Genetic variability of proteins involved in metronidazole metabolism and detoxification in *Giardia lamblia*

Christina Skår Saghaug

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

Genetic variability of proteins involved in metronidazole metabolism and detoxification in *Giardia lamblia*

Christina Skår Saghaug

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 09.12.2021

Year: 2021

Title: Genetic variability of proteins involved in metronidazole metabolism and detoxification in *Giardia lamblia*

Name: Christina Skår Saghaug

Print: Skipnes Kommunikasjon / University of Bergen

SCIENTIFIC ENVIRONMENT

This doctoral thesis project originates from the Bergen *Giardia* research group (BGRG) based at the University of Bergen (UiB), Bergen, Norway. The research was completed between 2016 and 2021 and was mainly conducted at the Department of Clinical Science, UiB, while other parts of the thesis were conducted at the Department of Biomedicine and The Center for Medical Genetics and Molecular Medicine, UiB. Finally, some parts of the thesis were carried out at the Department of Infectious Diseases, Unit 16 Mycotic and Parasite Agents and Mycobacteria, Robert-Koch-Institute, Berlin, Germany. Analytical expertise for Illumina whole genome sequencing was provided by Bioinformaticians at the Genomics Core Facility (GCF), UiB.

The Faculty of Medicine, UiB, provided the PhD grant for this thesis. Further financial support was granted by The Western Norway Regional Health Authority, Department of Clinical Science (K2) at University of Bergen, Norwegian Surveillance System for Antimicrobial Drug Resistance at University Hospital of North Norway and Norwegian PhD School of Pharmacy. Finally, The Norwegian National Advisory Unit on Tropical Infectious Diseases (Tropesenteret), Department of Medicine at Haukeland University Hospital has contributed with financial support.

The research in this thesis has been conducted under the supervision of Associate Professor Kurt Hanevik, M.D PhD, and Professor Nina Langeland, M.D PhD.

NASJONAL KOMPETANSETJENESTE
for tropiske infeksjonssykdommer

UNIVERSITAS BERGENSIS

ROBERT KOCH INSTITUT

ACKNOWLEDGEMENTS

After fulfilling a Master's degree of Pharmacy and getting my foot inside the 'researching gates', I decided to start a PhD. At that time, it was rather difficult for me to land a PhD position; especially without any published articles and nothing but my grades and an early research project-draft proposal in my hands. I was lucky enough to receive the opportunity of having a temporary 50 % position at Tropesenteret to continue part-time research and publish an article based on my master's degree results. This temporary part-time job helped me tremendously to get a PhD position. For this I am forever grateful to **Kristine Mørch,** leader at Tropesenteret. Writing a significant scientific thesis is hard work and it could not have been possible without support and encouragement from various people.

First and foremost, I would like to express my deep and sincere gratitude to my main-supervisor, **Kurt Hanevik**. The tireless effort, encouragement, and availability you have shown me have been hugely appreciated. The thesis and publications would not have been written successfully without your continuous support and guidance. You are extremely precise, and I have appreciated all your feedback, even after returning draft number 23 with track changes. All the drawings you have made, and your way of explaining difficult things in a simple manner – make you a true teacher and supervisor indeed.

Next, I wish to express my greatest appreciation towards my co-supervisor **Nina Langeland** for the intellectual guidance and your valuable advice that have carried me through all the stages of my PhD. I am extremely lucky to have been supervised by such enthusiastic and bright-minded people. My supervisors are an inspiration to continue research and continue to strive for success.

I would also like to give a special thanks to my all my co-authors who have contributed with their expertise, guidance, and knowledge within different fields: **Hans-Richard Brattbakk**, **Tomasz Stokowy**, **Juha P. Kallio**, **Inari Kursula**, **Juha Vahokoski**. **Christian Klotz** and **Toni Aebischer**. I wish to send a special thanks to Christian and Toni. You put in a lot of effort and guided me during my three months stay at the Robert-Koch Institute in Berlin to collect data for my second article. I got the

opportunity to learn new methods, new lab routines and maybe most importantly, adapt to a new laboratory and fulfilling my experimental goals.

A thank you goes out to my family, especially my Mother, **Siv Skår Saghaug**, Father, **Tore Saghaug**, and Sister, **Tone Skår Saghaug**, for always being encouraging, motivating and generous, and asking me how the PhD is going, even though you may not understand what I have been doing, you have always been there.

The scientific environment at the 5th floor, laboratory building has been a big part of my PhD. I would like to thank all the friendly and hardworking people that work there, especially those who have contributed with help, guidance and passing on skills: **Steinar Sørnes, Helene Sandnes**, **Sonja Ljostveit, Audun Nerland, Christel Gill Haanshuus** and **Marit Tellevik Gjerde**, to name a few. I thank everyone for all your contributions.

Further, I wish to thank my master students **Astrid L. Gamlem**, **Kirsti B. Hauge** and **Ole-Morten L. Morvik** for your immense contributions to my final project in my PhD, where especially Astrid's efforts provided crucial data for Paper III. I would also like to thank all the new friends and acquaintances in the research community for all contributions, tips, lovely lunches, conferences and sharing frustration over failed experiments: **Sunniva**, **Sam**, **Katrin**, **Seehe**, **Sarah-Marie**, **Oda**, **Tara**, **Chris**, **Sarah**, **Karl-Erik**, **Steffen** and **Sabrina**.

Finally, I would like to thank my supportive, inspiring and wonderful friends for all the exciting things we have been through, including hikes, travelling, dinners, and dancing: **Aslaug**, **Alise**, **Hanna**, **Marte**, **Aasta**, **Iril**, **Maria**, **Mariya**, **Oliver**, **Ghaith**, **Lisa**, **Sîan**, **Arun**, **Tine**, **Lisbeth** and **Morten**. A special thanks to **Johan** for emotional support, motivation and for providing a lovely home office "up in the mountains" of Odda. I further would like to humbly thank all the wonderful and inspiring people I've encountered in the sports I do. They have kept me sane and down-to earth, even when I'm several thousand feet up in the air, hanging on to a parachute, or in the middle of a crowed kizomba dancing floor, in the late hours of the night.

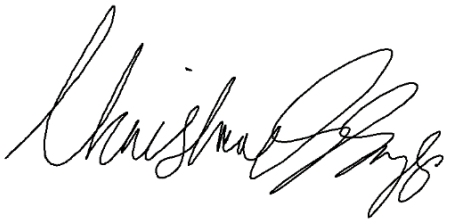
Writing a PhD is truly a teamwork-based project, and without all the support, assistance, and advice I have been so lucky to receive, this PhD would not have been possible. You put all your effort into your thesis, your time, your focus, your spark, your

frustration, your anger, your joy, and last but not least, your hope. Hoping that what you write will be high enough quality for publication, be validated, be honorable, be remembered and be reproduceable in future research.

"*Finishing a PhD is like finishing a group project where your partner made a ton of mistakes in the beginning of the assignment. Except your partner is just you 4 years ago.*"
John M. Mola, May 5th 2019

Bergen, August 2021

Christina Skår Saghaug



"All these described particles lay in a clear transparent medium, in which I have at times seen very prettily moving animalcules, some rather larger, others somewhat smaller than a blood corpuscle, and all of one and the same structure. Their bodies were somewhat longer than broad, and their belly, which was flattened, provided with several feet, with which they made such a movement through the clear medium and the globules that we might fancy we saw a pissabed running up against a wall. But although they made a rapid movement with their feet, yet they made but slow progress.", Anton Van Leeuwenhoek, 1681 [1].

SUMMARY

Metronidazole (MTZ) treatment-refractory *Giardia lamblia* infections is an emerging global problem. Recent studies show that prevalence rates of refractory infections are 10-20 % and can be as high as 46 %. The recent global increase of treatment refractory *Giardia* infections, suggest an emergence of strains with certain genetic inheritable traits. Up until now, no specific genetic marker of MTZ-resistance has been discovered in *Giardia* and most *in vitro* studies have utilized laboratory-derived sub-assemblage AI isolates, which rarely infects humans.

The overall aim of this thesis was to investigate and analyze single nucleotide variants (SNVs) in proteins that are responsible for activation or inactivation of MTZ, and involved in the detoxification of reactive oxidative species and reactive nitrosative species in clinical isolates of *Giardia*.

In this thesis we address the presence of genetic variation, especially amino acid altering, non-synonymous(ns) SNVs, of 29 genes. The variation was identified by Illumina whole genome sequencing of 20 recent, clinical, cultured sub-assemblage AII and assemblage B isolates, while cloning and Sanger sequencing was performed to identify alleles of the three genes flavohemoprotein (*gFlHb*), nitroreductase (*NR*) *1* and *2*, the two latter were explored in eight non-cultured clinical isolates as well. Further, *gFlHb* was investigated for its potential variable copy number by mining PacBio de novo assembled genomes in eight isolates.

Our findings disclose high frequencies of genetic variations in many of the 29 genes, where nsSNVs in assemblage B were especially high in ferredoxins (*fds*), *NR1* and *NR*2, thioredoxin peroxidase and NADH oxidase. *gFlHb* was found to be a variable copy number gene (2-6 copies/isolate) with high genetic variation. Moreover, we identified truncated *NR1* alleles in one MTZ refractory isolate, while two susceptible isolates, one homozygotic, had truncated *NR2* alleles.

The considerable genetic variation, coupled with findings of variable copies of *gFlHb* and dysfunctional NR proteins in recent clinical isolates of *Giardia* show a potential for identification of genetic markers for MTZ susceptibility/resistance. Future studies will help to decipher the effects of the mutations and could include functional assays and obtaining samples from a larger, more well-defined patient cohort.

ABBREVIATIONS

AI	Sub-assemblage WB
AII	Sub-assemblage DH
AMR	Antimicrobial resistance
ASH	Allelic sequence heterozygosity
B	Sub-assemblage GS
bg	Beta-giardin
bp	Base pair
CDS	The coding region of a gene
CNV	Copy number variation
DNA	Deoxyribonucleic acid
FAD	Flavin adenine dinucleotide
Fd	Ferredoxin
FDP	Flavodiiron protein
Fe-S	Iron-Sulfur
gdh	Glutamate dehydrogenase
gFlHb	Flavohemoprotein from *Giardia*
H_2O_2	Hydrogen peroxide
Indel	Insertions and deletions
HC	High confidence
LC	Low confidence
MIC	Minimum inhibitory concentration
MTZ	Metronidazole
NAD	Nicotidamide adenine dinucleotide
NO	Nitric oxide
NO_3^-	Nitrate
NR	Nitroreductase
nsSNV	Non-synonymous SNV (amino acid altering)

NTR-1	Oxygen-insensitive nitroreductase (also NR3)
NTZ	Nitazoxanide
O_2	Oxygen
PCR	Polymerase chain reaction
PFOR	Pyruvate ferredoxin oxidoreductase
qPCR	Quantitative PCR
RKI	Robert Koch-Institut
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
sSNV	Synonymous SNV (non-amino acid atering)
SOR	Desulfoferredoxin/superoxide reductase
T_m	Melting temperature
tpi	Triosephosphate isomerase
TrxR	Thioredoxin reductase
VSP	Variant-specific surface protein
WGS	Whole genome sequencing
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF PUBLICATIONS IN THIS THESIS

This thesis is based on three publications (one in manuscript format), which will be presented and referred to in the text as follows:

Paper I

Christina S. Saghaug, Christian Klotz, Juha P Kallio, Hans-Richard Brattbakk, Tomasz Stokowy, Toni Aebischer, Inari Kursula, Nina Langeland and Kurt Hanevik. "Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical *Giardia lamblia* assemblage A and B isolates".

-*Infect Drug Resist*. 2019:12 1221-1235

Paper II

Christina S. Saghaug, Christian Klotz, Toni Aebischer, Nina Langeland and Kurt Hanevik. "Genetic diversity of the flavohemoprotein gene of *Giardia lamblia* – evidence for high allelic heterozygosity and copy number variation".

-*Infect Drug Resist*. 2020:13 4531-4545

Paper III

Christina S. Saghaug, Astrid Gamlem, Juha Vahokoski, Christian Klotz, Toni Aebischer, Nina Langeland and Kurt Hanevik

"Genetic diversity in the metronidazole metabolism genes nitroreductase 1 and 2 in susceptible and refractory clinical samples of *Giardia lamblia* ".

-*In manuscript*

All of the published versions of the articles are reprinted with permission from the publishers.

LIST OF PUBLICATIONS OUTSIDE THE THESIS

Kurt Hanevik, Rasmus Bakken, Hans-Richard Brattbakk, **Christina S. Saghaug** and Nina Langeland. "Whole genome sequencing of clinical isolates of *Giardia lamblia"* -*Clin Microbiol Infect*. 2015 Feb;21(2):192.e1-3.

Christina S. Saghaug, Steinar Sørnes, Demitra Peirasmaki, Steffan Svärd, Nina Langeland and Kurt Hanevik. "Human Memory CD4+ T Cell Immune Responses against *Giardia lamblia*".
-*Clin Vaccine Immunol*. 2015 Sep 16;23(1):11-8

Matej Radunovic, Christian Klotz, **Christina S. Saghaug**, Hans-Richard Brattbakk, Toni Aebischer, Nina Langeland and Kurt Hanevik. "Genetic variation in potential *Giardia* vaccine candidates cyst wall protein 2 and α1-giardin".
-*Parasitol Re*s. 2017 Aug;116(8):2151-2158

1. INTRODUCTION

1.1 Preface

The two nuclei looked like dark deep eyes, while the median body resembled a little smile. I found myself gazing mesmerically into these hollow, dark and eye-like organelles, without being able to turn my gaze away. The cheeky smile of the parasite frowned upon me, and a great desire to investigate this little creature emerged within me. This fascinating protozoan parasite, Giardia lamblia, has been the topic of my master's degree, and now, lastly but not least, my PhD thesis. Wiser I now feel, but the parasite's smile is still haunting my mind, and in some mysterious ways it seems a little unreal.

1.2 History

1.2.1 It must be *Giardia*! A brief historical perspective

Giardia lamblia, was discovered 340 years ago by using a home-made hand lens on November the 4, 1681 [2]. It was the 'father of microbiology', the Dutchman Anton van Leeuwenhoek who discovered the parasite. Van Leeuwenhoek found *Giardia* in his own stool sample during an investigation he started, to find the cause of his ill stomach and diarrhea. *Giardia* was first and foremost described as an animalcule with paws [3, 4]. This was the first morphological description of the parasite, and *Giardia* was the first human infecting protozoan parasite to be identified [2, 3]. Further, in 1859, a Czech physician named Vilem Lambl would redescribe the parasite, and published self-made drawings of the protozoa [5] (see Figure 1). At this time point scientists disagreed whether the parasite was a true pathogen or just a commensal, i.e. an innocent organism merely colonizing the intestine [5]. The genus name *Giardia* was given by the researcher J. Künstler in 1882 and was given in honor of the zoologist Alfred M. Giard [6]. Not much later the species name lamblia was given to the parasite to honor Lambl (1888) [6]. A little while later, in 1915-16, clinical reports and research would indeed associate *Giardia lamblia* with diarrheal disease, thus being a pathogen, and the species name *Giardia lamblia* was introduced [5].

It is not yet been proven whether *Giardia* developed before or after the mitochondrial acquisition, and it has been proposed that *Giardia* acquired mitochondria, but lost them at a later phase during evolution [7].

There are several sub-species and assemblages of *Giardia*, some are specific to human hosts whereas others have animal hosts. The classification and species are further described in the taxonomy and zoonosis sections.

Two important publications have contributed to understanding the transmission of parasite cysts and culturing. One of the studies concluded that the infective dose of the parasite was as little as ten cysts [8], whereas the other managed to culture *Giardia* axenically [9].

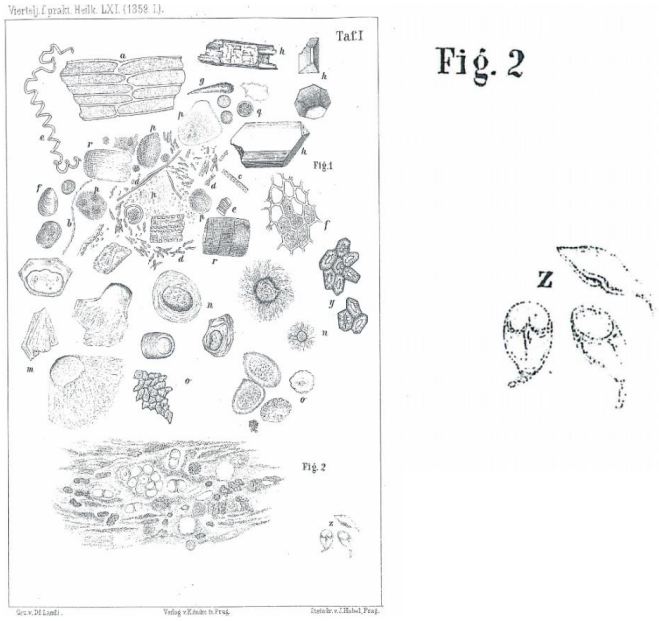


Figure 1: First microscopic drawing of *Giardia lamblia* by Vilem Lambl in 1859. Figure obtained from [10].

1.3 Taxonomy

Although the timespan since discovery has been long, the epidemiology and nomenclature of this antediluvian diverging organism can still be considered somewhat confusing, with its numerous subspecies still under debate [11, 12]. There are currently three species names used for the parasite, specifically: *G. intestinalis*, *G. lamblia* and *G duodenalis*. There is a disagreement relating to the classification of the parasite, and a

recent proposal of a revision of the nomenclature of *Giardia* was published a little over ten years ago [13]. The name *Giardia lamblia* or *Giardia* will be used to describe the parasite in this doctoral thesis.

Giardia belongs to the phylum metamonada, order of diplomonadida and the family of hexamitidae. *Giardia* further belongs to the binucleated flagellates group, which inhabit anaerobic and/or microaerophilic environments [6, 14]. The metabolism of the parasite somewhat resembles bacteria, and may offer exceptional opportunities to understand essential cellular pathways and cellular differentiations of eukaryotic cells [14].

1.3.1 Species and assemblages

The genus *Giardia* consist of several different flagellate species that may all be responsible for diarrheal disease and are classified based on their animal host and morphological appearance [15, 16]. As newer methods developed, such as polymerase chain reactions (PCR), it was possible to characterize different species based on genetic differences [15].

There are currently six species that can cause disease in animals, *Giardia agilis, Giardia ardeae, Giardia muris, Giardia microti, Giardia psittaci* and *Giardia lamblia*. *Giardia lamblia* has the broadest host range and can infect humans and mammals, *Giardia muris* will infect rodents, birds and reptiles while *Giardia agilis* infects amphibians [6, 17].

