17	20	19	27	13	21	19	9	21
IL18	*Homo sapiens*	672.67	20	1142	58.18	1049	645	20	30	711	630	409	749	1081	484	1122	1142
IL1B	*Homo sapiens*	148.42	23	283	51.78	283	186	26	23	156	133	101	152	199	103	216	203
IL1R1	*Homo sapiens*	278.75	24	505	57.85	449	252	24	39	330	281	150	266	416	172	461	505
IL33	*Homo sapiens*	20	13	30	27.72	13	23	20	14	26	13	19	18	19	27	18	30
IL6	*Homo sapiens*	196.5	22	503	66.75	503	245	24	22	148	153	115	204	272	128	269	275
NFKB1	*Homo sapiens*	127.67	13	245	57.72	245	124	13	18	143	126	84	130	179	63	216	191
PTGS2	*Homo sapiens*	201.75	20	393	58.98	393	203	26	20	210	177	118	194	282	128	327	343
PYCARD	*Homo sapiens*	1039.33	28	1837	62.12	1718	992	36	28	1024	967	535	1117	1837	682	1741	1795
RELA	*Homo sapiens*	1262.08	16	2256	59.69	2256	1241	16	26	1459	1278	749	1504	1838	716	2057	2005
RPL19	*Homo sapiens*	33569.5	27	59349	58.66	59349	32732	27	28	37931	34976	18976	39228	51670	24285	50358	53274
SDHA	*Homo sapiens*	967.17	18	1919	63.63	1557	848	19	18	1079	944	485	1098	1474	557	1919	1608
SIRT1	*Homo sapiens*	146	21	261	55.24	189	129	26	21	144	131	105	156	261	92	246	252
SOCS1	*Homo sapiens*	213.5	16	386	60.04	365	184	28	16	213	204	117	245	354	117	333	386
SOCS2	*Homo sapiens*	57.33	6	113	56.75	113	61	9	6	45	51	41	64	96	39	87	76
SOCS3	*Homo sapiens*	1100.08	13	1956	58.1	1956	1068	16	13	1265	1101	631	1418	1522	776	1786	1649
SOCS4	*Homo sapiens*	570.33	21	1071	65.08	904	518	21	21	574	469	297	587	1021	338	1071	1023

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
STAT6	*Homo sapiens*	722.75	24	1296	60.74	1269	671	24	32	831	698	388	810	1087	428	1296	1139
TRAF2	*Homo sapiens*	244.67	9	428	56.24	428	202	15	9	277	244	166	299	372	182	405	337
cgd1_1660	*C. parvum* Iowa II	56	9	507	253.8	9	11	507	25	11	18	20	11	9	19	12	20
cgd1_2270	*C. parvum* Iowa II	61.67	13	524	236.27	18	21	524	34	15	13	19	16	16	18	21	25
cgd1_2400	*C. parvum* Iowa II	195.92	53	1533	215.05	67	53	1533	101	59	83	79	60	66	96	80	74
cgd1_2880	*C. parvum* Iowa II	34.5	22	93	54.96	27	26	93	37	34	25	28	31	22	25	32	34
cgd1_300	*C. parvum* Iowa II	191.83	10	2058	306.38	24	15	2058	34	19	23	21	10	14	28	23	33
cgd1_3810	*C. parvum* Iowa II	147.17	10	1560	302.41	10	16	1560	50	14	16	18	20	12	22	16	12
cgd1_640	*C. parvum* Iowa II	712.17	12	8220	332.02	12	16	8220	120	18	26	27	21	14	17	33	22
cgd1_880	*C. parvum* Iowa II	102.58	17	958	262.7	23	17	958	43	18	21	21	26	23	20	31	30
cgd1_910	*C. parvum* Iowa II	116.67	7	1176	286.08	7	17	1176	49	20	21	14	17	11	24	22	22
cgd2_1070	*C. parvum* Iowa II	43.42	14	280	172.1	20	19	280	33	29	18	25	15	20	27	14	21
cgd2_20	*C. parvum* Iowa II	270.33	23	2789	293.47	27	33	2789	85	30	44	39	23	43	53	40	38
cgd2_200	*C. parvum* Iowa II	513.83	12	5844	326.76	13	12	5844	148	14	18	21	20	12	26	22	16
cgd2_2200	*C. parvum* Iowa II	47.42	13	332	189.78	17	13	332	42	16	24	24	17	31	18	20	15
cgd2_2540	*C. parvum* Iowa II	40.25	6	346	239.65	10	6	346	27	7	16	15	10	13	16	10	7
cgd2_3000	*C. parvum* Iowa II	59.92	6	569	267.81	14	7	569	30	7	11	20	11	11	15	6	18
cgd2_3070	*C. parvum* Iowa II	221.17	30	2188	280.14	33	45	2188	86	42	38	38	30	48	31	39	36
cgd2_350	*C. parvum* Iowa II	67.83	10	621	257.07	12	19	621	40	16	21	11	13	10	20	16	15
cgd2_3730	*C. parvum* Iowa II	197.92	8	2174	314.48	8	12	2174	54	15	19	18	13	12	19	19	12
cgd2_430	*C. parvum* Iowa II	323.67	9	3721	330.55	17	9	3721	16	10	13	16	13	14	24	13	18

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
cgd2_4320	*C. parvum* Iowa II	48.08	16	255	136.1	33	26	255	29	35	21	34	16	35	28	29	36
cgd2_820	*C. parvum* Iowa II	141.67	14	1471	295.53	22	21	1471	36	22	15	15	14	21	20	20	23
cgd3_1300	*C. parvum* Iowa II	224.75	78	1483	176.42	121	78	1483	117	105	100	119	95	114	122	112	131
cgd3_1400	*C. parvum* Iowa II	26.58	14	51	36.7	23	23	51	34	17	26	28	14	25	27	18	33
cgd3_1570	*C. parvum* Iowa II	299.17	14	3290	314.89	14	19	3290	85	20	19	21	20	30	23	24	25
cgd3_1940	*C. parvum* Iowa II	236.5	25	2394	287.33	41	31	2394	74	42	40	36	25	41	45	31	38
cgd3_2250	*C. parvum* Iowa II	63.42	8	603	268.06	21	11	603	20	16	13	10	8	15	13	8	23
cgd3_2940	*C. parvum* Iowa II	25.67	9	93	84.63	20	17	93	23	16	20	24	15	25	9	21	25
cgd3_330	*C. parvum* Iowa II	53.67	16	343	171.02	16	17	343	59	22	28	26	26	28	26	21	32
cgd3_3770	*C. parvum* Iowa II	519.92	11	6001	332.02	15	15	6001	85	12	11	12	13	18	15	23	19
cgd3_3790	*C. parvum* Iowa II	37.42	8	312	231.25	8	10	312	13	16	18	11	8	15	13	11	14
cgd3_3930	*C. parvum* Iowa II	202.67	18	2120	297.95	29	20	2120	47	27	33	32	19	18	28	31	28
cgd3_4150	*C. parvum* Iowa II	130.5	9	1349	294.12	14	18	1349	44	27	14	20	9	14	20	15	22
cgd3_4260	*C. parvum* Iowa II	38.83	6	324	231.73	11	6	324	20	7	15	27	9	12	13	10	12
cgd3_510	*C. parvum* Iowa II	265.25	7	2968	321.02	10	7	2968	97	9	15	9	7	13	11	22	15
cgd3_760	*C. parvum* Iowa II	89.67	11	893	282.19	15	19	893	28	18	11	20	11	13	17	18	13
cgd4_1940	*C. parvum* Iowa II	32.67	9	168	133.07	17	9	168	39	17	19	18	10	25	32	15	23
cgd4_2260	*C. parvum* Iowa II	71.67	5	719	284.56	5	11	719	25	19	6	15	8	13	8	15	16
cgd4_3080	*C. parvum* Iowa II	25.42	11	81	72.63	15	16	81	19	24	30	26	11	22	28	14	19
cgd4_3160	*C. parvum* Iowa II	35.75	18	120	76.02	40	25	120	28	31	18	28	29	35	20	26	29
cgd4_3550	*C. parvum* Iowa II	93.08	15	885	267.97	18	15	885	28	16	26	19	20	29	26	18	17