Giardia lamblia has a total of eight different genotypes, known as assemblages (A-H) divided according to single nucleotide variants (SNVs) [15, 18]. Assemblages A and B are responsible for human infections [6, 19, 20]. The assemblages may further be divided into sub-assemblages such as AI, AII, AIII and AIV and BI, BII, BIII and BIV [18, 21, 22]. Sub-assemblage AII nearly exclusively infects humans [17, 22]. The sub-division of assemblage B is not as specific as for assemblage A, most likely due to inconsistency in DNA sequence analysis and more genetic variability [17]. Further, assemblages C and D has more frequently been found in dogs and canines, assemblage E in cattle, sheep and pigs, while assemblages F and G mostly have been found in cats and rodents and assemblage H in seals [17].

A well-known difference between assemblage A and B is their adaptability of *in vitro* and *in vivo* growth, where assemblage A seems to grow at a steadier pace than assemblage B. Further, assemblage B has been associated with a higher risk of infecting HIV patients, and assemblage B was more pathogenic than assemblage A in a gerbil model [23-25]. Additionally, large genetic discrepancies including chromosomal and genomic conformations found in different genotypes of *Giardia,* have led to a debate whether they should be divided into different species, rather than assemblages [13, 14, 26, 27]. Assemblage B is more heterozygotic than assemblage A (all subtypes), which has been noticed in whole genome sequenced *Giardia* [26, 28, 29].

1.4 Epidemiology

1.4.1 Prevalence, cases and outbreaks

There are several waterborne pathogens that may cause diarrheal disease in humans in industrialized and LMIC countries. Diarrheal disease is one of the leading causes of deaths in children under the age of five, and as many as 1.66 million deaths are reported annually (2016) [30].

Giardia has been estimated to cause up to ~300 million symptomatic human infections annually [31]. The prevalence rate of infection in the industrialized world is estimated to be between 2 % - 5 %, while the prevalence is considerably higher in LMIC countries (20 % - 30 %) [32, 33]. Giardiasis is more prevalent in children under the age of five, compared to other age groups [32, 34].

Giardia has often been identified in waterborne outbreaks and the infection also spreads through food contaminated with *Giardia* cysts, person-to-person contact (fecal-oral route), and it may also spread via animals as a zoonosis [17, 35]. Transmission of *Giardia,* and the infection known as giardiasis due to contaminated drinking water, was first reported in the 1970s in the USA [2, 36]. Sporadic outbreaks may occur in industrialized countries, such as the outbreak in Bergen, Western Norway in 2004 where an estimated 5000 cases were reported [37]. Cases of giardiasis in industrialized countries are often acquired through travelling in *Giardia* endemic areas [38].

The Norwegian Scientific Committee for Food and Environment (VKM)

reported between 179 and 577 cases annually from 2009 to 2019, where 61 % of cases were reported to be acquired through travelling abroad, 21 % were acquired in Norway and 18 % had an unknown route of transmission [39].

1.4.2 Zoonosis

Giardia lamblia assemblages A and B can infect humans and various animals [5, 27, 40], while assemblages C-H may infect domestic animals and pets (cats, dogs and cattle) [27]. Former publications have addressed the potential zoonosis of *Giardia* [17, 22, 41-43], where contact with domestic animals has been associated with human *Giardia* zoonosis. The infection rate of human giardiasis has been found to be 23 % higher in rural areas of New Zealand compared to more urban areas [17]. Further, farm visitations have been associated with higher frequencies of giardiasis in the United Kingdom [17]. A study investigating *Giardia* infection among humans and cattle sharing the same habitat (farm) in India identified the same assemblage (AI) in both [44]. Further, dogs have been proposed to be a reservoir for *Giardia* in rural temple areas of Bangkok, Thailand, where 79 % of *Giardia* infections in dogs and 73 % of infections in humans were genotyped as assemblage A [17]. Finally, beavers have been identified as a potential source of zoonosis by contaminating surface water with cysts, leading to “beaver-fever” (see Figure 2) [45]. Still, we do not know if the beavers are the original contaminating source of *Giardia*, or if they are more prone to infections from fecal material containing cysts from humans or domestic animals [17].

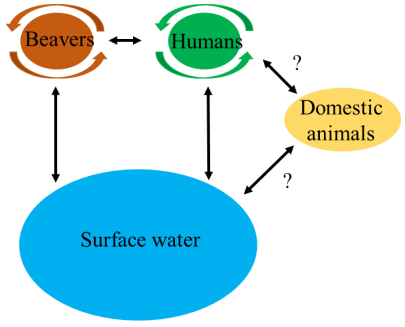


Figure 2: Illustration of zoonotic transmission of *Giardia*. Figure is made by Christina Saghaug 2021 and inspired by Tsui et al 2018 [45].

1.4.3 Mixed assemblages infections

Most human infections are caused by sub-assemblage AII and assemblage B, where assemblage B has been more frequently reported [15, 46, 47]. Some studies have found presence of mixed infections of the two assemblages, A and B, particularly in LMIC countries [47-49]. Still, mixed infections could potentially be underreported as the detection method used, PCR, may be biased towards detection of the most abundant assemblage during amplification [46]. A recent study assessed if the beta-giardin (*bg*) gene could be used for detecting mixed *Giardia* infections by using either Sanger sequencing or next generation sequencing (NGS) and found that the latter sequencing method was superior for detecting mixed infections [50].

1.5 Biology of *Giardia*

Giardia is a eukaryote with certain prokaryotic characteristics, including bacterial-like metabolic enzymes, that most likely have been acquired by lateral transfer, and it lacks mitochondria and golgi apparatus [6, 51]. The parasite may be regarded an anaerobic protozoan, but the term microaerophilic is more appropriate, as it is able to grow and thrive in the upper intestinal lining where O_2 levels may be up to 60 µM [52, 53].

1.5.1 The cyst and the trophozoite

Giardia may be termed a luminal extracellular parasite, as it infects the upper intestinal tract of its hosts, where it adheres to the epithelial microvilli, but does not invade the cells or mucosal tissue in most cases [54]. *Giardia* can vary between two unique morphological forms during transmission and infection. The immotile cyst is responsible for the transmission, while the motile trophozoite is the disease-causing form [14] (Figure 3).

The cyst measures 12 µm in length and 7 µm in width. The cyst is covered by a characteristic smooth and clear substance known as cyst wall. The cyst wall will protect *Giardia* from environmental changes, and it may survive for several months outside of the host, if the temperature is cold and the conditions are wet [14, 55]. *Giardia* is

relatively resistant to water treatment such as chlorine [56]. The cyst contains total of four nuclei, fragments of its disc and axonemes that resemble microtubules [6, 57].

The *Giardia* trophozoite is approximately 15 µm long and 8 µm wide and is shaped like a pear. Further, the trophozoite has organelles and structures such as flagella, an adhesive/ventral disc and mitosomes (Figure 3) [57]. The adhesive disc facilitates attachment to mucosal surfaces of the small intestine of a host, where it eventually may lead to a local disruption of the intestinal mucosa [54, 57].

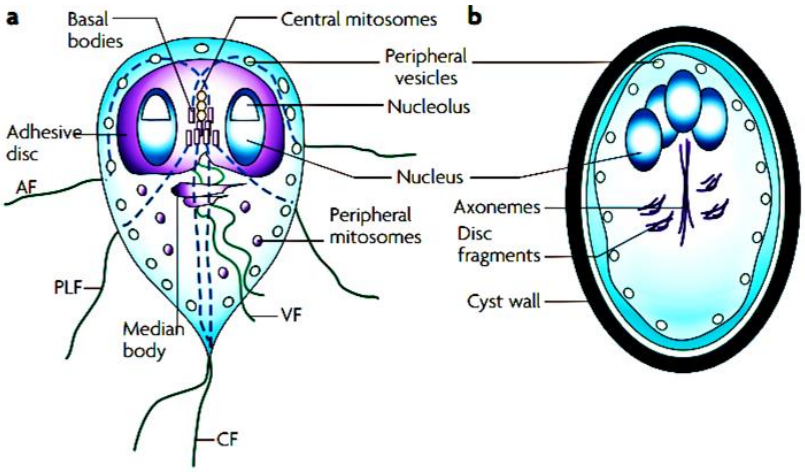


Figure 3: The trophozoite form of *Giardia* (left) and the cyst form of *Giardia* (right).
The trophozoite has two nuclei, an adhesive disk, mitosomes, four pairs of flagella: the anterior flagella (AF), ventral flagella (VF), posterior/lateral flagella (PLF) and caudal flagella (CF). The illustration is made by Ankarklev et al 2010 [14] and colors modified by Saghaug 2021.

1.5.2 Life cycle of Giardia

An infection with *Giardia* starts by ingestion of cysts from contaminated water/food or direct contact with an infected individual. The life cycle of the parasite is represented in Figure 4.

1.5.2.1 Excystation and encystation

Two different processes are crucial in the life cycle of the parasite, namely the excystation where a *Giardia* cyst differentiates to a trophozoite, and the encystation

where a trophozoite differentiates to a cyst [14]. Excystation of *Giardia* cysts to its trophozoite form is thought to first be triggered by the low pH of the gastric acid in the stomach, followed by exposure to the pancreatic secretions of the small intestine while passing through [14]. Trophozoites reproduce by cytokinesis, and are thought to reproduce asexually [58, 59]. However, recombination of chromosomes and meiosis proteins, usually associated with sexual reproduction, have been found [58, 60, 61].

As the trophozoites move further down the intestines, encystation will occur. The encystation is triggered by higher pH and bile salts [5]. The cysts can now be shed in feces and spread to a new host, or even cause a re-infection of the host (see life cycle Figure 4) [5].

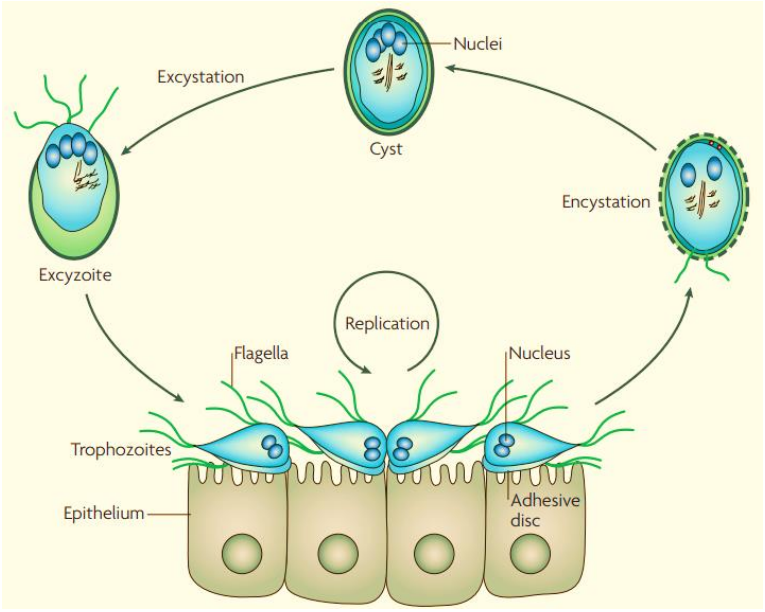


Figure 4: Life cycle of *Giardia*

Giardia cysts will go through excystation when exposed to the low pH of the gastric acid of the host, via the short-lived excyzoite form. The excyzoite will then divide two times, resulting in four trophozoites with two diploid nuclei each. The trophozoite can attach to the intestinal epithelial cells, replicate and cause infection of the host. Encystation will be triggered if trophozoites migrate further down the small intestine. The cell will go through a DNA replication and stop at the G2 stage, leaving the cell with two tetraploid nuclei. At a later stage of encystation these two nuclei will divide and result in four diploid nuclei [14]. Illustration is made by Ankarklev et al 2010 [14] and colors are modified by Saghaug 2021.

1.5.3 The antioxidant system of the parasite

Giardia is a microaerophilic organism, meaning it thrives in an environment with low levels of oxygen. As the gut habitat in the host release substances such as molecular oxygen (O_2), hydrogen peroxide (H_2O_2), superoxide ($O_2^{\cdot}$-) and nitric oxide (NO), the parasite must be able to withstand these reactive oxygen species (ROS) and other harmful substances such as reactive nitrogen species (RNS) [62, 63]. *Giardia* lacks common eukaryotic antioxidant enzymes such as peroxidase, glutathione peroxidase, catalase and superoxide dismutase [64]. Alternative antioxidant enzymes exist in *Giardia* to help it neutralize free radicals, including flavoprotein, desulfoferredoxin/superoxide reductase (SOR), peroxiredoxins (gPxrds), NADH oxidase and flavohemoprotein (gFlHb) [62, 65] (Figure 5). *Giardia* also has large quantities of cytosolic cysteine (a low-molecular weight thiol), that probably plays an important role in detoxification of oxygen [66], and activated MTZ has been hypothesized to be able to react with cysteines in cytosol and create adducts [67, 68]. Indeed, cysteine has been found to be able to protect *Giardia* trophozoites from thiol-blocking chemicals, thus indicating a potential role as a reducing agent for the protection of important thiol groups [69].

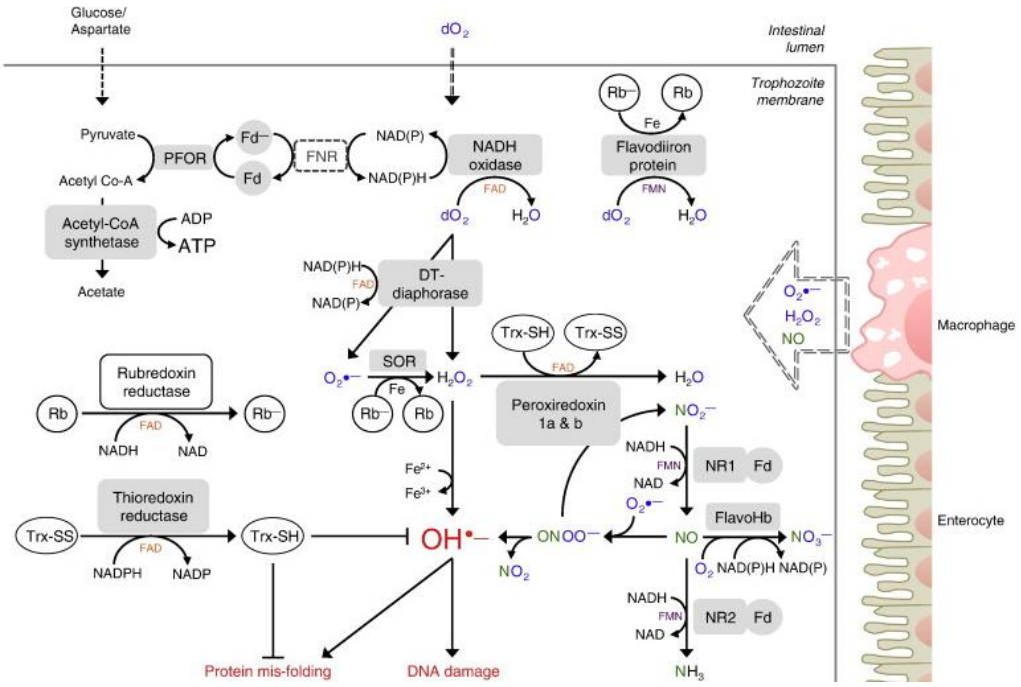


Figure 5: The antioxidant system and ATP production in *Giardia lamblia* during an infection.
Molecules from the intestinal lumen from the host such as dissolved molecular oxygen (here dO_2), macrophage-mediated H_2O_2, $O_2^{\cdot-}$ and NO will reach into *Giardia* by diffusion. NADH and NADPH (NAD(P)H) provide electrons to reduce dO_2, NO and other reactive species and radicals. Functionally characterized enzymes are shaded in gray, while enzymes with a black outline have not yet been characterized in *Giardia* (functionally analogous enzymes are assumed to exist based on *Escherichia coli* enzymes). The enzyme FNR which is represented with a perforated outline, has neither genomic nor functional evidence to exist in *Giardia.* Addionaly abbreviations outside the thesis: Fe^{2+}: ferrous iron; Fe^{3+}: ferric iron; NAD(P): oxidized nicotinamide adenine dinucleotide (phosphate); FAD: flavin adenine dinucleotide; FMN: flavin mononucleotide; FNR: ferredoxin NAD(P) reductase; Rb: rubredoxin; Trx-SH/SS: reduced/oxidized thioredoxin; NO_2^-: nitrite; NO_3^-: nitrate; $ONOO^-$: peroxinitrite; $OH^{\cdot-}$: hydroxyl radical. Figure is created by Ansell et al 2015 [70].

1.5.4 Flavohemoprotein

gFlHb was characterized a little over a decade ago and is an important RNS detoxification enzyme [65, 71]. The gFlHb enzyme has been predicted to have nitric dioxygenase characteristics, meaning the enzyme may induce the formation of NO to the less harmful nitrate (NO_3-) by utilizing O_2 as a co-factor [72, 73] (Figure 6). gFlHb may supposedly have NADH/NAD(P)H oxidase characteristics, similar to the

antioxidant enzyme flavodiiron protein (FDP), where FDP can catalyze O_2 to H_2O in environments with low levels of NO [71]. Interestingly, current genome data suggest that the *gFlHb* gene may be missing in *Giardia* assemblage C and D, indicating that different *Giardia* species have other ways of neutralizing NO and perhaps O_2 [74].

Several studies have found up-regulation of gFlHb during oxidative stress exposure (e.g., O_2 and H_2O_2) and also during nitrosative stress exposure [63, 65, 75]. Further, the protein levels of gFlHb have been found to be increased during MTZ exposure in a *Giardia* strain (713) resistant to MTZ and nitazoxanide (NTZ) [76]. The transcriptional induction of gFlHb, as an antioxidant stress response, may represent an alternative MTZ detoxification strategy [76].

Bulky imidazoles such as azoles, miconazole, econazole and ketoconazole have shown to be inhibited by enzymes with similar functions to gFlHb (FlHb homologs in bacteria), by being able to bind to the proteins' "heme pockets" and subsequently create ROS [77, 78]. This raises a hypothetical suspicion that gFlHb possibly could be linked to *Giardia*'s ability to handle and withstand the harmful effects of MTZ [77, 78].