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
cgd4_3620	*C. parvum* Iowa II	57.67	9	507	246.32	10	12	507	55	15	13	17	12	9	17	9	16
cgd4_4310	*C. parvum* Iowa II	83.33	14	758	255.19	14	19	758	49	22	15	19	22	21	17	18	26
cgd4_4460	*C. parvum* Iowa II	28.42	16	107	88.4	28	16	107	25	18	19	18	18	22	20	20	30
cgd5_10	*C. parvum* Iowa II	246.25	15	2681	311.45	21	23	2681	79	15	26	20	20	17	16	19	18
cgd5_1470	*C. parvum* Iowa II	102.83	9	1070	296.22	10	14	1070	27	17	15	16	12	15	15	14	9
cgd5_2370	*C. parvum* Iowa II	49.75	10	365	199.94	10	19	365	30	24	17	23	20	18	26	14	31
cgd5_2800	*C. parvum* Iowa II	66	11	582	246.43	16	11	582	37	18	12	24	17	17	25	16	17
cgd5_3040	*C. parvum* Iowa II	91.17	10	895	277.74	19	15	895	32	21	22	18	10	15	17	18	12
cgd5_4590	*C. parvum* Iowa II	47.42	14	318	179.99	25	22	318	27	23	20	21	17	31	22	29	14
cgd6_10	*C. parvum* Iowa II	197.25	20	2099	303.63	20	28	2099	25	30	25	23	21	22	26	24	24
cgd6_1070	*C. parvum* Iowa II	14.5	5	81	145.6	7	9	81	14	6	7	8	8	6	12	5	11
cgd6_1080	*C. parvum* Iowa II	17.08	7	71	101.53	17	12	71	11	13	10	7	10	11	20	9	14
cgd6_2170	*C. parvum* Iowa II	97.25	19	888	256.14	20	20	888	35	26	20	31	19	23	25	22	38
cgd6_2330	*C. parvum* Iowa II	31.58	12	176	145.47	12	13	176	35	19	19	18	14	12	26	18	17
cgd6_2600	*C. parvum* Iowa II	348.67	8	3991	329	8	15	3991	61	13	9	17	10	14	18	12	16
cgd6_3010	*C. parvum* Iowa II	59.08	4	557	265.5	14	12	557	20	13	4	19	11	20	13	11	15
cgd6_3800	*C. parvum* Iowa II	54.33	29	180	74.91	36	37	180	60	49	36	43	34	59	44	29	45
cgd6_3850	*C. parvum* Iowa II	83.17	8	818	278.36	10	16	818	31	10	11	21	15	24	16	18	8
cgd6_3920	*C. parvum* Iowa II	480.75	14	5504	329.09	20	14	5504	98	15	14	19	14	16	22	17	16
cgd6_3990	*C. parvum* Iowa II	136.67	8	1466	306.35	8	17	1466	33	13	13	18	13	11	21	11	16
cgd6_40	*C. parvum* Iowa II	411.17	25	4580	319.3	33	26	4580	25	37	38	44	30	27	33	30	31

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
cgd6_4190	*C. parvum* Iowa II	42.58	9	370	242.24	11	10	370	19	11	13	17	9	15	11	14	11
cgd6_4620	*C. parvum* Iowa II	92.75	11	909	277.32	16	16	909	46	15	11	16	14	22	15	17	16
cgd6_4630	*C. parvum* Iowa II	128.42	9	1348	299.11	22	9	1348	29	14	15	15	14	17	25	15	18
cgd6_4860	*C. parvum* Iowa II	161	13	1673	295.82	21	13	1673	55	18	23	18	15	25	23	27	21
cgd6_4910	*C. parvum* Iowa II	207.58	11	2262	311.7	19	23	2262	46	24	13	20	11	15	18	19	21
cgd6_5400	*C. parvum* Iowa II	109.17	2	1239	325.95	7	6	1239	17	2	8	5	5	3	2	7	9
cgd6_5410	*C. parvum* Iowa II	637.33	16	7242	326.4	16	21	7242	146	20	19	33	29	28	36	28	30
cgd7_120	*C. parvum* Iowa II	66.75	13	507	208.07	18	13	507	46	30	30	30	22	28	26	21	30
cgd7_130	*C. parvum* Iowa II	104.33	23	920	246.31	32	24	920	50	28	28	23	25	29	25	28	40
cgd7_1460	*C. parvum* Iowa II	99.42	8	983	279.94	8	14	983	29	14	17	19	20	24	20	21	24
cgd7_1830	*C. parvum* Iowa II	14.33	5	55	95.49	7	9	55	20	8	9	19	5	14	12	6	8
cgd7_2110	*C. parvum* Iowa II	99.67	11	968	274.44	22	21	968	35	22	11	24	18	19	22	12	22
cgd7_2250	*C. parvum* Iowa II	38.42	20	157	98.8	30	25	157	39	21	38	34	22	31	20	20	24
cgd7_2280	*C. parvum* Iowa II	73	5	680	262.14	19	15	680	41	13	18	14	5	18	24	9	20
cgd7_2300	*C. parvum* Iowa II	421.5	8	4839	330.06	8	26	4839	55	12	13	23	18	14	21	16	13
cgd7_2430	*C. parvum* Iowa II	377.33	12	4246	322.91	12	17	4246	80	12	32	30	18	29	17	19	16
cgd7_2540	*C. parvum* Iowa II	99.08	7	1020	292.79	15	14	1020	37	13	13	12	7	13	11	19	15
cgd7_3010	*C. parvum* Iowa II	66.25	13	574	241.53	18	13	574	36	15	23	23	19	23	15	18	18
cgd7_3020	*C. parvum* Iowa II	48.92	13	304	164.83	26	13	304	39	29	25	31	16	20	26	28	30
cgd7_3240	*C. parvum* Iowa II	57.17	8	483	234.85	12	17	483	33	23	20	20	8	17	15	24	14
cgd7_3250	*C. parvum* Iowa II	117.67	10	1228	297.21	11	10	1228	32	21	16	21	15	15	14	11	18

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
cgd7_4050	*C. parvum* Iowa II	101.75	11	1020	284.26	15	15	1020	26	15	11	13	13	27	24	16	26
cgd7_480	*C. parvum* Iowa II	503.33	11	5707	325.68	13	17	5707	161	24	14	23	18	15	11	20	17
cgd7_700	*C. parvum* Iowa II	18.92	6	70	90.12	16	14	70	28	6	13	17	9	16	10	10	18
cgd7_790	*C. parvum* Iowa II	75	6	759	287.25	7	18	759	15	14	11	16	6	12	13	17	12
cgd8_1840	*C. parvum* Iowa II	86.83	23	618	192.88	40	34	618	54	26	40	36	23	43	40	41	47
cgd8_2340	*C. parvum* Iowa II	44.5	9	334	205.27	9	15	334	23	16	29	18	11	17	26	18	18
cgd8_3230	*C. parvum* Iowa II	220.67	7	2475	321.75	7	9	2475	42	8	18	20	11	13	16	13	16
cgd8_3390	*C. parvum* Iowa II	61.83	13	518	232.53	26	19	518	32	18	23	13	13	17	16	18	29
cgd8_3460	*C. parvum* Iowa II	14.25	8	25	40.27	17	15	25	8	22	15	8	8	9	10	17	17
cgd8_3480	*C. parvum* Iowa II	111.75	3	1204	307.86	12	10	1204	31	3	16	11	7	11	17	9	10
cgd8_3520	*C. parvum* Iowa II	982.58	14	11423	334.64	25	21	11423	151	14	15	32	16	22	24	29	19
cgd8_3770	*C. parvum* Iowa II	119.83	19	1114	261.34	28	19	1114	44	27	21	42	25	30	32	32	24
cgd8_4050	*C. parvum* Iowa II	141.5	19	1403	280.78	23	24	1403	31	28	21	37	19	30	24	24	34
cgd8_4220	*C. parvum* Iowa II	79.67	23	631	218.04	34	30	631	43	23	27	24	26	31	28	27	32
cgd8_4320	*C. parvum* Iowa II	304.58	14	3362	316.2	21	24	3362	98	21	16	16	24	18	14	21	20
cgd8_4500	*C. parvum* Iowa II	34.5	7	237	186.15	16	7	237	37	12	15	15	9	17	19	19	11
cgd8_60	*C. parvum* Iowa II	141.5	21	1375	274.55	21	24	1375	37	30	35	37	24	35	25	31	24
cgd_2_140	*C. parvum* Iowa II	270.83	9	3039	321.9	21	15	3039	49	20	9	17	16	17	15	20	12
cgd2_3270	*C. parvum* Iowa II	26.83	6	169	170.01	15	12	169	40	13	13	15	6	9	12	6	12
cgd5_3160	*C. parvum* Iowa II	55.25	7	427	214.71	14	16	427	81	16	12	17	22	18	7	16	17
cgd6_2090	*C. parvum* Iowa II	19.33	5	115	158.31	5	11	115	23	9	8	19	9	7	13	7	6

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S.typhimurium +	S.typhimurium +	C. parvum Spors	C. hominis Spors	Cells alone	Cells alone	C. parvum +	C. hominis +	C. parvum +	C. hominis +	C. hominis +	C. hominis +
cgd6_4270	*C. parvum* Iowa II	34.67	14	188	139.94	15	21	188	31	22	14	21	16	20	26	20	22

Table E.6. Raw mRNA reads from timeseries experiment.

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. homins +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
ARG1	*Homo sapiens*	12.17	9	17	20.39	11	17	14	14	11	15	10	9	13	12	9	11
ARG2	*Homo sapiens*	175.25	9	327	54.08	249	14	9	194	160	209	274	107	327	196	178	186
BCL2	*Homo sapiens*	15	6	22	35.05	20	18	9	20	6	20	13	9	14	12	22	17
CAMP	*Homo sapiens*	13.42	9	18	20.71	9	15	15	14	18	10	17	10	12	14	13	14
CASP1	*Homo sapiens*	511.58	11	958	59.04	286	17	11	595	534	796	958	294	552	617	794	685
CASP3	*Homo sapiens*	388.08	13	687	54.33	370	17	13	417	403	552	687	207	436	461	580	514
CASP4	*Homo sapiens*	1163.17	22	1875	53.01	926	30	22	1402	1306	1732	1875	736	1320	1504	1608	1497
CASP5	*Homo sapiens*	53	11	89	49.42	37	11	13	45	52	72	89	32	63	83	80	59
CCL5	*Homo sapiens*	200.33	10	1128	148.99	1128	10	12	149	96	150	154	46	179	147	174	159
CD36	*Homo sapiens*	31.92	7	67	51.59	23	14	7	24	37	54	67	23	33	40	29	32
CISH	*Homo sapiens*	27.67	7	40	39.27	24	7	12	20	32	34	40	23	27	34	40	39
CX3CL1	*Homo sapiens*	262.58	11	853	82.6	853	13	11	251	251	341	349	102	242	257	274	207