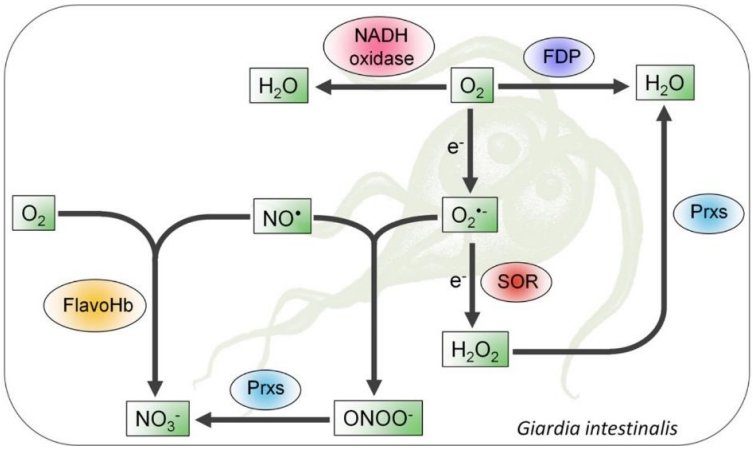


Figure 6: Important antioxidant enzymes in *Giardia lamblia.*

In addition to NADH oxidase, the antioxidant enzymes include flavodiiron proten (FDP), flavohemoprotein (gFlHb), superoxide reductase (SOR) peroxiredoxins (Prxs). Illustration created by Mastronicola et al 2016 [62].

1.6 Genetics in *Giardia*

The four nucleotides A, G, T and C make up the sequence of the DNA. A combination of three nucleotides is called a codon and encodes one specific amino acid. Nucleotides encoding amino acids of a protein, with a established start and stop codon, is known as the coding sequence (CDS) [79]. The start codon (usually ATG) marks the site at which translation of DNA into protein begins, while a stop codon (usually TAA and TAG) marks the site at which the translation ends.

1.6.1 The reference genomes of *Giardia lamblia*

Cysts from patient samples, specifically assemblage B, have proven difficult or in some cases impossible to culture [80, 81]. There exist other methods for obtaining DNA from *Giardia* than culturing, e.g., immunomagnetic separation (IMS) and sucrose filtration, which enables cysts from clinical samples to be isolated [82]. The two methods may be combined and can result in good quality DNA, which can be used for whole genome sequencing (WGS) [82].

WGS and comparison of genomes are beneficial methods to decipher genetic variation between different species and assemblages of *Giardia* [26]. Up until now, five of the eight *Giardia lamblia* assemblages have been sequenced. At first, the genome of *Giardia* the sub-assemblage AI (strain WB) was available in 2007 [83]. Secondly, *Giardia* assemblage B (strain GS) was first sequenced in 2009 [29], then in 2013 (strain GS-B) [26] and in 2015 (isolate BAH15c1) [84], followed by assemblage A isolates (DH in 2013 [26] and AS175 and AS98 in 2015 [85]). Further, a pig isolate (P15) from assemblage E was sequenced in 2013 [26], and more recently assemblages C and D (dog) in 2019 [74]. Last year, the sub-assemblage AI was re-sequenced, resulting in longer contigs (overlapping DNA segments representing a consensus of DNA) compared to the older genomes from 2013. The genes have also been better annotated in the recent genome [86]. The new WB reference genome is more complete and assembles into near-complete chromosomes.

The genomes of the WGS *Giardia* assemblages vary in size: 11.7 Megabases (MB) for sub-assemblage AI, 10.7 MB for sub-assemblage AII and 12.0 MB for

assemblage B [26, 83]. The published genomes of *Giardia* are fragmented into hundreds of contigs, and this consequently impacts the interpretation of how the chromosomes and genes are orchestrated [26, 29, 84-86]. AI has 306 contigs, sub-assemblage AII has 239 contigs, while B consists of 544 contigs. Further, 5147 annotated genes can be found in the assembled AII, whereas the assembled B contains 6098 annotated genes [26].

In recent years, new sequencing techniques have been introduced, making it possible to obtain longer reads. One of the newer sequencing platforms is Pacific Biosciences (PacBio). with an average read length of 13.5 kilobases (kb) [87-89] and Oxford Nanopore Technologies (ONT) where the whole cDNA molecule may be sequenced end-to-end [90]. Illumina sequencing is more accurate than the long-read sequencing methods [91], and PacBio and Nanopore may be combined with short reads from Illumina to obtain higher quality of the sequences [86, 92]. Other sequencing methods include DNBSEQ ™ Rapid Whole Genome Sequencing (rWGS), Low-Pass Whole Genome Sequencing, and high-quality amplification for single cell sequencing, which was done on cryptosporidium in 2016 [93], and the more recent technology, 10x genomics and BD Rhapsody, where single cells can be sequenced [94].

1.6.2 Ploidity and chromosomes

A eucaryote organism's genome is normally distributed into several chromosomes. Ploidy is a term that is used to describe how many sets of chromosomes an organism has (haploid =1 N, diploid = 2N etc). *Giardia* is considered to be of functionally tetraploid nature [95], because the trophozoite harbors two diploid nuclei that are estimated to have equal size (see Figure 7), each containing five chromosome pairs [6, 96-98]. Single-copy genes may therefore exist in up to four different versions, termed alleles, in each pure *Giardia* isolate [58, 99].

However, studies have shown that extra or even missing copies of chromosomes or genes, aneuploidy, is common in individual trophozoites [29, 96, 99, 100].

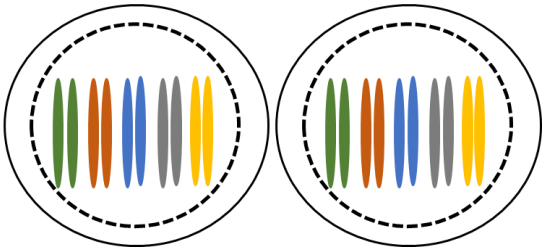


Figure 7: The two diploid nuclei of *Giardia*.

Each of the nuclei contains five chromosome pairs. Figure is made by Christina Saghaug 2021 and inspired by Gamlem 2020 [101].

1.6.3 Genetic variation in *Giardia*

1.6.3.1 Introduction to mutations

Genes are regions of DNA coding for proteins and consists of sequences of nucleotides. Genetic variation will most frequently be present in the form of single nucleotide polymorphisms (SNPs) and point mutations/single nucleotide variants (SNVs) [102, 103]. This type of genetic variation means that a DNA sequence in one genome either diverges between members of the same or different species or diverges from an established reference genome. For a mutation to qualify for the term SNP, the variant should be recognized as common, and is usually present in at least 1% of the population. The word SNV is used if the frequency is unknown [102].

There are two types of single nucleotide variations (SNVs) that can occur in the CDS of a gene. One is synonymous (sSNV) not leading to an amino acid change, and the other is non-synonymous (nsSNV) causing amino acid changes. A nsSNV can be a missense mutation, where the amino acid in of a protein sequence is changed, or it may result in a premature stop codon, usually called a nonsense mutation (change into a stop codon) resulting in a shorter and unfinished protein product [104], or it may also be a nonstop mutation, meaning the stop codon is changed, resulting in a longer protein product. Insertions (addition of one or more nucleotides to the CDS) and deletions (removal of one or more nucleotides) (abbreviated indels) may also occur. Indels may contribute to severe downstream changes in the protein, as the amino acids are shifted and may lead to stop codons and truncated proteins. An amino acid substitution or an

indel may, or may not affect the protein function, and may be investigated by using predictive tools such as PROVEAN, NCBI NR (non-redundant) protein databases and UniProtKB/Swiss-Prot Protein database [105].

An allele is a variant form of a specific gene located at the same position (locus) of a chromosome, that may be identical (homozygous) or different due to mutations (heterozygous) [106]. Alleles on the same chromosome are normally inherited together in the same order, meaning inheritance of a cluster of SNVs, and is referred to as a haplotype [107].

1.6.3.2 Genetic variation

The genetic diversity, especially in *Giardia* assemblage B, has been found for some of the genes used for genotyping and housekeeping in the parasite and include: *bg*, glutamate dehydrogenase (*gdh*), elongation factor 1-alpha (*ef*-1), mlh1 (*mlh*), the FLORF-C4 (*C4*) and triosephosphate isomerase (*tpi*) [15 , 108, 109]. Most studies investigating full length or parts of these genes have done so by direct PCR investigation of the genes, while other studies have used a clone/subclone based approach [60, 110-114]. These studies, have reported high numbers of nsSNVs, especially for B isolates, except for two studies which reported low allelic sequence divergence, one for axenic cultured assemblage B parasites and one for cloned clinical samples of *Giardia* [60, 113]. The observed nsSNVs in the *gdh*, *TPI* and *bg* genes have been validated in other studies and mostly lead to conservative amino acid changes, probably with small or no effect on the protein function or structure [111, 115].

1.7 Giardiasis

1.7.1 The disease

The *Giardia* trophozoites absorb nutrients from the lumen in the upper part of the small intestine of the host [54]. Trophozoites are thus not associated with invasive infection, as the parasite does not enter and penetrate the epithelial cells or neighboring tissue of the host, or release any known toxins [14, 116]. After ingestion of the cysts, an incubation period of 6 to 15 (average 7 to 10) days will follow before the *Giardia*

infection giardiasis is established [14]. The disease is quite heterogenous and may vary from acute symptomatic to asymptomatic [32, 40]. Symptomatic giardiasis is frequently recognized as acute, but can turn into a chronic infection (lasting> 2 months) persisting for months or years, both with and without symptoms [14, 117, 118]. Most commonly the symptoms stop when the parasite is cleared, either through host immune response with or without medical treatment [32]. Even if the infection can be self-limiting or cured with antibiotics, it may influence the quality of life for one individual to a great extent [20].

Symptoms of a *Giardia* infection include persistent and watery diarrhea, stomach pain/cramps, abdominal discomfort such as bloating, nausea, dehydration, weight loss, in some cases malabsorption, and the infection may even lead to post-infectious sequelae [17, 40, 119]. Further, the infection may cause damages to the small intestinal surface, where atrophy of the villi has been observed by microscopy, as well as deficiency of intestinal enzymes [120, 121]. However, *Giardia* has also been hypothesized to contribute some protection against diarrhea from other pathogens in young children, as it has been observed more frequently in asymptomatic cases than in diarrhea cases [122, 123]. Complications of giardiasis include micro- and macronutrients deficiencies and even retarded growth and development [124-127].

More knowledge regarding post-infectious sequelae of *Giardia* was obtained after the 2004 Bergen outbreak, where irritable bowel syndrome (IBS), chronic fatigue and self-reported lactose intolerance lasting for years after eradication of the parasite was reported [128-130].

1.7.2 Diagnostics

Historically, microscopic analysis looking for cysts in fecal samples has been the most common diagnostic tools to determine a *Giardia* infection [40, 131]. This method is still used, although not as sensitive as the PCR where amplification and detection of the 16S-like gene has shown to have both a higher specificity and sensitivity than other diagnostics and may be used to differentiate between assemblages or sub-assemblages [131, 132]. There are additional ways to detect a *Giardia* infection as well, such as rapid

tests that react to *Giardia* antigens in feces [133].

1.7.3 Host immunity

Protection against *Giardia* is most likely dependent on antibodies produced by B cells and T cell mediated immune responses during the host and pathogen interaction [33, 134]. Indeed, individuals who have encountered the parasite before have lower infection rates, indicative of acquired immunity [135, 136]. Subsequently T cell immune responses have been found to be present up to five years after an initial infection [137].

Studies on mice have shown that the $CD4^+$ T cell response against *Giardia* is crucial, as $CD4^+$ T cells deficient mice cannot control the infection properly, while the antibody response did not seem to be as essential for the regulation of an acute infection [138]. Further interferon gamma (IFN-γ) was found to be released by human intestinal and blood $CD4^+$ T cells when stimulated with trophozoites [139]. While the cytokine IL-17 has been associated with a protective role for both mice and humans [140]. Antibodies are mostly mediated against the variable surface proteins (VSPs) that cover the parasite. However, *Giardia* has the ability to evade this immune response by shifting the VSPs by expressing different ones [141].

Still, individuals with immunodeficiency such as the common variable immunodeficiency are more prone to chronic infection with *Giardia* indicating that decreased B cell mediated antibody production is hampered by less effective $CD4^+$ T cells [138, 142].

1.8 Treatment of *Giardia* infection

In some cases, giardiasis will be self-limiting [143], and individuals may use antimotility drugs, hydration and electrolytes to reduce the symptoms and reduce the severity of diarrhea [101]. There is no effective and approved vaccine to protect an individual from giardiasis, and prevention is based on sanitary and hygienic measures while treatment is dependent on antiparasitic drugs [70]. There exist several drugs that are effective against *Giardia*, but only some of them are commonly used in practice or

licensed in Norway [144, 145].

1.8.1 Metronidazole

Giardiasis is usually treated with the synthetic 5-nitroimidazole drug MTZ, which is a common first line drug choice globally [144, 145]. MTZ has a treatment efficacy of 70-100 % [70, 146]. MTZ was first discovered in 1959, is derived from *Streptomyces* sp and was first used for giardiasis in 1961 [68]. MTZ is also used to treat other organisms such as *Trichomonas vaginalis*, *Entamoeba histolytica*, *Clostridium difficile* and *Helicobacter pylori* [68, 101, 147, 148].

The common treatment regime of MTZ for giardiasis is 500 mg one to three times daily for five to ten days [144, 149]. Side effects of MTZ are usually mild, where the most common ones are nausea, metallic taste in the mouth, abdominal pain and diarrhea [101]. Usage of MTZ during pregnancy is controversial and should not be used in the second and third trimester due to caution [150], however, more recent updates do not show that the drug increases teratogenicity [101, 151].

1.8.2 Other available treatments

In cases where MTZ treatment fails, combination therapy or other drugs are recommended [144, 145, 152]. Several treatments may be used and includes: tinidazole as a single dose treatment (efficacy 74-100 %), quinacrine (efficacy 92-100 %), albendazole (efficacy 24-100 %), furazolidone (efficacy 80-100 %), NTZ (efficacy ~78 %), paromomycin (efficacy 55-90 %) and quinacrine (efficacy 90-100%) [145, 153, 154]. For pregnant women paromomycin has been recommended [145], while NTZ is an option in children [154]. Combination therapy may be used if solo treatment with MTZ fails, and include albendazole in combination with MTZ as a common second treatment option for refractory cases. A third line choice may be quinacrine alone, or in combination with MTZ, and is usually an effective treatment [145, 152, 153, 155-158].

In 2019 a new class of drug; aminoguanidines related to robenidine, against *Giardia* showed to have a potential as an antigiardial agent [159] Auranofin has also been proposed to be a potential agent to be used in treatment-resistant giardiasis [64].

1.8.3 Unscrambling nitroheterocyclic drugs' mode of action

The intracellular environment of *Giardia* is highly reduced in order to manage oxidative stress, and this environment is partly responsible for the susceptibility to the nitroheterocyclics (MTZ, ornidazole, tinidazole, C17, nitrothiazolide, nitrofuran, furazolidone and NTZ) [70]. The nitroheterocyclics are all 'redox-active' prodrugs with low redox potentials [70]. The drugs are thought to reach the inside of the cell by passive diffusion, before they will be activated enzymatically. MTZ is the best-studied nitroheterocyclic drug and will be explained in further details.

MTZ is dependent on intracellular reduction (activation) for successful effect [160] (see Figure 8 for schematic representation). The activation of MTZ has been found to occur by three (maybe even four) main enzymes: the nitroreductase 1 (NR1) [161], pyruvate ferredoxin oxidoreductase (PFOR) 1 and 2 (if same effects in the cell as PFOR1) [162] and the enzyme thioredoxin reductase which is associated with cycling of thiol [163]. During activation; reduction, of MTZ at its nitro group, toxic intermediates are formed (see Figure 9) [144]. The drug is considered to be working most efficiently in microaerophilic or anaerobic environments, due to MTZ's risk of conversion back to its prodrug form if O_2 is present. This mechanism is known as 'futile cycling' [160, 164].

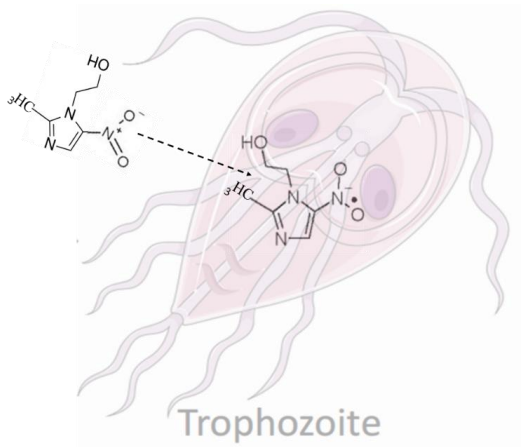


Figure 8: Schematic representation of MTZ activation from a prodrug into a free radical toxic form. Figure is made by Christina Saghaug 2021.

The toxic intermediates are thought cause several effects in the cell, such as disruption of the DNA, oxidative stress and ultimately cell death [64, 149, 165]. Further the toxic effects of MTZ include the generation of a nitro anion which is responsible for the oxidative stress response [70]. This anion may react with free thiol groups or proteins with the capacity of creating cysteine adducts. Additionally, MTZ may lead to degradation of elongation factor 1-gamma, which is an important enzyme in protein translation [166]. MTZ probably has a pleiotropic mode of action, as large numbers of molecules in the cell are affected by the drug, rather than a specific target, that other antimicrobials may have [101, 167, 168].

A recent study has evaluated MTZ for its antioxidant activity in *in vitro* global systems. No antioxidant activity was found, though it was found to possibly affect the homeostasis of ROS-independent mechanisms by reducing free radical reactions that further may stop the "respiratory burst" defence system in an organism [169].

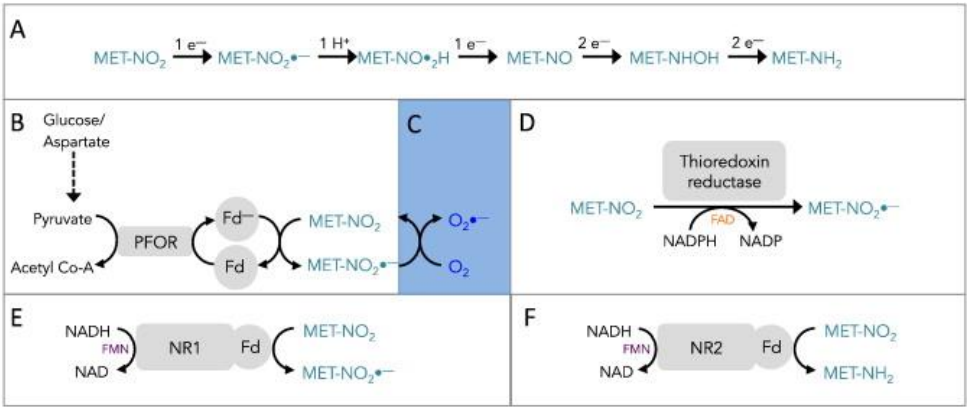


Figure 9: Activation of metronidazole.

The activation pathway of MTZ (MET). **A**: MET-NO_2 is the prodrug form. The different intermediates of MET are presented where the fully reduced form is the inert MET-NH_2. **B**: The pyruvate ferredoxin reductase (PFOR) enzyme's activation of MTZ, where the co-factors ferredoxins (fd) are involved. **C**: free radicals of oxygen created during activation. **D**: Thioredoxin reductase's (TrxR) activation of MTZ, where NADPH is an electron donor. **E**: NR1 activating MTZ and **F**: NR2 fully metabolizing MTZ to inactive form. The NRs obtain electrons from the co-factors NADH and FMN. The Figure is made by Ansell et al 2015 [70].

1.8.4 Nitroreductases

Nitroreductases (NRs) are a class of enzymes responsible for reducing nitro groups and other nitrogenous compounds, such as R-NO, R-NO_2- and R-NO_3- in *Giardia* [170-172]. The genes have most likely been obtained from anaerobic bacteria or archaebacteria by a process known as lateral transfer [51, 170].

1.8.4.1 The activator vs. the inactivator

There are several different nitroreductases in *Giardia*, and there are two important NRs, NR1 and NR2, that are considered to be paralogs with different capacities of MTZ metabolism [173] (see Figure 10). NR1 activates MTZ by partial reduction to reactive toxic metabolites, whereas NR2 seems to have a protective effect by fully reducing MTZ with six electrons to an inert and non-toxic form of the antibiotic [171, 173-175] (See Figure 9A).

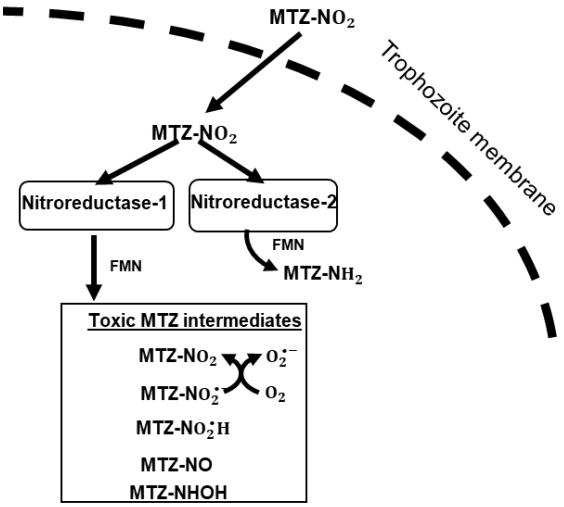


Figure 10: Activation and inactivation of MTZ by the two nitroreductases in *Giardia*. Figure is created by Christina Saghaug 2021 and inspired by [28].

In turn, down-regulated NR1 expression in MTZ-resistant laboratory cell lines has been found, while NR2 expression has been found to be up-regulated [68]. Further, *E. coli* expressing both NR enzymes from *Giardia* is not susceptible to MTZ, indicating that

NR2's MTZ inactivating capabilities surpasses the NR1's activating ones [173]. Overexpression of NR2 likely makes the parasite more tolerant to MTZ [160]. *Giardia* trophozoites overexpressing NR1 has been found to have a higher susceptibility to the two nitro drugs NTZ and MTZ [174].

The NR enzymes both contain cysteine-rich domains which has been demonstrated to be ferredoxin (fd) domains that can bind iron-sulfur (Fe-S) clusters [51]. The fd-domains are thought to be important in the activation process of MTZ in *Giardia*. Activation of MTZ can happen after the drug's nitro group is reduced by electrons mediated by the activated/reduced fd domain of the NR [172].

Further, a "third" giardial nitroreductase, (oxygen insensitive nitroreductase (NTR-1 or NR3)), without the characteristic fd domains, that has been shown to increase the susceptibility during protein expression in disk diffusion assays of MTZ where *E. coli* have been used [170]. Still, it is not known if this enzyme has a role in MTZ activation, although, it may contribute to the protection against oxidative- and nitrosative stress, due to the enzyme's O_2 insensitivity. Interestingly, the enzyme has been found to be up-regulated during MTZ exposure [69].

1.8.5 Other metronidazole activating enzymes

1.8.5.1 Pyruvate ferredoxin oxidoreductases

There are two different PFOR genes in *Giardia:* PFOR1 and PFOR2. PFOR1 has been sequenced and purified in one study [162] and the enzyme has been inhibited by a hammerhead ribozyme study a while ago [176]. The PFOR gene name is often used generically, and may be associated with the PFOR1 gene [162]. Still, the PFOR2 may potentially be involved in the activation of MTZ as well [162]. The PFOR is an essential enzyme in the oxidative decarboxylation, where pyruvate is decarboxylated to acetyl-CoA and CO_2 during the energy production in *Giardia*. The PFOR functions as an initial enzyme in electron transport in the parasite, which is crucial for MTZ activation [177]. Both of the PFOR enzymes contain fd domains capable of forming (Fe-S) clusters [177]. The PFOR co-factor, fd protein, is reduced during the activation of MTZ [177, 178]. Former studies have proposed a correlation between down-regulation of PFOR and increased MTZ resistance, while other studies have not been

able to confirm this hypothesis [162, 176, 177, 179].

1.8.5.2 Thioredoxin reductase

Thioredoxin reductase (TrxR) has been proposed to display disulphide reduction and to have NADPH oxidase activites. The enzyme is thought to use NADH as an electron donor during the MTZ activation [162, 180]. In addition to being a MTZ metabolizing enzyme, TrxR may be targeted and inhibited by MTZ and the drug's reduced intermediates and be a target for auranofin [163, 180, 181]. In the parasite *T. vaginalis* MTZ has been shown to target TrxR and inhibit the function of the enzyme [144, 180].

1.9 Refractory *Giardia* and ongoing research

1.9.1 Refractory *Giardia*

World health organization (WHO) has defined antimicrobial resistance (AMR) as "Antimicrobial Resistance (AMR) occurs when bacteria, viruses, fungi and parasites change over time and no longer respond to medicines making infections harder to treat and increasing the risk of disease spread, severe illness and death". A result of AMR is that standard medical treatments become less effective, and infections will persist and have a greater risk of spreading to others [182]. The term "refractory" will be used in this thesis to describe laboratory confirmed *Giardia* infections not responding to MTZ treatment by *Giardia* being eradicated from stool samples. Re-infection of *Giardia* after a successful treatment or recurrent infection due to low patient compliance will not be termed as refractory *Giardia.*

MTZ resistance in *Giardia* is an emerging problem**,** and an increasing rate of nitroimidazole-treatment refractory *Giardia* infections has been reported over the last few years [32], especially from travelers coming back from different regions Asia [149]. Treatment of *Giardia* may be hampered by the increasing number of treatment-failures and several studies have shown that the prevalence of MTZ-refractory *Giardia* is usually 10-20 % [157, 183, 184] and even reaching as high as 46 % in some cases [184]. A retrospective study of nitroimidazole-refractory giardiasis in returned travelers was carried out at the Hospital for Tropical Diseases, London between 2011 and 2013 [156].

This study showed a rapid increase of treatment refractory giardiasis. A total of 70 % of the travelers with refractory *Giardia* had been to regions of India, indicating a higher risk of treatment refractory *Giardia* here than elsewhere in the world [156]. The findings from London are in fact supported by another study that was conducted at a travel clinic in Spain, where high rates of giardiasis and refractory infection was observed in travelers from Asia as well [157, 183]. More recently, a Cuban study showed that up to 46 % of patients experienced MTZ treatment failure [184].

1.9.2 Resistance mechanisms

Several studies have investigated the up-and down-regulation of enzymes in the MTZ metabolism, and most studies are based on laboratory *Giardia* strains such as the WB isolate (sub-assemblage AI). *Giardia* sub-assemblage AII, and especially assemblage B are difficult to culture, and consequently, standardized susceptibility assays for resistance such as agar-diffusion and minimum inhibitory concentration (MIC) testing are difficult to accomplish [185].

The suggested resistance mechanisms in *Giardia* consist of down-regulation of MTZ activation enzymes that reduce MTZ into reactive toxic intermediates and is termed passive, or alternatively, up-regulation of enzymes that have high enough redox potential to inactivate MTZ into an inert amine that has been named active resistance mechanism [186]. One of the most familiar passive resistance mechanism include down-regulation of the PFORs, followed by down-regulation of NR1 in addition to down-regulation of oxygen detoxification enzymes [186]. Down-regulation of oxygen detoxification enzymes has a more indirect effect on MTZ metabolism, as it allows more oxygen to be present in the cytosol, which leads to futile redox cycling, thus re-oxidation of MTZ back to its original, inert form [186, 187]. This could mean that some parasites may have increased ability to withstand oxidative stress during MTZ exposure (by e.g., having multiple copies of detoxification enzymes, such as gFlHb) [144, 188, 189]. Lower PFOR activities has been associated with MTZ resistance, although MTZ resistant *Giardia* lines with normal levels of PFOR/fd has been described as well [144]. Still, the oxygen scavenging capability of MTZ resistant *T. vaginalis* has been found to be impaired, meaning that this parasite could handle higher levels of oxygen, but be susceptible during low oxygen levels [144].

Oppositely, active resistance mechanism includes up-regulation of the MTZ-inactivating enzyme NR2 [186]. Further overexpression of the enzyme TrxR has in one study rendered *Giardia* more susceptible to MTZ, although contradictory, as TrxR was associated with a protective role in a former study [163, 190].

Moreover, clinical MTZ refractory *Giardia* strains from patients have not always been found to be resistant in laboratory settings [191], meaning eradication is likely to be affected by host physiology and immune response as well [144].

In the bacteria *H. pylori*, nonsense mutations in the *rdxA* genes (oxygen-insensitive *NR*) are a plausible reason for MTZ resistance [192, 193]. Moreover, MTZ resistance in *T. vaginalis* is possibly linked to mutations in the *NR* genes *ntr4Tv* and *ntr6Tv* [194]. Considering this, genetic variation, e.g mutations, may as well be one of the reasons for treatment failures in *Giardia*. A recent study published by Ansell et al in 2017 found a nonsense mutation in some transcripts of the NR1 in the laboratory-induced MTZ resistant isolate, 106-MtzR [186]. Up until now, no specific genetic marker of resistance has been discovered in MTZ treatment-refractory *Giardia*, and the resistance is probably complex as MTZ does not target a single protein or mechanism, and the proposed resistance mechanisms may be incomplete. The recent increase in MTZ-treatment refractory *Giardia* cases observed during the last few years, suggest that strains with certain genetic inheritable traits are spreading.

2. AIMS OF THE THESIS

2.1 Overall aim of the thesis

The overall aim of the thesis was to investigate genetic variation, and the potential effect on protein function and metronidazole (MTZ) tolerance, in genes of *Giardia* proteins that are responsible for activation or inactivation of MTZ or important in detoxification of reactive oxidative species (ROS) and nitrosative stress responses.

2.2 Specific research objectives

Paper I:

The aim of this paper was to investigate the general genetic variation in a set of 29 genes important in the metabolic pathways related to MTZ resistance or management of oxidative and nitrosative stress substances by whole genome sequencing of cultured, recent clinical *Giardia* isolates.

Paper II:

The *Giardia* flavohemoprotein (*gFlHb*) gene showed large genetic variation, and indications of being a multicopy gene in Paper I. The aim of Paper II was to explore the copy number variability, allelic diversity, and SNVs potential effect on protein function by a cloning based approach in addition to Pacbio and Illumina sequenced genomes in a set of recent, cultured, clinical *Giardia lamblia* sub-assemblage A2 and assemblage B isolates.

Paper III:

The aim of the final study was to analyze the genetic variation and allelic composition of the two *Giardia* nitroreductases 1 and 2 genes in recent cultured and in non-cultured assemblage B clinical isolates, some of them from MTZ treatment-refractory patients. Identified genetic variants were evaluated in relation to clinical information where this was available, and for their potential effect on protein function.

3. MATERIALS AND METHODS

3.1 Genes of interest (Paper I)

A PubMed literature search was performed during 2016-2017 in order to obtain a list of genes that could be of potential interest for MTZ metabolism, for protection against ROS and RNS and antioxidant properties. The word *Giardia* was used in combination with metronidazole, refractory, resistance, and oxidative stress.

Further, the publications and their references were used to make a list of 29 candidate genes (Paper I, Table 1 [28]). The candidate genes were included according to having a direct or indirect effect on MTZ metabolism, being co-factors of MTZ metabolism genes or involved in oxidative and nitrosative stress responses. Additionally, supplementary data from Adam et al [26] and The *Giardia* DataBank (http://giardiadb.org) was used to identify pairwise shared genes and ORFs for *Giardia* sub-assemblage AII and assemblage B.

3.2 Clinical *Giardia* isolates and samples

3.2.1 Cultured clinical Giardia isolates (Paper I, II and III)

20 clinical isolates of both human infecting *Giardia* assemblages, A and B, were successfully cultured from human clinical stool samples between 2011 to 2015 at the Robert Koch-Institut (RKI), Berlin, Germany (see Table 1 for overview of the isolates). 19 of the stool samples, were collected by the Institute of Tropical Medicine and International Health, Charité, Berlin, Germany, whereas isolate P324, was obtained from the Tropenmedizinische Ambulanz, University hospital Düsseldorf, Germany. Until extraction, the samples were kept in the refrigerator and the cysts were extracted from the stool samples as soon as possible using a two-step sucrose gradient filtration with centrifugation. The cysts were then excysted into vegetative trophozoites by using a well-known two-step excystation method [195]. Trophozoites were then cultured in Diamond's TY-S-33 medium supplemented with bile at 37°C in confluent cultures in 11 ml tubes, according to the methods of Keister [196]. After the trophozoites had been growing for a couple of days, they were collected, and DNA

was extracted with Maxwell® 16 FFPE Plus LEV DNA Purification Kit (Promega BioTech AB, Sweden, cat AS1135) and the Maxwell® 16 instrument (AS1000). Nanodrop (Thermo Fisher Scientific) was used to identify the amount of DNA while Qubit 2.0 (Life Technologies, Grand Island, NY) was used to validate the DNA concentration. No clinical data was collected from the clinical samples of *Giardia* obtained by RKI.

Table 1: Clinical isolates obtain from stool samples of *Giardia* assemblages A and B and year of collection.

Isolate	Assemblage	Year of collection
P033	AII	2011
P034	AII	2011
P064	AII	2011
P316	AII	2013
P324	AII	2013
P361	AII	2013
P368	AII	2013
P392	AII	2013
P403	AII	2014
P407	AII	2013
P478	AII	2013
P506	AII	2014
P344	B	2014
P387	B	2014
P413	B	2014
P424	B	2014
P427	B	2014
P428	B	2014
P433	B	2014
P458	B	2014

3.2.2 Collection of *Giardia* DNA from clinical non-cultured samples (Paper III)

Non-cultured clinical samples of *G. lamblia* assemblage B were collected from 2004-2017 at Haukeland University Hospital, Bergen, Norway from giardiasis patients included in ongoing studies or being followed at the out-patient department. Clinical treatment responses, routine and study specific lab results were recorded in patient journals or in specific study forms.

Cysts from a total of 18 non-cultured isolates were purified from patient stool samples by sucrose floatation alone or in combination with IMS (Dynabeads® G-C combo kit, Life technologies) following the protocol from the manufacturer and Hanevik et al 2015 [82], but 200 ul of H_2O was added after the final bead wash step,

and the samples were frozen at -80 C before DNA extraction was carried out.

DNA from the isolated *Giardia* cysts was extracted using either mini stool kit (Qiagen) according to the manufacturer's descriptions, except for using 70 degrees warm nuclease free water for eluation of DNA, or the by using the kit MagAttract® HMW DNA (Qiagen). The HMW kit is based on magnetic separation of high molecular DNA, and to avoid the magnetic beads from the previous IMS extraction to interfere with the DNA binding beads, the IMS beads were removed before DNA extraction. This was carried out by freezing and thawing three times before adding 20 µl protein kinase K followed by 150 µl AL buffer and 4 µl RNase A from the MagAttract® kit. The samples were then incubated at 56 degrees for 1 hour with gentle agitation according to Ogren et al 2016 [197]. The samples were then frozen and thawed two more times followed by a final heat incubation at 98°C for 15 min and the remaining beads and cyst wall debris were pelleted and removed by centrifugation at 11 000 g for 2 min. The supernatants were transferred to a clean Eppendorf tube before following the kit's instructions. The samples were frozen and thawed for a total of five times following DNA extraction according to our previous experience [82].

Cysts from a sample originating from the 2004 Bergen *Giardia* outbreak (see Table 1:VA) was obtained by salt flotation and IMS, while the DNA was extracted by QIAamp DNA mini kit (QIAGEN GmbH, Germany) according to Robertson et al 2006 [198].

3.3 Polymerase chain reactions (Papers II and III)

3.3.1 Obtaining genes from *Giardia* using polymerase chain reactions

Polymerase chain reaction (PCR) was developed by the American biochemist Kary Mullis in 1985 [199]. The PCR method has the advantage of creating up to millions of copies of a gene of interest from a relatively small starting amount of DNA [199]. The reaction requires template DNA from the organism of interest, a forward and reverse primer, a polymerase, nucleotides for generating DNA copies, nuclease free water and a buffer [101]. The PCR reaction consists of three steps; namely the denaturation, the annealing and the extension, the two latter steps may be combined in a single step [199].

The DNA-concentration of a PCR is crucial, as high amounts of DNA can inhibit a reaction, by increasing the risk of amplification of unspecific products or by presence pf nucleases/inhibitors in the DNA [200]), and a too small amount may hamper the primers from binding to the target. Inhibitors are normal in an original sample (polysaccharides in feces), but inhibitors may be added later point during e.g., sample processing and DNA extraction [101, 200].

3.3.2 Primers

National Centre for Biotechnology Information (NCBI) also known as GenBank™ primer tool software was used together with Geneious Prime to design forward and reverse primers in conserved regions of the genes, flavohemoprotein (*gFlHb*) (DHA2_154000 and GSB_151570) for *Giardia* assemblages A and B and nitroreductase 1 (*NR1*) (GSB_22677) and *NR2* (GSB_153178) for assemblage B (most primers were designed by Gamlem [101]). The cloning primers were made based on whole genome sequences from isolates from Paper I. All primers were designed approximately 50-200 bp up- or downstream of the CDS. In order to choose the best pair of primers obtained in the programs, the Eurofins Genomics oligo analysis tools was used to identify self-annealing characteristics and risk of primer-dimer formations. The primer pairs with the best characteristics were chosen. All primer pairs were visually inspected in Geneious to check for SNVs in the primer binding regions. For the *NR1* gene, it was not possible to design a reverse primer in a conserved region, due to mutations, that otherwise would present adequately good primer qualities [101]. Two different reverse primers were created, one included the SNV and the other one was complementary to the reference sequence. Both of the reverse primers were used together with the forward primer, see Table 2. For the *NR2* gene, some of the non-cultured *Giardia* samples were not successfully amplified with the primers designed to bind to the cultured *Giardia* samples' sequences, and a new set of primers was designed to achieve a successful PCR (see Table 2).