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. hominis +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
CXCL8	*Homo sapiens*	9395	10	44779	124.13	44779	18	10	7634	5502	6776	9273	2944	9103	8488	9462	8751
DEFA1	*Homo sapiens*	5.75	3	10	31.57	7	6	6	6	6	5	6	5	3	10	3	6
DEFB1	*Homo sapiens*	550	12	1234	65.32	302	17	12	538	648	1056	1234	437	459	576	711	610
DEFB4B	*Homo sapiens*	8.5	4	14	42.71	14	11	5	8	5	6	8	5	10	14	12	4
FAS	*Homo sapiens*	83.08	5	130	50.25	130	8	5	100	75	89	126	49	104	101	113	97
FASLG	*Homo sapiens*	8.17	3	13	34.57	5	7	10	10	5	11	9	3	13	9	8	8
GUSB	*Homo sapiens*	811.08	13	1438	54.22	642	16	13	859	874	1159	1438	564	954	944	1151	1119
HPRT1	*Homo sapiens*	1805	12	3124	55.32	1549	12	13	1965	1822	2429	3124	1069	2168	2286	2852	2371
ICAM1	*Homo sapiens*	2698.58	3	12153	117.14	12153	3	4	2126	1508	2005	2589	796	2855	2626	3037	2681
IFNG	*Homo sapiens*	11.25	6	21	39.23	11	21	11	8	9	8	18	6	8	11	14	10
IL18	*Homo sapiens*	450.42	12	760	54.17	451	18	12	464	419	553	710	253	581	544	760	640
IL1B	*Homo sapiens*	495.25	16	2781	149.87	2781	18	16	351	200	260	258	113	450	427	569	500
IL1R1	*Homo sapiens*	274.33	18	511	53.7	176	19	18	290	326	410	511	209	299	321	368	345
IL33	*Homo sapiens*	29.25	12	49	36.21	27	17	12	34	29	31	33	15	29	35	49	40
IL6	*Homo sapiens*	133	16	694	135.91	694	24	16	81	80	74	102	45	140	105	110	125
NFKB1	*Homo sapiens*	184	8	556	75.54	556	16	8	152	139	214	240	103	174	192	212	202
PTGS2	*Homo sapiens*	440.42	8	1573	90.88	1573	8	14	332	372	514	626	221	389	398	439	399
PYCARD	*Homo sapiens*	1059.92	20	1912	55.94	650	33	20	1119	1210	1550	1912	724	1233	1298	1557	1413
RELA	*Homo sapiens*	1721.75	12	2842	52.79	2417	14	12	1771	1841	2443	2842	1044	1961	2002	2154	2160
RPL19	*Homo sapiens*	37761.08	24	64785	54.67	31880	24	24	42970	40816	56030	64785	23803	43444	45243	54611	49503
SDHA	*Homo sapiens*	791.67	6	1473	55.51	634	13	6	766	889	1052	1473	511	975	1002	1094	1085
SIRT1	*Homo sapiens*	110.17	13	182	46.29	130	20	13	110	126	153	182	78	104	147	140	119
SOCS1	*Homo sapiens*	186.92	12	336	53.03	236	12	14	186	187	232	336	97	233	190	263	257