3.3.3 Primer design

A PCR requires two primers, one forward and one reverse to produce an amplified DNA sequence. The forward primer is in the same direction as the DNA (3'–5') while the reverse follows the opposite direction (5'-3'). Effective primers are a necessity for a successful PCR, as the polymerase used in a PCR experiment requires a primer that is bound to the target DNA to begin synthesis of a new strand [201]. Further, SNVs in the target DNA sequence, especially the primer binding region (nucleotide mismatch) may hamper the primer binding efficacy [202]. Numerous primer characteristics increase the probability of a successful PCR [199] (listed by Lorenz et al [199] and according to Gamlem's thesis [101]):

- Low probability of self-annealing (meaning the primer binds to itself instead of the target DNA).
- Similar T_m of forward and reverse primers, ideally, not more than 5°C apart.
- G-C content around 40 – 60%.
- The 3'end (binding site of the DNA polymerase) contains G or C, known as a GC-clamp (as the C-G bond is physically stronger than the A-T bond).
- Di-nucleotide repeats (e.g., ATATATATA) or single base repeats (e.g., CCCCC) increases the risk of hairpin loop structures.
- 18-22 nucleotides long primers prevent binding to other similar sequences in the target DNA [203].

The target specificity of a primer is yet another critical characteristic in order to obtain amplification of the target gene. Primers may bind and unspecifically amplify other similar DNA sequences in the sample, including from bacterial or human contaminant DNA. The risk for this was reduced by blasting promising target primers on NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Self-annealing characteristics were checked with the online oligo calculator (https://eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/).

3.3.4 Gene PCR primers

Table 2: Primer sets for obtaining the *Giardia* flavohemoprotein (*gFlHb*) gene DHA2_154000 (sub-assemblage AII) and GSB_151570 (assemblage B), and primer sets for obtaining the *NR1* GSB_22677 and *NR2* (GSB_153178) from *Giardia* assemblage B isolates. Primers were ordered from either Eurofins Genomics, Luxembourg or TIB Molbiol, Berlin, Germany.

Gene	Direction	Primer Sequence, 5'–3'	Bp	[Tm][1]	Position related to CDS	Amplicon length	Reference
DHA2_154000 *gFlHb*	Forward	CGCCACCACAAGCGATCATT	20	59.4	-50	1473 bp	Paper II
DHA2_154000 *gFlHb*	Reverse	GTGTGTAGAGCGATTACAT	19	52.4	+46		Paper II
GSB_151570 *gFlHb*	Forward	CGGCCTTCAGGTACTTCCCC	20	63.5	-23	1850 bp	Paper II
GSB_151570 *gFlHb*	Reverse	GAGACCAAAAGTCCATATGAACT	23	57.1	+39		Paper II
GSB_22677 *NR1*	Forward	GTGATGGAGCAAAGTCGC	18	64°C	-162	1135 bp	Paper III
GSB_22677 *NR1*	Reverse1	GTGGATGGGGCTCTTGAATA	20	64°C	+178		Paper III
GSB_22677 *NR1*	Reverse2	GTGGATGAGGCTCTTGAATA	20	61°C	+178		Paper III
GSB_153178 *NR2.1*	Forward	ATCTACATAAGATCCGCGCACT	22	65°C	-174	1212 bp	Paper III
GSB_153178 *NR2.1*	Reverse	TACTCTGCACCTCATCGCCG	20	69°C	+171		Paper III
GSB_153178 *NR2.2*	Forward[2]	GACTCACAGAGTGGCAACGA	20	67°C	-251	1271 bp	Paper III
GSB_153178 *NR2.2*	Reverse[2]	CGCCGAGCAATGTAGTGGTT	20	68°C	+156		Paper III

[1]Tm is estimated from NEB T_m calculator with 400 nM primer concentration and Q5® polymerase.
[2]New primers for non-cultures isolates of *Giardia.*

3.3.5 PCR protocol using Q5® polymerase

The protocol for the gene PCRs was based on New England Biolabs' (NEB) recommendations. The proofreading polymerase, Q5® (NEB, M0491S), was used for all of the gene PCRs. Reaction volumes of 25 µL was used.

3.3.5.1 master mix for gene PCR

The reagents for one PCR reaction are shown in Table 3. The master mix was added to PCR-tubes (Thermo Scientific, 3418), followed by addition of 1 µl sample DNA, and nuclease-free H_2O (Qiagen, 129112) for negative controls. Then the PCR tubes were placed in a thermocycler (Thermo Fisher Scientific, A24811).

Table 3: Gene PCR Master mix for one reaction

Component	Amount
5X Q5® Reaction Buffer	5.0 µL
10 mM dNTPs	0.5 µL
10 µM Forward Primer	1.0 µL
10 µM reverse primer	1.0 µL
Q5® High-Fidelity DNA Polymerase	0.25 µL
Nuclease-free water	11.25 µL
5X Q5® High GC enhancer	5.0 µL
Total	24.0 µL

3.3.5.2 Gradient PCR

Gradient PCR was done to find the optimal annealing temperature for each of the three genes. Different annealing temperatures were tested (e.g., 55°C, 58°C, 60 C and 64°C), while the other thermocycling conditions were kept constant. Gradient PCR was conducted with *Giardia* positive reference DNA samples: GS (assemblage B) and WB (sub-assemblage AI).

3.3.5.3 Thermocycling conditions for PCR

The thermocycling programs used were depended on genes (e.g., length and T_m). Thermocycling conditions used for the genes in this thesis are presented in Tables 4 and 5.

Table 4: Thermocycling conditions for gene PCR with Q5® polymerase for *gFlHb.*

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
35 cycles	98°C	10 seconds
	55°C[1]/60°C[2]	30 seconds
	72°C	45-60 seconds
Final extension	72°C	3 minutes
Hold	12 °C	∞

[1]Sub-assemblage AII
[2]Assemblage B

Table 5: Thermocycling Conditions for gene PCR with Q5® polymerase

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
39 cycles	98°C	10 seconds
	64°C	20 seconds
	72°C	31 seconds
Final extension	72°C	2 minutes
Hold	12 °C	∞

3.3.5.4 qPCR detection of Giardia DNA (paper III)

qPCR detection of the *Giardia* glutamate dehydrogenase (*gdh*) was carried out for the 18 non-cultured cyst samples with the LightCycler 480 II instrument (Roche Diagnostics GmbH, Mannheim, Germany), according to [204], except for using a *Giardia* assemblage B-specific forward primer (Table 6), FAM fluorochrome was used for probe, and a standard curve using 10 fold dilutions of a quantified pUC57 based plasmid (Genscript, NJ, USA) containing the target sequence, *gdh*, (accession number MT108431.1). The primers and probes for *Giardia* qPCR are listed in Table 6.

Isolates with ~900 gene copies per µl sample, or higher, were selected for downstream cloning experiments, resulting in a total of eight samples that produced visible bands on the agarose gel after gene PCR amplification, and included one sample with unknown treatment status, three clinically MTZ susceptible samples and four MTZ refractory samples (see Table 7). The MTZ refractory samples originated from patients that were unsuccessfully treated with MTZ, confirmed by *Giardia* positive follow-up stool specimens. These three patients were later cured using secondary or tertiary drug regimens according to the 'treatment ladder' proposed by Mørch et al 2008 [145], and a small subunit ribosomal RNA gene (16S-like) qPCR of

Giardia according to [205] was carried out to ensure that the patients were successfully treated.

Table 6: qPCR primer overview

Gene	Direction	Primer Sequence (5' – 3')	Bp	[Tm] [1]	Position related to CDS	Amplicon length	Reference
GDH	Forward	GGGCAAGTCGGACAACGA	18	61	434	262 bp	[204] and this study
GDH	Reverse	GTCTACTTCCTGGAGGAGATGTGC	24	62.2	696		[204]
GDH	Probe	6FAM-TCATGCGCTTCTGCCAG-BBQ	17	62.3	454		[204]

Table 7: Available clinical isolates of *Giardia* assemblage B, year of collection and type of sample

Isolate	Year of collection	Type of sample	Treatment status
0099	2014	Clinical non-cultured sample	No data
Ag15	2013	Clinical non-cultured sample	Susceptible
Ag30	2017	Clinical non-cultured sample	Susceptible
VA	2004	Clinical non-cultured sample	Susceptible
Ag10	2013	Clinical non-cultured sample	Refractory
Ag13	2013	Clinical non-cultured sample	Refractory
Ag20	2014	Clinical non-cultured sample	Refractory
Ag22	2014	Clinical non-cultured sample	Refractory

3.4 Genome (Illumina and Pacbio) Sequencing

3.4.1 Illumina Miseq sequencing (Papers I-III)

Fragmented *Giardia* DNA (2 µg, 550bp insert size) from the twenty cultured trophozoites was first prepared and indexed using TruSeq DNA PCR-Free Low Throughput Library Prep Kit and TruSeq DNA Single Indexes Set A (12 indexes, 24 samples) (Cat. Nr 20015962 and 20015960S, Illumina, San Diego, Ca, USA). Library size distribution was validated using the 2100 BioAnalyzer and the High Sensitivity DNA chip (Agilent Technologies, Santa Clara, CA, USA). Quantification was done using the KAPA Library Quantification Kit (Roche, Basel, Switzerland) for Illumina sequencing platforms. Normalized (libraries diluted to the same concentration) whole genome *Giardia* DNA-libraries were then sequenced using the Illumina MiSeq paired-end (PE) technology (2 x 300 bp). Further, FASTQ-files from the sequencing were

assembled and aligned to the respective reference genomes of *Giardia* sub-assemblage AII and assemblage B using the aligner Bowtie 2-2.2.3 [206]. The reference genomes of *Giardia* (versions 26) were obtained from *Giardia* DB in January 2016 [207].

3.4.2 PacBio de novo sequencing (Paper II)

High-molecular-weight DNA was extracted from eight of the twenty RKI cultured *Giardia* isolates using the Genomic-tip kit (Qiagen, Hilden, Germany). PacBio sequencing service was carried out at GATC Biotech (Konstanz, Germany) with at least 50-fold coverage of the expected 12 MB genome size of *Giardia*. De novo assemblies were generated for the eight *Giardia* isolates by the bioinformatic support team at the RKI using the software tools HGAP 2.0 (PacificBiosciences). More details are explained in Klotz et al. 2021 (manuscript in preparation).

3.5 Pre-cloning methods (Papers II and III)

3.5.1 Gel electrophoresis

Gel electrophoresis may be used to separate DNA fragments (or PCR products) according to their size [208]. DNA is negatively charged and the fragments will move towards the positive electrode of the electrophoresis chamber [208]. Smaller fragments will move through the agarose gel at a faster pace than larger molecules. DNA-binding stains (e.g GelGreen® (Paper II) and GelRed® (Paper III)) are either added directly to the gel before casting, to the PCR products directly or as an after-stain, and is used to develop images of the PCR products in an imaging machine such as iBright™ system.

3.5.1.1 Agarose gel and visualization of PCR products

Agarose gels were used together with 1X TAE buffer. The PCR products were mixed with 6X Loading dye (Biolabs, B7024S/B7021S) prior to loading of the wells, while a 1 kB (kilobase) DNA ladder (Biolabs, N3232S/N0552L) was used to determine size of the PCR products (Figure 11). The gels were run for 45 minutes to one hour (using a power supply from Bio-Rad, 1645050). The gel was finally visualized with an iBright™ CL1500 Imaging System (Thermo Fisher Scientific, A44240).

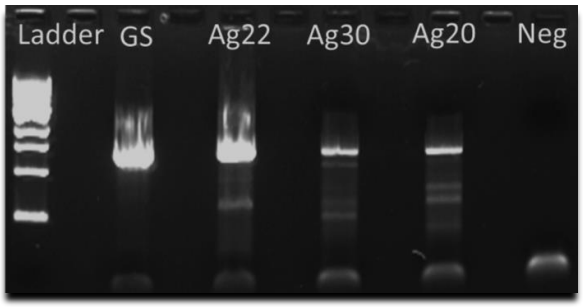


Figure 11: Agarose gel with NR2 PCR-products, where a 10 000 bp is the largest band and 500 bp is the smallest band of the ladder.

3.5.2 Purification of PCR products

The Wizard® SV Gel and PCR Clean-Up System (Promega, A9282) was used for purification of PCR-products. The manufacturer's protocol was followed, except for using 70 °C nuclease free water in the final step of the purification.

3.5.3 Sequencing of PCR products

To confirm amplification of the correct in gene in the PCRs, some of the purified PCR would be sequenced using Big Dye 3.1 (Thermo Fisher Scientific, 4336917). Parameters for the thermocycler was set to 28 x (96º C 1 min, 96 ºC 10 seconds, annealing temperature 5 sec. and 60 ºC 4 min.). The PCR products were either Sanger sequenced at the Sequencing Laboratory of the RKI, Berlin, Germany (Paper II) or at the 6th floor of the laboratory building, University of Bergen, Bergen, Norway (Paper III).

3.6 Cloning methods (Papers II and III)

In cloning, the gene of interest is inserted into a circular piece of DNA known as plasmid or vector by a method known as ligation [209]. The vector is further introduced into bacteria by transformation [209, 210]. The most widely applied cloning vectors are *E. coli* plasmids containing 3 functional regions: 1) replication origin, 2) antibiotic resistance gene and 3) a region where foreign DNA can be inserted [101, 209].

Competent cells are used in cloning experiments, as these cells are more permeable to foreign DNA uptake. Competent cells have different transformation efficiency, meaning how effective a cell is at incorporating extracellular DNA into its cell [101, 211].

3.6.1 Ligation

The CloneJET PCR Cloning Kit (Thermo Fisher, K1231) was used to ligate the gene of interest into the cloning vector. The protocol was carried out as described by the manufacturer. The mixture was briefly vortexed and centrifuged before transformation was carried out.

The vector contains both a lethal restriction enzyme and a gene coding for ampicillin (AMP) resistance. If a DNA-ligation is unsuccessful, the lethal gene will kill the bacteria. Ergo, only bacteria that receive DNA/gene inserted into the vector can form colonies on an agar plate with AMP.

3.6.2 Transformation

The ligation mixtures were introduced to *E. coli* DH5alpha competent cells (BioLabs, C2987U, 50 µL) following a standard transformation method [210]. The bacteria were carefully thawed on ice and transferred to a 11 mL tube (Pyrex ®, 99445-15). 5 µl ligation mixture was added to the bacteria, followed by an incubation on ice for 30 min. and a heat-shock reaction for 30 seconds at 42°C using a water bath (Grant GD120). The cells were left to rest for 2 min on ice before 950 µL room temperature LB medium was added to the bacteria and placed in a CO_2-incubator (Labnet, I5311-DS) for at least one hour with 250 rpm shaking at 37 °C. After the shaking incubation the bacteria was plated on pre-heated AMP LB agar plates in triplicates and placed in a CO_2-incubator (Sanyo) over-night at 37°C.

3.6.3 Verification of insert by colony PCR

To confirm successful transformation of the gene into *E. coli,* colony-PCR was performed. *For each experiment, 20-30 of the visible colonies for each gene per Giardia*

isolate were tested. A ready-made master mix solution was used for the colony PCRs (HotStarTaq Master Mix Kit, Qiagen, 203445). The pJET1.2 forward and reverse sequencing primers (ThermoFisher) were used to determine size of insert. The reaction mixture for one reaction is presented in Table 8.

Table 8: Colony PCR reaction mixture

Reagent	Amount (µl)
Master mix	12.5
10 µM Forward primer	0.5
10 µM Reverse primer	0.5
Nuclease-free water	8
Coral load	2.5
Total	24

The thermocycling conditions for the colony PCR are presented in Table 9. Annealing temperatures for the primers were found using the NEB online calculator. An image of a typical colony PCR can be seen in Figure 12.

Table 9: Colony PCR thermocycling conditions

Initial activation step	15 min.	95 °C
3 step cycling: 30 cycles		
Denaturation	45 sec.	94 °C
Annealing	45 sec.	57 °C
Extension	1.2 min.	72 °C
Final extension	10 min.	72 °C

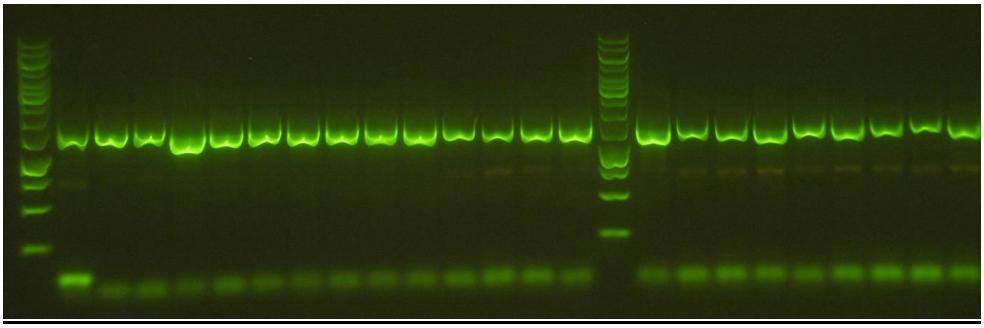


Figure 12: A typical colony PCR of the gFlHb gene. The first well represents the DNA ladder, followed by 14 positive PCR reactions, a repeated DNA ladder and 9 positive PCR products.

3.6.4 Incubation

Positive colonies were cultured over-night (37 °C, 250 rpm shaking) in 1-2 mL LB medium supplemented with AMP. Approximately 20 clones of each gene per isolate were cultured before plasmid extraction using Zyppy Plasmid Miniprep Kit (Zymo research, D4037). The manufacturer's protocol was followed. TE-buffer (Haukeland University Hospital) was used to resuspend bacterial pellets before Lysis Buffer from the kit was added.

3.7 Sanger sequencing (Papers II and III)

3.7.1 Plasmid sequencing

The *gFlHb* plasmids were sequenced at the Sequencing Laboratory of the RKI, Berlin, Germany and by the company Genewiz (Leipzig, Germany) using a technique known as Sanger sequencing, while the *NR* plasmids were all sequenced by Genewiz.

3.8 Analysis of genetic variation

3.8.1 Sequence analysis (Papers I, II and III)

The reference genomes of *Giardia lamblia* sub-assemblage AII (DH) and assemblage B (GS_B) were obtained from the Giardiadb.org [207] and imported into the genome analysis software Geneious v.10.2.4 (Paper I) and Prime ® (Papers II and III). One FASTA-file and one general feature format (GFF) file were combined in Geneious to obtain annotated genes. Illumina sequenced isolates (20 previously described cultured assemblage A and B isolates [28]) were imported into Geneious and analyzed (Paper I) or used to validate corresponding cloned sequences (Papers II and III).

Aligned data were analyzed using Geneious and variant calling was done to identify single nucleotide variants (SNVs) where a minimum variant frequency of 0.1 and a minimum coverage of 10 nucleotides was used (Paper I). Coding regions (CDS) of 29 candidate genes (including pre, 150 bp, - and post-coding regions, 50 bp) were analyzed.

Sequenced *gFlHb, NR1* and *NR2* forward and reverse sequence chromatograms from clones were aligned and compared to the reference genes of interest, *gFlHb;*

DHA2_154000 and GSB_151570, *NR1*; GSB_22677 (*Fd-NR2* in *Giardia*DB) and *NR2*; GSB_153178 (*Fd-NR1* in *Giardia*DB) for the cloned sequences in Papers II and III.

For the cultured isolates' clones the SNVs were validated using criteria based on whole genome sequences and clone sequences [188], whereas in non-cultured isolates any SNV present in more than one of all clone sequences from any isolate were considered as valid. SNVs were thus divided into high confidence (HC) and low confidence (LC), and only HC SNVs were included in the analysis. The consensus sequences from all the clones from one isolate were aligned together and finally sorted into haplotypes based on the SNVs in the sequences. Excel files were used to examine all SNV positions and make tables inspired by Lecova et al [115]. SNVs were divided into synonymous(s) or non-synonymous (ns).

For Paper II, DNA Sequence Polymorpism (DnaSP) [212, 213] v6 was used to calculate nucleotide diversity (defined as the average number of nucleotide substitutions, per site between two sequences within a population for pairwise comparisons [214]), in addition to haplotype diversity (here defined as allelic diversity, which is an estimate of the uniqueness of an allele in a population, or the likelihood that two alleles are different from one another) [188].

For Paper III, haplotype hybrids (chimeric products created during PCR amplification [215]) were identified using the programs Bellerophon, Geneious and Excel [216].

3.8.2 Data analysis of amino acid changing positions (Papers I, II and III)

The universal resource databank, Uniprot.org, was used to identify important domains of the proteins (for gFlHb: heme binding globin domain, FAD-binding domain and C-terminal NAD binding domain. For NR proteins we analyzed the ferredoxin (fd) domains separately.

3.9 Homology modelling (Papers I, II and III)

Conservative replacements were identified as amino acid substitutions with rather small physicochemical distance, whilst a radical amino acid included large physicochemical characteristics that can lead to potential rearrangement of the secondary structure (e.g., truncation, rearrangements of secondary structure or changes in important domains).

3.9.1 Ferredoxin homology modelling (Paper I)

Homology modeling of the NRs and fds 1 to 6 was based on the reference sequences of the genes. Homology models were created using the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) server with intensive mode [217]. Further, the models were visualized and superimposed in PyMol (The PyMOL Molecular Graphics System, Version 2.1.0 Schrödinger, LLC).

3.9.2 Flavohemoprotein modelling (Paper II)

The crystal structure of the gFlHb is not known, and homology models were used to estimate whether amino acid changes would affect protein structure and function. Homology modeling was carried out using the program Protein Homology/analogy Recognition Engine V2.0 (Phyre2) [217] with the help of a co-author experienced in this method. Two protein sequences, E2RTZ4 and A0A482ESB4, were obtained from Uniprot and served as models for gFlHb assemblage A and B, respectively. The model was created using the single highest scoring template, and a crystal structure of the *E. coli* flavohemoglobin (PDB-ID 1GVH) (40% sequence identity with the gFlHb [218]) was used.

3.9.3 Nitroreductases modelling (Paper III)

Open reading frames of the *NR1* (GSB_22677) and *NR2* (GSB_153178) genes were modeled with the modern and simplified AlphaFold 2 [219] using Google Colab implementation (colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFo

ld.ipynb), and visualized in Chimera [220].

3.10 Analysis programs and software (Papers I, II and III)

-Word 2019 (Microsoft Corporation®, Redmond, USA) for Windows, was used to write the thesis and make tables.

-Excel 2019 (Microsoft Corporation®, Redmond, USA) for Windows, was used to analyze sequences, calculations and sorting clones into haplotypes based on SNV position.

-Endnote X8 (Clarivate Analytics Philadelphia, USA) for Windows, was used as the citation manager in this thesis.

-PowerPoint 2019 (Microsoft Corporation®, Redmond, USA) for Windows, was used to draw figures.

-Geneious v.10.2.4 (2017) and Prime® 2020 (Biomatters Ltd., Auckland, New Zealand), version 2.1 for importing reference genomes, sequence analysis, SNV calling and making phylogenetic trees.

-The primer designing tool (NCBI, Bethesda MD, USA) was used for designing primers.

-DnaSP v6 [212, 213] was used to calculate nucleotide diversity (pi) and haplotype diversity (Paper II).
- Bellerephon 2004 [216] was used together with visual inspection to identify chimeric sequences.

3.11 Accession numbers in genbank (Papers I, II and III)

Paper I: MK043361 - MK043940
Paper II:MT713149-MT713229, MT713231-MT713243, MT713245-MT713263, MT713267, MT713269.
Paper II: Will be made available at publication.

3.12 Ethical aspects (Papers I, II and III)

For Paper I and II no patient data was used in the analysis, and any link between individual parasite data and patient information had been removed before the whole genome sequencing of the trophozoites was carried out.

For seven of the eight clinical isolates obtained from Haukeland University Hospital and presented in Paper III, the study was approved by the Regional Committee for Ethics in Medical Research (REK) (2013/1285/REK vest). The final sample, VA, was obtained from a Combined research and clinical biobank of the infectious diseases department (REK vest 165.04) (Diagnostisk og Forskningsbiobank for primære/ervervede immunsvikttilstander og alvorlige infeksjonssykdommer). Informed consent was obtained from all patients who donated stool samples and clinical data.

4. SUMMARY OF RESULTS

The first study, Paper I, aimed to investigate the general genetic variation in 29 genes important in the metabolism of metronidazole (MTZ) and detoxification of nitrosative and oxidative stress substances. The second study, Paper II, was a follow-up study of Paper I, where the nitrosative stress detoxification gene, flavohemoprotein (*gFlHb*) was analyzed for genetic variation and copy number variation. Finally, Paper III utilized clinical cultured isolates in addition to non-cultured *Giardia* samples to investigate the genetic variation of the two genes nitroreductase (*NR*) 1 and 2 to assess the genes' potential as markers of MTZ resistance.

Recent clinical cultured *Giardia* isolates were used in all three studies, while non-cultured MTZ-refractory and susceptible clinical samples were included in the final study.

Paper I: "Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical *Giardia lamblia* assemblage A and B isolates".

The study describes findings of the general genetic variation in a set of 29 MTZ-metabolizing and detoxification genes. The genetic variation was obtained from Illumina whole genome sequenced 20 recent clinical cultured sub-assemblage AII and assemblage B *Giardia* isolates.

The average number of nonsynonymous (ns) single nucleotide variants (SNVs) for the 29 candidate genes was 0.3% for AII and 1.2% for assemblage B. The genes with the highest genetic variation in sub-assemblage AII were ferredoxin (*Fd*)*2*, *Fd6* and nicotinamide adenine dinucleotide phosphate (NADPH) oxidoreductase putative genes. In assemblage B isolates, several genes in the MTZ activation or oxidative stress management (thioredoxin peroxidase, *NR2*, *Fd2*, NADH oxidase, *NR1*, alcohol dehydrogenase, *Fd4* and *Fd1*) had the higher numbers of nsSNVs compared to the average variation. On the contrary, *NR2* and *NTR-1* was found to be fully conserved in sub-assemblage AII, while *Fd6* and Histone H2A were fully conserved in assemblage B.

Homology modelling showed that amino acid changes close to the Fe-S cluster binding sites (ferredoxin domains) in the NR1 and NR2 and Fd2 and Fd4 may affect these proteins' abilities to interact with other proteins, or perhaps even affect intermolecular electron transfers.

The flavohemoprotein gene (*gFlHb*) was found to have a substantially higher coverage (87.6 (54.2)) in 17 of the 20 *Giardia* isolates, indicating that this gene could potentially be a variable multicopy gene.

Paper II: "Genetic diversity of the flavohemoprotein gene of *Giardia lamblia* – evidence for high allelic heterozygosity and copy number variation".

Re-analysis of the Illumina sequence data from the 20 clinical isolates in Paper I, showed variable, but generally higher coverage along the whole coding regions (CDS) of the *gFlHb* gene. PacBio sequence data was available for eight of the isolates and showed that *gFlHb* was present in two to six copies in these isolates. Generally, a higher Ilumina coverage fold difference would correspond to a higher number of gene copies in PacBio sequence assemblies.

A total of 37 distinct alleles (sub-assemblage AII), and 41 distinct alleles (assemblage B) were identified in the isolates by cloning. In some single isolates, up to 12 unique alleles were identified. The allelic diversity, or the probability that two alleles are different from one another, was estimated to be high for both assemblages (>0.9) while the nucleotide diversity was <0.01. The genetic variation (SNVs per CDS length) was calculated to be 4.8 % and 5.4 % (AII and B respectively). The average number of nsSNVs was 1.6 % in AII and 3.0 % in B.

We further wanted to evaluate how and if the distinct *gFlHb* alleles obtained by cloning were represented in the *gFlHb* genes discovered by de-novo assembled consensus genomes from PacBio sequenced isolates. Several allelic forms of *gFlHb* were found to be present in different isolates, for example one specific allele was found to be present in four sub-assemblage AII isolates. Three PacBio copies were identical to alleles obtained by cloning in their corresponding isolates. Fewer alleles were shared

between the assemblage B isolates compared to sub-assemblage AII isolates.

Some of the identified amino acid changes could potentially alter the protein structure and possibly function of the enzyme, where the most drastic change was a premature stop codon at position 49 in sub-assemblage AII and a change from Y at position 211 to H. For assemblage B, a deletion of three amino acids (position 75-77) was identified, in addition to a change from E to G at position 80. Both the amino acid changes in assemblage B could potentially be helix breakers and affect the heme binding.

Paper III: "Genetic diversity in the metronidazole metabolism genes nitroreductase 1 and 2 in susceptible and refractory clinical samples of *Giardia lamblia*".

In this study, 7 of the clinically cultured isolated were used in addition to 8 non-cultured clinical samples, all from assemblage B. Four of the non-cultured *Giardia* samples were MTZ-refractory, while three were susceptible in addition to one sample with unknown treatment status.

Full length *NR1* and *2* genes were successfully amplified, cloned and sequenced for characterization of SNVs and allele identification in all 15 isolates. Both cultured and non-cultured sample groups contained a spectrum of variability, from highly variable, potentially mixed infections to homozygous samples. The total number of combined SNV positions for all isolates per gene length, was similar for the *NR1* (11.2 %), and *NR2* (10.7 %) genes. The percentage of total SNV positions for individual isolates was an average of 3.5 % (0.3-5.9 %) for *NR1* and 2.6 % (0.7-5.1%) for *NR2*. However, the higher percentage found in *NR1* was most likely due to exceptionally high SNV numbers in a few isolates. The percentages of nsSNV per CDS was found to be slightly less for the *NR1* gene (4.3 %) with 35 variable positions compared to 6.6% and 58 positions for the *NR2* gene.

A total of 69 unique alleles were identified in the *NR1* gene, while 90 different alleles were found in the *NR2* gene. Four of the isolates in the study were homozygotic at both genes, however, none of the refractory isolates were homozygotic.

Chimeric sequences were identified in both NR genes in most of the eleven heterozygous samples. These sequences were omitted from the allele identification analysis. However, the number of distinct alleles and genetic variation remained high in three cultured isolates and in three uncultured samples.

Further, in a refractory sample, a nsSNV at codon 195 in NR1 lead to a premature stop codon, causing a truncated, most likely dysfunctional protein. We also identified three different single nucleotide deletions in a total of seven distinct *NR2* alleles causing a frameshift mutation. These premature stop codons were located at codon position 99 in two isolates, one MTZ susceptible non-cultured sample and one cultured isolate. This particular stop codon was found in a homozygous isolate, meaning that all NR2 proteins most likely would be dysfunctional. The other stop codons, at codon positions 14 and 127, was found in two of the three alleles of one susceptible isolate. Further, in two of the alleles of *NR2* in a MTZ-refractory isolate, a nsSNV was found at codon position 57, leading to a switch from cysteine to arginine in the fd domain.

5. GENERAL DISCUSSION

5.1 Methodological issues

5.1.1 DNA sequencing

Illumina MiSeq sequencing was used to obtain whole genome sequences of 20 cultured *Giardia* assemblage A and B isolates in the current thesis. High concentrations of DNA were obtained from the *Giardia* trophozoites, and whole genome amplification (WGA) of the DNA was not needed before sequencing. Illumina sequencing technology has an advantage of creating paired-end (PE) sequences, up to 300 bp, improving ability for detection of indels and creating more accurate alignments to a reference genome [221, 222]. Sequencing errors may happen during the runs of Illumina and may lead to potential false positive single nucleotide variants (SNVs). Illumina sequencing has previously been estimated to have an error rate of 0.1-0.24 % per base [91, 223]. The coverage of the genes has a high impact on the potential introduced errors, as higher sequence coverage will counterbalance erroneous nucleotides or noise.

The coverage of the 29 candidate genes in Paper I was variable, and a lower sequence coverage was noted for two of the assemblage B isolates and for one sub-assemblage AII isolate (mean coverage of 11.3-14 compared to the average coverage of 31.1). The correct variant calling of a SNV relies both on high-quality sequencing data and high coverage, in order to be statistically relevant [224]. The recommended SNV coverage cutoff has been recommended to be 30X for a diploid genome by Illumina, Inc. [224]. We used a SNV cutoff value of >10 in coverage and a minimum variant frequency >0.1 for the Illumina SNV analysis. For regions with too few sequence-reads to give a satisfactory coverage or many poor-quality reads, the SNVs would potentially not be identified. As *Giardia* is a tetraploid organism, we used a rather low threshold coverage for what is accepted as a SNV, however, rare variants only present in one of four alleles could still have been missed. The relatively short Illumina reads (35-301 bp), although PE, did not allow for identifying full-length genes/alleles in Paper I [28].

For Paper II, sequences obtained by PacBio technology were available, and de novo assembly enabled us to identify longer stretches of the genome to identify the *gFlHb* gene at a variable number of locations in the *Giardia* assemblage A and B genomes [188]. However, PacBio has a much higher error rate than Illumina, where the rate may be as high as 15 % for continuous long reads [89]. The sequencing error rate of PacBio may as well be the reason to why some of the sequenced clones were not identical to the PacBio identified copy sequences.

Sanger sequencing technology was used for the identification of alleles in Papers II and III has a much higher accuracy than the Illumina and PacBio sequencing technologies, where the accuracy has been suggested to be as high as 99.999 % [225].

5.1.2 Primers

The forward and reverse primers used in Papers II and III were designed to bind upstream or downstream of the coding regions (CDS) of the genes. Sanger sequencing files usually had poor sequence quality at the start/end of a sequence but having primers binding outside the CDS led to high quality protein coding sequences. Further, during primer design using NCBI software, the suggested primer binding regions of the target genes were checked for presence of SNVs using sequences from the Illumina sequenced *Giardia* isolates, to avoid nucleotide mismatches that could affect primer binding.

The primers in papers II and III were designed according to guidelines for good primer properties as previously described by Lorenz [199]. However, some properties were difficult to achieve, due to many SNVs in otherwise good primer binding regions. The region where the reverse primer of the *NR1* gene would bind, had two potential SNVs, one at position 8 of the primer (C->T) and one at position 11 (G->A) [101]. Consequently, two variant reverse primers were designed where one nucleotide differed between them (only the SNV with the highest variability was accounted for). The *NR2* gene was successfully obtained from the cultured clinical isolates of *Giardia*, however, were not easily obtained from non-cultured isolates. Several rounds of PCR were tried, by e.g., changing the number of amplification cycles and adding more template DNA to each reaction. After several failures, new primers at a new location were designed for the *NR2* gene and PCR-products could be obtained. The reason for failure was probably

due to SNVs in the original primer binding regions of non-cultured isolates. SNVs in the primer binding region can potentially favour alleles matching the primers, and lead to an unrealistically low number of alleles included in the analysis. However, we did find many isolates with more than one allele.

Most of the primers contained the recommended GC content of 40-60 %, and was within the recommended maximum of 5°C difference in melting temperature ($T_{m)}$. However, the primers that were used to obtain the *gFlHb* gene (GSB_151570) varied from the recommended GC-content of 40-60 %, where the forward primer had a GC-content of 65 %, while the reverse had a lower percentage of 39 %. The T_m were usually less than 5°C apart, except for the primers used to obtain the *gFlHb* genes, where the difference was as high as 7°C for DHA2_154000, while the difference in T_m for GSB_151570 was a little above 6°C. Although the T_m differences were high, we experienced successful amplification for the *gFlHb* gene from 17 of the 20 clinical isolates (DNA samples were not available for one sub-assemblage AII and one assemblage B isolate, while DNA from sample P433 was available at a later timepoint, thus not included due to estimated low copy numbers of the gene).

Nucleotide repeats were tried to be avoided, however, four C's in a row was found for the forward primer of GSB_151570, while four consecutive A's were found in the reverse primer of GSB_151570 and four G's in a row was present in the NR1 (GSB_22677) reverse primer. Further, all except one primer (GSB_151570 reverse) followed the recommended primer length of 18-22 nucleotides.

Finally, the self-annealing characteristics of the primers were checked. and all of the primers except for the reverse primer of *gFlHb* with a self-complementary value of 10 were within the recommendations. However, self-annealing of the primers did probably still happen in most of the PCR reactions, as primer dimers bands could be observed on the agarose gels. Primer dimers can happen due to primers preferentially self-annealing or annealing to other primers in the reaction mix, or it can be a result of a higher concentration of primers compared to the DNA concentration, which leads to unspecific binding [199].