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. hominis +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
SOCS2	*Homo sapiens*	57.75	3	113	56.73	113	8	3	47	54	72	88	28	67	75	85	53
SOCS3	*Homo sapiens*	637.83	14	1038	51.63	795	16	14	677	719	852	1038	369	780	725	815	854
SOCS4	*Homo sapiens*	491.75	10	867	54.35	373	22	10	509	574	712	867	304	568	614	665	683
STAT6	*Homo sapiens*	667.83	7	1142	53.67	602	17	7	692	746	1034	1142	440	717	825	901	891
TRAF2	*Homo sapiens*	246.58	7	394	53.34	209	12	7	246	279	362	392	153	272	321	394	312
cgd1_1660	*C. parvum* Iowa II	41.92	3	389	261.12	10	389	27	8	7	13	10	3	8	9	10	9
cgd1_2270	*C. parvum* Iowa II	39.25	5	372	267.22	8	372	20	6	10	6	10	5	14	10	5	5
cgd1_2400	*C. parvum* Iowa II	117.92	3	1260	305.1	3	1260	39	12	13	11	12	15	12	11	10	17
cgd1_2880	*C. parvum* Iowa II	20.33	10	55	58.84	15	55	29	13	10	19	21	16	13	20	15	18
cgd1_300	*C. parvum* Iowa II	142.67	10	1539	308.24	12	1539	30	18	12	17	14	10	12	14	17	17
cgd1_3810	*C. parvum* Iowa II	97.08	3	1057	311.51	10	1057	35	4	3	7	12	4	4	12	12	5
cgd1_640	*C. parvum* Iowa II	466.08	8	5297	326.44	8	5297	85	16	14	27	47	21	18	21	17	22
cgd1_880	*C. parvum* Iowa II	63.5	7	614	273.13	12	614	15	11	9	21	23	13	12	18	7	7
cgd1_910	*C. parvum* Iowa II	70.67	7	685	274.05	12	685	40	12	12	10	22	14	9	12	13	7
cgd2_1070	*C. parvum* Iowa II	32.5	9	203	165.62	21	203	23	13	16	17	20	9	14	18	19	17
cgd2_20	*C. parvum* Iowa II	166.75	10	1785	305.69	21	1785	51	16	13	15	23	11	11	22	23	10
cgd2_200	*C. parvum* Iowa II	394	5	4372	318.19	8	4372	180	9	5	23	40	35	11	19	14	12
cgd2_2200	*C. parvum* Iowa II	38.58	9	307	219.71	12	307	31	14	12	10	22	15	9	12	10	9
cgd2_2540	*C. parvum* Iowa II	31.5	4	286	254.72	4	286	15	6	5	7	13	14	10	5	7	6
cgd2_3000	*C. parvum* Iowa II	40.67	1	398	276.96	5	398	19	7	7	9	12	1	12	9	6	3
cgd2_3070	*C. parvum* Iowa II	139.08	24	1299	262.72	30	1299	49	31	25	40	28	47	24	45	26	25
cgd2_350	*C. parvum* Iowa II	46.75	2	453	273.94	2	453	26	10	9	9	7	5	9	11	12	8
cgd2_3730	*C. parvum* Iowa II	121	4	1317	311.39	9	1317	42	9	10	4	19	11	6	9	7	9

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. homins +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
cgd2_430	*C. parvum* Iowa II	267.17	6	3103	334.27	10	3103	6	7	7	8	21	6	8	11	9	10
cgd2_4320	*C. parvum* Iowa II	30.5	9	177	151.85	15	177	19	18	16	18	26	13	9	17	20	18
cgd2_820	*C. parvum* Iowa II	103.67	7	1084	297.87	13	1084	34	12	11	12	16	11	16	15	13	7
cgd3_1300	*C. parvum* Iowa II	119.25	11	1176	279.22	17	1176	54	15	22	21	29	11	19	27	22	18
cgd3_1400	*C. parvum* Iowa II	17.33	8	35	41.11	8	35	23	15	15	11	17	11	18	19	14	22
cgd3_1570	*C. parvum* Iowa II	216.5	10	2402	317.91	11	2402	22	13	16	17	30	20	14	26	17	10
cgd3_1940	*C. parvum* Iowa II	150.5	16	1543	291.48	21	1543	58	18	16	19	30	18	16	16	30	21
cgd3_2250	*C. parvum* Iowa II	49	6	480	277.05	11	480	15	8	9	7	10	9	13	9	6	11
cgd3_2940	*C. parvum* Iowa II	18.08	11	56	70.13	11	56	26	15	18	11	12	11	14	14	12	17
cgd3_330	*C. parvum* Iowa II	35.58	10	252	191.83	19	252	24	16	11	19	17	13	17	14	10	15
cgd3_3770	*C. parvum* Iowa II	375.83	5	4332	331.52	9	4332	57	9	12	10	27	21	9	13	5	6
cgd3_3790	*C. parvum* Iowa II	28.67	1	224	215.83	7	224	27	13	8	17	14	5	1	10	11	7
cgd3_3930	*C. parvum* Iowa II	157.33	9	1692	307.2	17	1692	30	9	14	18	22	13	21	9	23	20
cgd3_4150	*C. parvum* Iowa II	93.5	5	1000	305.42	11	1000	32	5	10	11	16	7	8	5	11	6
cgd3_4260	*C. parvum* Iowa II	21.75	3	188	241.02	8	188	4	11	10	5	10	6	4	5	7	3
cgd3_510	*C. parvum* Iowa II	180.42	3	1984	314.89	9	1984	51	3	10	16	21	28	8	12	12	11
cgd3_760	*C. parvum* Iowa II	65	5	651	284.14	6	651	32	5	9	6	18	7	11	13	11	11
cgd4_1940	*C. parvum* Iowa II	21.67	6	136	166.98	14	136	18	6	11	14	13	10	7	7	14	10
cgd4_2260	*C. parvum* Iowa II	53.33	6	522	276.96	8	522	29	11	8	10	9	6	10	9	7	11
cgd4_3080	*C. parvum* Iowa II	14.67	5	69	118.89	6	69	14	7	9	8	12	5	14	13	13	6
cgd4_3160	*C. parvum* Iowa II	30.42	11	98	73.49	31	98	30	11	25	22	27	29	15	20	35	22
cgd4_3550	*C. parvum* Iowa II	61.67	6	599	274.56	10	599	26	10	9	15	6	8	15	14	8	20
cgd4_3620	*C. parvum* Iowa II	34.08	2	291	237.93	9	291	23	6	7	10	15	11	10	17	8	2