5.1.3 Amplification cycles

In order to obtain high enough concentrations of the PCR products in the last two studies, we used more than 30 amplification cycles [188, 226]. Indeed, it is common to use more than 30 amplification cycles for genetic and haplotype analysis in *Giardia* [60, 111, 112, 114]. The number of amplification cycles used in a thermocycler directly affects the error rate in the PCR-product, as any error introduced in earlier cycles becomes more pronounced when more cycles are performed [199]. A higher number of amplification cycles may also increase the risk of falsely introduced errors due to the polymerase's inherent error rate, and the more PCR product created, the higher the likelihood of errors (cumulative errors) [199]. However, one previous study using 45 cycles for amplification of *Giardia* DNA did not observe increased number of variations in the PCR products [60].

Some measures were taken to minimize errors in amplification which included using a high-fidelity proofreading polymerase Q5®. The Q5® is considered to be a very accurate polymerase [227] and up to 100-200 times [228, 229] more accurate than a Taq-polymerase, and may vary depending on the number of cycles used in the PCR and lastly, the length of the target gene. The error rate of Taq has previously been found to be ~1 in 3,500 bases while the error rate of Q5® is as low as 1 in 1 million bases [228, 229]. Clone-based approaches to identify and analyze SNVs in *Giardia* genes have been performed before [60, 111, 230]. Some of these studies have used Taq polymerases, that could have contributed to false-positive SNVs. One study of the bacterial *lacZ* gene has shown that a Taq polymerase can introduce as many as 28 % additional SNVs (recombination products) during 16 amplification cycles [228], and the importance of using a high fidelity polymerase should not be underestimated in future studies exploring SNVs.

5.1.4 SNV validation

Sanger sequencing of the cloned genes (*gFlHb, NR1* and *NR2*) in this project first relied on DNA amplification by PCR. Mistakes introduced during this initial PCR experiment and sequencing mistakes occurring during Sanger sequencing will appear in the sequence file data of later experiments and could potentially be falsely introduced

mutations, which can impact the genetic analysis [228].

Validation of each SNV was carried out for the *Giardia* isolates where both Illumina and Sanger sequences were available for the two clone-based studies. SNVs were categorized as high-confidence (HC) and low-confidence (LC) inspired by Boutte et al 2015 [231]. For the non-cultured isolates, without Illumina sequences, any SNV present in more than one of all the cloned sequences from any isolate was considered valid. The reported number of SNVs in both of the cloning studies may therefore be considered conservative.

We also observed that some of the SNVs identified by cloning were not classified as a SNV in the Illumina sequencing data. For the *NR1* gene, two additional SNVs were detected in isolate P344, while for *NR2* gene, one additional SNV was detected in isolate P413 and seven additional SNVs in isolate P344. For the *gFlHb* gene, more "clone-only" SNVs were found in both assemblages (AII: P033 with three, P064 with four, P368 with seven, P392 with ten, P407 with 11 and P478 with 19. B: P344 with three, P387 with six, P413 with nine and P427 with seven). Sanger sequencing supplements genetic analysis by being able to identify alleles and possibly more SNVs than Illumina sequencing.

The validation of SNVs using a combination of Illumina sequences and Sanger sequences may be termed a quality measure in the two cloning studies conducted in this thesis, as erroneous SNVs may be eliminated from the analysis. Still, the high number of LC SNVs identified for some of the isolates may indicate that the methods used are prone to errors and should be improved in future clone-based studies. In light of this, the isolate with most LC SNVs, P344, is most likely a mixed isolate.

5.1.5 Chimeras

Hybrid sequences that are partly one allele, and partly another, also called chimeric sequences, have been noticed in PCR reactions over 25 years ago, and the main contributor to this phenomenon is the number of amplification cycles used in an experiment. If over 30 cycles are used, a 30 % increase of chimeras can be noticed [232]. A recent study compared the two sequencing methods Illumina and Sanger to identify mixed *Giardia* infections. Illumina sequencing was found to be a better

method for detecting mixed infections. Further, the study reported a total of 21 % of the Illumina sequences to be chimeric sequences, while the number of Sanger sequencing chimeras was not reported [50]. It has been shown that WGA before sequencing leads to an increase of chimeras [233, 234], thus making chimera detection and removal in next generation sequencing (NGS) application an important measure. PCR-free kits were used prior to Illumina sequencing in this thesis, thus limiting the likelihood of chimeric sequences in the whole genome sequenced *Giardia* isolates.

In Paper III, chimeras were a particular challenge in heterozygous and mixed samples for the *NR* genes, where up to 31 % of sequences from a single isolate were chimeric. The cloned sequences of *gFlHb* were not originally checked for presence of chimeras, and a new analysis was done (data not shown) to check for this phenomenon. No chimeras were detected in the *gFlHb* sequences, and it's tempting to speculate that the alleles could have been too different from one another to function as primers, or that the alleles may be so similar that chimeras not would be noticed (not very likely according to the variation reported in Papers I and II).

Nevertheless, we identified non-chimeric homozygotic alleles of the single-copy *NR* genes in some of the *Giardia* isolates in addition to no chimeras in the *gFlHb* analysis, providing evidence that the method of PCR amplification, ligation, cloning and sequencing was suitable for obtaining specific and correct sequences.

5.1.6 Potential contamination

Some of the SNVs found in the clones were not expected to exist, based on the Illumina sequence data. One of the reasons why additional SNVs were identified, could be due to contamination of samples. Contamination may happen at any stage during the experiments, and cannot be out-ruled, as sterile conditions have not been used for all parts of the methods. Although, some countermeasures were used to avoid cross-contamination, and included the usage of filtered pipette tips, exposure of UV-light to LAF benches before master mixes were made, "DNA AWAY™" spray to disinfect equipment and running PCR-products in every other well on the agarose gels.

Some isolates were found to have identical alleles to other isolates' alleles such as P413 and P424 (*gFlHb* and *NR2* assemblage B) and P433 and P458 (*NR1* and *NR2*

assemblage B), which could mean contamination. However, some alleles of genes may be common in the parasite population and does not necessarily mean contamination. The cultured isolates of *Giardia* showing identical alleles were not cultivated and processed at the same day, or in some cases, the same year, which reduces the likelihood that identical alleles in different isolates were due to cross-contamination.

5.2 Study results

5.2.1 SNVs as potential markers of resistance

In the first study (Paper I), we showed that nsSNVs are common in the 29 genes related to MTZ-metabolism and detoxification in culturable clinical isolates of *Giardia* assemblages A and B. *Giardia* assemblage B presented higher numbers of SNVs in the 29 candidate genes, and this may indicate that assemblage B isolates can mutate faster than sub-assemblage AII isolates. Further, the higher frequencies of nsSNVs in assemblage B may be associated with a more diverse functional variation, higher prevalence of disease and even higher rates of MTZ treatment-refractory cases [17, 235].

Up until now, four different enzymes have been associated with MTZ metabolism in *Giardia*. Enzymes that contribute to activation of the prodrug MTZ include NR1, PFOR1 and (potentially PFOR2) and TrxR [70]. On the flip side, only a single enzyme, NR2, has been correlated to inactivation of the drug [70]. Taking this into consideration, nsSNVs leading to dysfunctional MTZ activating genes would potentially be beneficial to the parasite, while a dysfunctional inactivating gene, *NR2*, would most likely render *Giardia* more susceptible to MTZ. Genetic variation of the MTZ metabolism genes and detoxification genes in clinical samples of *Giardia lamblia* was the main focus of this thesis, and whether SNVs could serve as potential markers of resistance will be discussed below.

Both of the *NR* paralog genes in Paper I were found to have higher number of nsSNVs than the average. In the context of genetic variation, we further wanted to explore the alleles of these two genes in non-cultured clinical *Giardia* isolates as well. In Paper III the two *NR* paralogs of *Giardia* were investigated, and we found presence of mutations that potentially could alter the MTZ metabolizing capacities of the parasite, including dysfunctional NR1 and NR2 enzymes [226]. The dysfunctional NR1 enzyme

was found in a MTZ-refractory isolate, and could mean that the parasite will be less efficient at activating MTZ, and thereby lead to less toxic effects of the drug, and lastly make *Giardia* more tolerant to MTZ. A recent study conducted by Ansell et al in 2017 described a nonsense mutation in some transcripts of NR1 in a laboratory induced MTZ resistant isolate in addition to potentially increased transcription of NR2 and a MATE transmembrane efflux pump system [186]. Interestingly, the exact same mutation for NR1 was found in Sanger sequenced clones in our study as well [226]. Agreeably, premature stop codons have been detected in the two *NR* genes $ntr4_{TV}$ and $ntr6_{TV}$ of MTZ resistant *T. vaginalis* [194]. However, 11 NRs have been found in *T. vaginalis*, and some of these enzymes may be of higher relevance with respect to MTZ activation and inactivation [194].

To sum up, nonsense mutations in the MTZ-activating gene, *NR1*, should be explored for its relevance to MTZ resistance in future studies, preferably with a larger collection of clinical MTZ resistant *Giardia* isolates, to assess whether this specific mutation could serve as a potential MTZ resistance marker [226].

Two other studies have reported down-regulation of NR1 in three resistant *Giardia* strains [186, 236], thus supporting our perception that dysfunctional NR1, or for this specific case, lower levels of the enzyme, could protect *Giardia* against the toxic effects of MTZ. Correspondingly, genetic markers of resistance have been identified in *Helicobacter pylori* and *Escherichia coli*. In three studies, the MTZ-activating enzyme oxygen-insensitive nitroreductase, *rdxA,* (NTR-1 also known as NR3) was associated with MTZ resistance, due to either an inactivated NTR enzyme (null mutation) or presence of nsSNV [192, 193, 237]. An oxygen-insensitive nitroreductase (*NTR-1*), similar to NTR in bacteria, exist in *Giardia* as well [51], though different from NR1 and NR2 due to its lack of ferredoxin (fd) domains which consequently mean that this enzyme probably does not play a key role in the electron transport [51, 170]. In Paper I the number of nsSNVs in the NTR-1 gene was low compared to the average number of nsSNVs, and it is not fully known whether this enzyme has an important role in MTZ activation in *Giardia*. A more recent study did find that the NTR-1 may harbor similar characteristics as those of NR1, as *Giardia* overexpressing this gene was more susceptible to MTZ compared to the control [170]. Additionally, the enzyme may

contribute to the protection against oxidative- and nitrosative stresses, as the enzyme can handle high levels of oxygen, and the gene was found to be up-regulated during MTZ exposure [69].

A recent study by Galeh et al 2016 [238], where clinical *Giardia* isolates were investigated, suggested that nsSNVs in partial, short-stretch alleles of the two genes *NR2* (here more commonly known as NR1 according to the gene sequence though) and *PFOR* could be linked to MTZ resistance, although this must be interpreted cautiously, as only two MTZ treatment-refractory *Giardia* isolates were explored. The *NR2* gene investigated in Paper III was found to have a higher number of nsSNVs and more distinct alleles per sample compared to the *NR1* gene [226]. In *Giardia* samples where clinical patient information was available, we found a total of two susceptible samples with nucleotide deletions, leading to truncated, dysfunctional NR2 proteins. Interestingly, one of the susceptible isolates (VA) was homozygotic, meaning a total loss of NR2 function. The considerable variation, and findings of mutations leading to dysfunctional NR proteins in recent clinical *Giardia* assemblage B isolates show the potential for genetic alterations affecting MTZ susceptibility. A dysfunctional NR2 enzyme may potentially serve as a susceptibility marker of MTZ tolerance in *Giardia.* The effect on MTZ of genetic alterations should further be studied using various approaches that could include a combination of agar diffusion assays and exploration of the survival in MTZ drug exposure assays of cultured *Giardia* trophozoites. In an agar diffusion assays dysfunctional genes could be cloned into BL21 *E. coli* and the inhibition zones around MTZ disks could be measured to test the susceptibility of the NRs to further understand more about the resistance mechanism in *Giardia* [170].

In Paper I the MTZ-activating gene PFOR1 in addition to PFOR2 had lower than average numbers of nsSNVs, corresponding to preservation of the proteins' function, while the Fd co-factors of PFOR were highly variable (three of the five fds in assemblage B and two of the fds in sub-assemblage AII) [28]. The role of the MTZ-activating PFOR is believed to involve transfer of electrons, through the enzymes' bound Fe-S clusters to their co-factors fds [160]. Accordingly, lower levels of PFOR have been identified in MTZ resistant *Giardia* laboratory strains [162, 186, 187, 190]. Still, PFOR gene expression has not been found to be of significant difference in clinical

patient isolates with different tolerance to MTZ *in vitro*, and coupled with low levels of nsSNVs, these genes may not be likely to harbor markers of resistance [239].

As MTZ may only be activated in anaerobic or microaerophilic environments, *Giardia* may use other strategies to avoid the toxic effects of free radicals created during MTZ activation [144]. MTZ may be inactivated during the presence of oxygen due to a mechanism known as 'futile cycling', although by creating ROS that *Giardia* needs to neutralize through enzymes in its antioxidant network system [70, 187, 240]. In Paper I, high numbers of SNVs were found in the oxygen detoxifying protein NADH oxidase (GSB_9719) and in the hydrogen peroxide detoxifying protein thioredoxin peroxidase in *Giardia* assemblage B isolates. Interestingly, the ROS contributing enzyme, NADPH oxidoreductase putative (GSB_17150), had higher numbers of nsSNVs than the average for both *Giardia* assemblages.

Some of the genes of Paper I (fd6 and Histone H2A of assemblage B) were found to be highly conserved [28]. Introduction of nsSNVs in conserved genes, may lead to more severe consequences for the parasite as well as affecting biological functions including survival in culture, clinical infectivity, and possibly affecting tolerance to free radicals, and should be explored in future studies [28].

5.2.2 Amino acid changes

Previous genetic studies have reported high numbers of nsSNVs in house-keeping genes, especially for assemblage B. However, these SNVs most commonly caused conservative amino acid changes, and are not likely to affect protein function or structure [111, 115].

Both of the NR proteins in *Giardia* contain two domains, a fd domain and a reductase domain. The fd domain contains cysteine residues able to wind two Fe-S clusters which form part of redox-active centers responsible for biological electron transport [186, 241]. Most of the nsSNVs within and surrounding the fd domains will probably not alter the function of the enzymes, as the amino acid changes probably do not alter the basic properties, e.g., polarity or charge [28, 226]. However, one of the cysteines in the fd domain of NR2 alleles in a refractory *Giardia* isolate was substituted with an arginine, likely disrupting the protein's Fe-S binding capabilities [226].

Although this mutation contradicts our hypothesis with a potentially dysfunctional NR2 for a refractory isolate, this sample most like represents mixed isolates due to its total nine distinct alleles of NR2 [226].

Some other amino acid changes were located in close proximity to the cysteine part of the Fe-S binding motif, may hypothetically affect the geometry or potentially the electron transfer properties of the domain. The activated form of the MTZ has previously been hypothesized to be able to react with cysteines in the cytosol and further form adducts [160]. This potential adduct-forming effects on the proteins analyzed in Paper I and the NRs of Paper III should be investigated in future studies. Further, if amino changes close to or within the fd domains could affect the formation of adducts should be addressed in future studies using structural- and molecular biology studies [28, 226].

The *gFlHb* genes of sub-assemblage AII and assemblage B in Paper II was found to have a higher number of nsSNVs that potentially could alter the proteins' amino acid sequences. Still, the crystal structure of the gFlHb protein is not yet known, and all of the structural mapping of the location of the mutated amino acids was therefore based on homology models [188]. A premature stop codon was found in two of the eight alleles of sub-assemblage AII clones, meaning that this isolate, P478, harbored alleles of dysfunctional *gFlHb* [188]. Further, interpretation of mutations located on the surface of the gFlHb protein is more challenging with regards to predicting the potential effects. However, altering of the surface charge of the protein could hypothetically have an effect on inter-molecular interactions. Further research is needed to investigate whether the amino acid changes can affect the function of the identified protein variants.

5.2.3 Copy number variation of gFlHb

We first noticed, in the work with Paper I, that the antioxidant and detoxification gene, *gFlHb,* seemed to have a higher copy-number coupled with a high genetic variation than the other 28 genes, due to the gene's much higher coverage in Illumina sequencing reads [28].

gFlHb is an important NO and O_2 scavenging enzyme [62, 63, 65, 186], and one study has shown that gFlHb is up-regulated both during O_2 and H_2O_2 exposure [63]. We found evidence of *gFlHb* being present up to six times in one of the *Giardia*

isolates (P424) in Paper II [188]. It may be beneficial for *Giardia* to harbor several copies of the gene during oxidative stress created by drugs, and indeed, a recent study reported up-regulated gFlHb protein levels in an MTZ and nitazoxanide (NTZ) resistant *Giardia* isolate [76]. More numerous copies of the *gFlHb* could potentially promote survival in an O_2 rich environment if a *Giardia* strain is depending on 'futile cycling' of MTZ. In fact, gFlHb was found to be strongly induced in a MTZ resistant *Giardia* line (713-r), and the authors further proposed that gFlHb could constitute an alternative MTZ detoxification enzyme [186]. Still, this finding may depend on which assay is used, and it is probably plausible in specific strains of *Giardia*, as we showed that various *Giardia* isolates harbor variable copy numbers of *gFlHb* [188]. The variable copy numbers of *gFlHb* in a *Giardia* isolate could also mean that *Giardia* could utilize different MTZ tolerance formation strategies, meaning that isolates with high *gFlHb* copy numbers may more easily utilize an active detoxifying MTZ resistance strategy, while other *Giardia* isolates may utilize a reduced MTZ activation strategy [189]. We do not, however, know how, and if all of the *gFlHb* copies in *Giardia* would be expressed during oxidative stress [188], and further studies combining methods such as single cell DNA and RNA sequencing should be carried out to understand the effects of having multiple *gFlHb* copies. Digital droplet PCR of the *gFlHb* gene was tested to support our findings in Paper II, though due to methodological issues and reproducibility, the results could not be interpreted, and was not included in the study. Further southern blots could have supported our analysis, by being able to identify copy numbers, however, it might not be accurate enough to identify more than two copies of a gene [242].

Giardia is a tetraploid organism, and the maximum number of alleles that can be found in one parasite will be four times the number of copies of a gene. In some of the sub-assemblage AII isolates, we saw that a higher copy number correlated well higher numbers of alleles (P064; two copies and two alleles, P407; three copies and eight alleles and P392; four copies and eight alleles), however for assemblage B, this was not observed (P344 with two copies and nine alleles vs P424 with six copies and nine alleles and P458 with three copies and two alleles). Still, we have to consider that some rare alleles may have been missed due to e.g., SNVs in primer binding regions.