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. homins +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
cgd4_4310	*C. parvum* Iowa II	63.25	3	596	265.56	18	596	37	3	13	19	12	8	14	13	13	13
cgd4_4460	*C. parvum* Iowa II	20.67	10	68	75.65	20	68	27	10	11	17	17	16	12	21	14	15
cgd5_10	*C. parvum* Iowa II	166.75	2	1818	312	12	1818	65	4	2	17	15	21	15	17	7	8
cgd5_1470	*C. parvum* Iowa II	73.92	4	787	303.84	11	787	16	9	9	6	12	4	7	10	11	5
cgd5_2370	*C. parvum* Iowa II	31.25	4	259	229.89	4	259	16	12	13	12	17	6	8	13	5	10
cgd5_2800	*C. parvum* Iowa II	49.83	4	475	269.09	6	475	33	11	7	10	8	13	12	12	7	4
cgd5_3040	*C. parvum* Iowa II	57.92	3	584	286.25	10	584	27	3	8	12	14	3	8	10	8	8
cgd5_4590	*C. parvum* Iowa II	30.33	7	184	160.3	20	184	19	13	17	7	21	13	16	17	25	12
cgd6_10	*C. parvum* Iowa II	158.33	7	1736	313.8	20	1736	14	13	18	19	18	13	16	14	12	7
cgd6_1070	*C. parvum* Iowa II	10.17	4	41	102.55	6	41	18	8	8	5	5	4	7	4	6	10
cgd6_1080	*C. parvum* Iowa II	12.33	4	45	89.91	6	45	7	6	5	11	16	4	17	10	12	9
cgd6_2170	*C. parvum* Iowa II	70.58	7	706	283.57	9	706	21	12	10	13	12	7	15	12	21	9
cgd6_2330	*C. parvum* Iowa II	19.58	3	109	147.56	11	109	30	11	6	11	11	3	8	10	12	13
cgd6_2600	*C. parvum* Iowa II	241.33	6	2760	328.68	7	2760	33	6	9	10	17	12	9	18	8	7
cgd6_3010	*C. parvum* Iowa II	35	5	329	264.66	5	329	15	8	6	5	11	9	7	9	10	6
cgd6_3800	*C. parvum* Iowa II	35.67	12	147	100.92	31	147	41	18	15	30	33	12	22	24	30	25
cgd6_3850	*C. parvum* Iowa II	61.5	3	619	285.71	3	619	31	12	10	7	13	5	10	10	13	5
cgd6_3920	*C. parvum* Iowa II	329.5	6	3744	326.37	9	3744	63	6	7	15	31	18	11	23	17	10
cgd6_3990	*C. parvum* Iowa II	115.83	6	1265	312.47	8	1265	28	13	8	6	16	8	7	8	15	8
cgd6_40	*C. parvum* Iowa II	290.58	14	3229	318.46	25	3229	20	30	17	24	35	14	20	21	30	22
cgd6_4190	*C. parvum* Iowa II	26	3	234	252.18	4	234	14	6	7	7	5	7	10	9	6	3
cgd6_4620	*C. parvum* Iowa II	66.42	4	676	289.29	9	676	35	7	4	12	13	7	11	9	9	5
cgd6_4630	*C. parvum* Iowa II	93.25	2	999	305.98	2	999	30	16	12	6	7	9	8	8	10	12

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. homins +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
cgd6_4860	*C. parvum* Iowa II	99.92	7	1041	296.76	7	1041	41	9	22	9	16	8	13	14	9	10
cgd6_4910	*C. parvum* Iowa II	128.25	7	1385	308.67	11	1385	41	13	10	9	17	13	7	10	11	12
cgd6_5400	*C. parvum* Iowa II	68	1	758	319.6	2	758	14	5	3	7	9	1	4	6	6	1
cgd6_5410	*C. parvum* Iowa II	461.83	10	5223	324.72	10	5223	114	18	19	17	48	28	19	23	12	11
cgd7_120	*C. parvum* Iowa II	43.83	9	353	222.91	17	353	41	13	12	13	9	14	15	13	11	15
cgd7_130	*C. parvum* Iowa II	71.83	17	618	239.52	24	618	33	18	17	19	24	19	23	26	19	22
cgd7_1460	*C. parvum* Iowa II	70.33	8	713	287.83	13	713	25	9	10	15	9	8	13	10	10	9
cgd7_1830	*C. parvum* Iowa II	8.75	4	37	106.56	5	37	14	7	5	4	7	4	5	5	8	4
cgd7_2110	*C. parvum* Iowa II	67.42	8	672	282.52	11	672	27	8	10	12	10	9	18	9	12	11
cgd7_2250	*C. parvum* Iowa II	21	9	93	109.76	15	93	25	11	10	18	14	9	15	14	13	15
cgd7_2280	*C. parvum* Iowa II	59.75	6	582	275.38	11	582	24	8	14	8	17	11	9	15	12	6
cgd7_2300	*C. parvum* Iowa II	383.33	8	4389	329.1	8	4389	60	12	14	10	44	11	16	13	13	10
cgd7_2430	*C. parvum* Iowa II	269.42	7	2988	317.82	15	2988	66	20	14	13	27	24	17	20	22	7
cgd7_2540	*C. parvum* Iowa II	76.17	5	796	297.76	6	796	31	11	11	10	7	6	12	5	12	7
cgd7_3010	*C. parvum* Iowa II	43.58	3	399	257.05	11	399	13	8	6	10	23	13	12	3	12	13
cgd7_3020	*C. parvum* Iowa II	26.17	8	158	159.64	15	158	23	12	9	8	10	13	18	15	21	12
cgd7_3240	*C. parvum* Iowa II	39.33	8	340	241.11	8	340	27	8	10	8	17	10	11	12	9	12
cgd7_3250	*C. parvum* Iowa II	99.33	2	1064	305.92	8	1064	33	2	6	8	14	10	14	10	13	10
cgd7_4050	*C. parvum* Iowa II	71.25	8	717	285.55	12	717	31	13	10	9	10	8	8	15	11	11
cgd7_480	*C. parvum* Iowa II	326.83	5	3696	324.71	7	3696	86	6	5	17	27	32	9	21	7	9
cgd7_700	*C. parvum* Iowa II	12.67	3	49	96.06	14	49	15	11	6	6	6	3	6	8	15	13
cgd7_790	*C. parvum* Iowa II	45.25	2	441	275.54	9	441	13	9	6	12	14	9	6	14	8	2
cgd8_1840	*C. parvum* Iowa II	62.5	17	448	194.61	29	448	41	23	17	31	23	19	27	39	32	21

Probe Name	Species Name	Avg Count	Min Count	Max Count	%CV	S. typhimurium +	C. parvum Spors	C. hominis Spor	Cells alone	120 h C. parvum +	120 h C. hominis +	96 h C. parvum +	96 h C. hominis +	48 h C. parvum +	48 h C. hominis +	24 h C. parvum +	24 h C. hominis +
cgd8_2340	*C. parvum* Iowa II	36	3	291	223.35	17	291	19	15	3	13	16	11	11	12	13	11
cgd8_3230	*C. parvum* Iowa II	153.33	3	1691	315.91	6	1691	49	9	3	10	20	6	13	14	11	8
cgd8_3390	*C. parvum* Iowa II	43.5	6	366	233.96	23	366	29	13	7	9	15	13	6	14	10	17
cgd8_3460	*C. parvum* Iowa II	10.5	6	18	34.82	11	18	6	7	8	12	12	13	9	15	8	7
cgd8_3480	*C. parvum* Iowa II	73.17	3	785	306.43	7	785	17	5	8	8	8	3	13	10	10	4
cgd8_3520	*C. parvum* Iowa II	600	8	6855	328.35	11	6855	132	8	10	26	53	27	25	27	14	12
cgd8_3770	*C. parvum* Iowa II	65.92	10	608	259.11	14	608	30	19	17	12	20	12	10	18	12	19
cgd8_4050	*C. parvum* Iowa II	102.42	8	1005	277.63	24	1005	35	25	20	13	20	11	8	24	27	17
cgd8_4220	*C. parvum* Iowa II	56.75	17	419	201.18	20	419	34	25	25	25	28	21	17	26	18	23
cgd8_4320	*C. parvum* Iowa II	222.17	4	2447	315.47	4	2447	72	21	8	18	27	16	14	14	12	13
cgd8_4500	*C. parvum* Iowa II	21.33	5	133	167.46	15	133	28	6	5	13	6	5	10	13	11	11
cgd8_60	*C. parvum* Iowa II	101	4	1058	298.43	15	1058	15	12	15	4	21	8	20	20	12	12
cgd_2_140	*C. parvum* Iowa II	192.08	5	2186	326.93	5	2186	35	8	7	10	9	8	11	5	12	9
cgd2_3270	*C. parvum* Iowa II	19.67	3	118	164.44	6	118	39	10	10	9	6	7	11	12	3	5
cgd5_3160	*C. parvum* Iowa II	38.25	6	326	237.2	13	326	24	12	11	6	9	13	8	15	10	12
cgd6_2090	*C. parvum* Iowa II	11.83	1	58	133.95	7	58	26	8	5	5	11	6	7	5	1	3
cgd6_4270	*C. parvum* Iowa II	27.08	12	159	153.65	16	159	21	15	12	17	15	13	16	15	14	12

Fin