5.2.4 Alleles

In Paper II, *gFlHb* was found to have high degrees of genetic variation in both of the *Giardia* assemblages. One interesting finding was that some of the sub-assemblage AII isolates possessed more SNVs than assemblage B isolates. Further, relatively high allelic heterozygosity was found for both assemblages, while the nucleotide diversity was low (<0.01). Agreeably, other studies have investigated the nucleotide diversity of single-copy genotyping genes such as glutamate dehydrogenase (*gdh*), betagiardin (*bg*) and triosephosphate isomerase (*tpi*), and found similar nucleotide diversity values ranging from 0.003 to 0.02 [112, 114].

Several alleles of the two *gFlHb* genes were identified. We found a total of 37 alleles in eleven sub-assemblage AII isolates, while 41 alleles were found in six assemblage B isolates. Further, the number of *gFlHb* alleles in sub-assemblage AII isolates varied from two to eight, while two to twelve alleles per isolate were found in assemblage B isolates. Fewer *Giardia* assemblage B DNA samples with sufficient DNA concentrations were available and we were not able to collect as many clones for some of them, compared to sub-assemblage AII isolates. We did not find suspect any inter-assemblage recombination between the two assemblages, as no sub-assemblage AII derived alleles were detected in assemblage B isolates, or vice versa [188]. Further, we found presence of intra-assemblage recombination as some of the alleles were found to be present in more than one single isolate [188].

In Paper III, we identified a total of 69 distinct alleles for the *NR1* gene, and a total of 90 distinct alleles in the *NR2* gene in 15 clinical isolates of *Giardia.* Presence of inter-assemblage recombination was found in both of the genes, however, no shared alleles were detected between the non-cultured samples of *Giardia* for *NR1*, while a refractory and a susceptible isolate and a cultured and non-cultured samples shared alleles for *NR2*. The number of distinct *NR* alleles was exceptionally high for three of the cultured isolates (P344, P427 and P387) and also for two of the non-cultured *Giardia* samples (Ag10 and Ag13).

It has previously been shown that it is challenging to culture clinical isolates of *Giardia*, especially assemblage B isolates [23, 81]. *Giardia* isolates capable of culturing could potentially have an affected metabolism of MTZ as well as different

detoxification strategies of free radicals and O_2 and may only present a small selection of isolates capable of causing disease in humans. The cultured isolates used in this thesis could therefore be biased for the number of identified *gFlHb* copies, or be the reason for the low nucleotide diversity identified, and lastly; be the reason for many different, yet similar, alleles of *gFlHb* exist due to acquisition of SNVs during selective culturing [108].

All things considered, more than the expected maximum four haplotypes per gene has been found in previous studies investigating single copy genes (e.g., *bg*, *gdh*, *mlh* and *tpi*) [60, 111, 112, 114, 115]. At present, we cannot rule out that our findings of more than four alleles could be due to unidentified additional copies of the genes in the *Giardia* genomes. However, the high number of alleles detected in these isolates, may be explained by mixed isolates.

5.2.4.1 Mixed infections

Several studies of alleles of *Giardia* genes have assumed that the presence of more than the four expected alleles for single copy genes in single isolates, most likely occur due to mixed infections (e.g., two sub-assemblages) that potentially can take place by a subsequent infection of an already infected host, with a different *Giardia* parasite [22, 243, 244] or perhaps due to putative genic recombination [60, 111, 112]. The findings of more than four alleles for single isolates for the *gFlHb* and *NR* genes are probably due to a combination of reasons that can include mixed infections of sub-assemblages, intra-isolate genetic recombination or be artificial PCR-induced errors caused by a high number of amplification cycles. However, our findings support the idea that high numbers of distinct alleles reflect small genetic variations that most likely have occurred during culturing and given rise to co-existence of daughter lineages as presented by Choy et al [108]. The cultured isolates were obtained by limiting dilution, and the risk of having more than one isolate will be relatively small. Further, selective pressure during axenization of *Leishmania donovani* has led to remarkable genetic variation as well [245-247].

For one of the isolates included in all three Papers (P344) in this thesis, two different batches of DNA were used for the initial Illumina sequencing and the final

cloning experiments. This particular isolate may represent a mixed isolate as it showed the highest number of LC SNVs in Paper II and the highest numbers of alleles identified for the *NR* genes in Paper III.

5.2.5 Nucleotide diversity and haplotype diversity

In Paper II we analyzed the nucleotide diversity (pi) in addition to haplotype diversity. Selection during culturing may be found as deviations from neutrality by i.e., nucleotide diversity or the frequency of allele distribution [248].

The pi values of the NR genes (not calculated in Paper III, calculated for the purpose of comparison to *gFlHb,* and was 0.02 for both genes) were similar to those of *gFlHb.* The nucleotide diversities were low for all of the genes and this can be expected for cultured isolates, as populations that have gone through bottlenecks have been found to have a low nucleotide diversity [249]. Even if some of the isolates were not cultured, they represent a small fraction of all of the available alleles.

The haplotype diversity (referred to as allelic diversity in Paper II) showed relatively high values, and may correspond to many different alleles that are not very different from one another [188]. The haplotype diversities of the *NRs* was not calculated in Paper III, but were similarly high as *gFlHb* (0.99).

Further we ran Tajima's D calculations on the genes and found values of -0.9 for sub-assemblage AII and -0.8 for assemblage B for *gFlHb*, and 0.3 for NR1 and -0.4 for NR2 from assemblage B. No significant values were found.

Still, we have to take into consideration that we only included a few isolates in this thesis. Several genetic studies investigating nucleotide diversity, haplotype diversity have done so on a larger scale (population-based), thus including many organisms/isolates with one gene each rather than few isolates with many alleles [250-252].

5.2.6 The reference genomes of *Giardia*

The reference genes we have used in this thesis belongs to some particular laboratory strains of *Giardia.* The allele of a gene represented by the reference may not therefore be the major allele and could even present a minor allele.

The reference genomes of *Giardia* assemblages A and B used in this thesis may have affected the SNV analysis, as the isolates used for obtaining the reference genomes have been cultured over a long period of time. Additionally, the genomes are preliminary and consist of several hundred smaller contigs, instead of the five predicted chromosomes of the parasites. Better annotated reference genomes of the DHA and GSB, similar to that of the freshly sequenced WB genome, would certainly be important for future genetic diversity studies [86]. Further, MTZ tolerance studies that have used reference strains of *Giardia,* where MTZ has been slowly induced in the laboratory, may not reflect the mechanisms behind the increase of MTZ treatment refractory infections observed during the last few years [156, 183, 184].

5.2.7 Drug resistance in *Giardia*

Individuals of all ages may be infected with *Giardia* and the infection contributes significantly to the global burden of diarrheal disease, as well as post-infectious sequelae and possibly stunting [20, 32, 125, 184]. A former study has proposed that there are at least six potential scenarios that can lead to treatment failures of giardiasis and include I) reinfection, II) inadequate levels of drug, III) immunosuppression of patients, IV) microbial resistance to the drug, V) sequestration in the gallbladder or pancreatic ducts of the patient and VI) unknown reasons [253]. Some studies report that individuals with diseases such as common variable hypogammaglobulinemia and lymphoproliferative diseases involving the gastrointestinal tract [254-256] seem more prone to *Giardia* infections, and infections in such patients are likely also more difficult to cure [253]. Further, HIV and AIDS infected individuals seem more susceptible for infection with *Giardia* in some studies [142, 257], however this may also be contributed to epidemiological factors rather than immunosuppression [142]. Immunocompromised individuals infected with *cryptosporidium* can often suffer intractable diarrhea, sometimes with fatal outcomes [258].

Fortunately, giardiasis is a treatable infection, and patients not responding well to treatment will usually be offered multiple courses of standard drugs (see introduction section) in order to cure the infection [253, 259]. *Giardia* may be self-limiting in some individuals, and one small study with 14 participants has reported that the mean duration

of infection was approximately three weeks before 85 % (N=12) of the individuals would be cured, while 15 % (N=2) experienced chronic disease [8]. Despite the high self-limiting rate, it is rational to treat the infection to avoid chronic disease, sequelae and limit the spread of the parasite [260], at least in developed countries with less chance of re-infection. However, an increasing number of treatment-failures have been observed over the last few years [149, 152, 156, 261].

No single molecular resistance mechanism has been defined yet, and the term treatment refractory *Giardia* is better suited to explain treatment failures, due to all possible causes that may lead to persistence of *Giardia* infection (see above) [260]. Most studies investigating resistance mechanisms in *Giardia* have been performed on decades old laboratory strains of the WB isolate [262], and laboratory-induced resistance can be lost during ex/endocystation [263]. The laboratory-induced drug resistance, could potentially be due to metabolic adaptations and the word 'tolerance' may be better suited to explain this phenomenon, while "resistance" is normally used for lines/isolates that carry a well-documented ability to resist a specific drug [32]. In *Giardia*, no such markers are currently known, but could be genotypic changes, or drug deactivating characteristics, such as nonsense mutations observed in our final study [32, 226]. Still, clinical refractory *Giardia* isolates may represent other resistance mechanisms than the laboratory-derived ones, and a recent increase of treatment-refractory *Giardia* may be explained by occurrence of new genetic heritable traits [260]. Further, former gene expression studies on resistant *Giardia* have reported rather broad and variable responses, where posttranscriptional and posttranslational such as acetylation, methylation and phosphorylation and alterations of genes may be involved in the resistance mechanism in addition to redox-sensitive epigenetic alterations [68, 186, 236, 263, 264]. It is clear that further studies for treatment-refractory *Giardia* isolates are needed to investigate the rather rapid increase of treatment refractory *Giardia* cases [156, 184].

Treatment failure of giardiasis has previously been defined as the presence of the parasite in at least one of three consecutive stool samples of an infected patient, coupled with symptoms after completion of one or more courses of standard treatment [259]. Still, the methods for determining treatment refractory *Giardia* isolates are scare, mostly

due to difficulties of obtaining successful cultures of clinical patient samples, and different growth rates of the assemblages [185, 265, 266]. The symptoms of giardiasis are variable [32, 40], and it is important to distinguish between post-infectious symptoms and recurrence of infection. The treatment outcomes of *Giardia* infections should further be verified in studies investigating MTZ-refractory isolates, where a successful treatment would mean a negative PCR a few weeks after initial treatment.

Drug susceptibility assays for other organisms such as bacteria, may be more easily obtained as standard agar diffusion assays and MIC testing can be done *in vitro,* which can provide a relatively rapid test result. Bacteria are most commonly haploid organisms, have rapid replication time and may be grown to large quantities easily [267]. Bacteria have smaller genomes than eukaryotes, which means that genomic analyses may be carried out more easily [267]. As *Giardia* is a functionally tetraploid organism with up to four alleles of each gene, the genomics are more complex, and genetic variation regarding refractory disease may be difficult to interpret [101]. The general genetic variation of MTZ metabolism genes has been showed to be high in this thesis, and the parasite's potential for variability coupled with its tetraploidity may makes SNV analysis harder [28].

MTZ resistance induced in the laboratory will potentially have different manifestations than the resistance observed for clinical isolates, the use of *Giardia* isolates directly from patients, without undergoing culturing will most likely supplement former studies where isogenic laboratory strains have been used [260].

5.2.8 Limitations

The number of *Giardia* isolates used in this thesis is low, and for the whole genome sequenced isolates, clinical treatment data was not available. Still, none of those samples were suspected to be treatment-refractory. The non-cultured clinical isolates analyzed in Paper III represent only a few refractory and susceptible isolates, and several samples had to be excluded due to too low DNA concentrations to be used in cloning experiments. We propose that samples with few *Giardia* cysts may be successfully included in future genetic studies by performing multiple IMS purifications, and then combine them before DNA extraction.

All of the amplified *gFlHb* genes in Paper II and some of the *NR* genes in Paper III were obtained from trophozoites growing and replicating in culture, which could be termed as unnatural conditions. We have to admit that differences between *Giardia* from clinical samples and cultured *Giardia* cannot be excluded. The fact that the *Giardia* B assemblages used in this thesis have been successfully cultured, originating from a former study that found assemblage B isolates rather difficult to culture [185], may indicate that our study population could be biased.

For some of the isolates in Papers II and III, small numbers of clones were obtained, thus, rare alleles may have been missed in some isolates. One explanation for the low number of obtained clones could be SNVs in the primer binding regions of the genes. Potential falsely introduced SNVs induced by many amplification cycles in Papers II and III should also be noted as a weakness, even if several measures including using a high-fidelity polymerase and discarding LC SNVs were done to limit its impact during the experiment setup and analysis.

The nsSNVs analyzed in this thesis were all limited to the CDS of the genes, while the number of SNV outside the CDS was reported in Paper I. SNVs in non-CDS regions of genes could possibly be important for regulation and promotion of a gene, thus directly linking it to transcription, mRNA translation and even protein expression [28, 236, 268]. DNA sequences preceding the CDS, harbor transcriptional promoters, and these promoter regions include two AT-rich sequences crucial for transcription, approximately −65 to −29 base pairs upstream the ATG start codon, where one is known as TATA-like box, in addition to an element resembling a CAAT box approximately -50 bp upstream the CDS [6, 269-272]. Still, how the presence of SNVs in promoter regions may affect gene transcription is not yet known, and knowledge for interpretation of these SNV are limited. However, future studies should explore how SNVs in promoter regions may affect gene transcription, especially in MTZ refractory *Giardia* isolates.

Further, functional studies have not been carried out to decipher the potential effects of the amino acid changes of the proteins in this thesis. The effect of the amino acids changes are thus hypothetical and only based on homology models or the theoretical effects of the changes. Functional studies of the MTZ metabolism genes

could provide important information in future prospective studies.

6. CONCLUSION

The majority of previous studies regarding genetic variation and MTZ tolerance in *Giardia* have been carried out on old laboratory strains of the sub-assemblage AI, which rarely infects humans. The work performed in this thesis has added more knowledge about general genetic variation in MTZ metabolism and detoxification genes in clinical isolates of the parasite from both human infecting assemblages (AII and B). We have established some workflows and introduced analytical methods for investigating alleles of a variable copy number gene, and of two MTZ-metabolism genes, that have not been done before in *Giardia*. Further, some mutations may serve as markers of resistance in future experiments investigating MTZ tolerance in *Giardia*. A short summary of the key findings in this thesis is included below:

- Both of the human-infecting assemblages, A and B, of *Giardia* commonly present genetic variation in the form of single nucleotide variants (SNVs). The degree of genetic variation of assemblage B supports earlier studies, that this assemblage is more heterogenous than assemblage A. Nonsynonymous (ns)SNVs in assemblage B were especially high in genes related to MTZ metabolism and detoxification of reactive oxidative species (ferredoxins (*fds*), nitroreductase (*NR*) 1 and 2, thioredoxin peroxidase and NADH oxidase). Some of the identified nsSNVs in this thesis could potentially affect electron transfer properties of Iron-Sulfur clusters in some of the NRs and fd proteins.
- The flavohemoprotein (*gFlHb*) gene in both of the human infecting *Giardia* assemblages, A and B, is a variable copy number gene with high allelic diversity. The multi-copy nature of the *gFlHb* may affect how a *Giardia* strain can adapt to nitrosative and oxidative stress responses, as gFlHb is an important detoxification enzyme, and having several copies of this gene could theoretically affect the MTZ susceptibility.
- The two MTZ-metabolizing genes *NR1* and *NR2* from *Giardia* assemblage B were found to have considerable degrees of genetic variation. On one side, dysfunctional NR2 enzymes were found in susceptible isolates and may affect MTZ susceptibility in *Giardia*, while on the other side, dysfunctional NR1

enzymes found in a MTZ-refractory isolate may contribute towards resistance in the parasite. Further the identified amino acid changes in close proximity to the fd domains may affect the functions of the proteins and could possibly lead to increased MTZ tolerance.

7. FURTHER RECOMMENDATIONS

There are yet many unanswered questions about genetic variation in *Giardia,* and more importantly, about genetic markers of metronidazole (MTZ) resistance in this parasite. The following topics could be addressed in further studies:

Paper I can be regarded as a baseline for future studies into variability of genes in the MTZ-metabolizing pathways between refractory and susceptible *Giardia* isolates. A high degree of genetic variation was found in the NRs and fds, and future studies are needed to understand their functional properties. Analysis of the genetic variation in recent and circulating MTZ susceptible and especially treatment-refractory *Giardia* isolates will most likely be important to further understand the recent increase of MTZ resistant *Giardia*.

The multi-copy nature of gFlHb could be important for MTZ susceptibility, as this enzyme is capable of managing stress responses such as O_2 and NO. Future studies should test MTZ susceptibility for *Giardia* isolates harboring different copies of the gene, as gFlHb recently was found to be up-regulated in a MTZ-resistant laboratory strain of *Giardia*. Using RNA sequencing one can also profile which gFlHb alleles becomes upregulated.

In light of the findings from the NR cloning work, where dysfunctional NR1 and NR2 were observed, several follow-up studies have been initiated by The Bergen *Giardia* Research group. The follow-up studies includes analyzing alleles of the two PFOR genes and TrxR in addition to extraction of DNA from more refractory *Giardia* samples and clone genes such as *NR1*, *NR2*, *NTR-1*, *PFOR1*, *PFOR2* and *TrxR* in order to look for potential markers of resistance in a larger and better clinically defined patient cohort.

Better reference genomes of sub-assemblage AII and assemblage B (longer contigs and better annotated), would be important for future genetic studies of the parasite. More continuity in the contigs would make it easier to assess multi-copy genes as well.

A search for genetic markers of MTZ resistance should be carried out in larger

patient cohorts, followed by functional assays. This would be beneficial to examine whether, and how, genetic alterations such as variable copy number genes and nonsense mutations may influence MTZ susceptibility.

Some of the gene variants of *gFlHb* and *NR1* and *2* can further be tested in functional assays, where some specific genes could be edited using CRISPR/Cas9 methods in cultured reference strains of *Giardia* [273], or specific variants of NRs could be tested in MTZ disk diffusion assays [170]. Additionally, *in vitro* protein assays can be useful for characterizing purified protein variants, such as measuring the NO consumption by membrane-inlet mass spectrometry of the gFlHb enzyme [274].

8. REFERENCES

1. Dobell, C. and A.v. Leeuwenhoek, *Antony van Leeuwenhoek and his "Little animals"; being some account of the father of protozoology and bacteriology and his multifarious discoveries in these disciplines*. 1932, New York: Harcourt, Brace and company.
2. Boreham, P.F., J.A. Upcroft, and P. Upcroft, *Changing approaches to the study of Giardia epidemiology: 1681-2000.* Int J Parasitol, 1990. **20**(4): p. 479-87.
3. Dobell, C., *The Discovery of the Intestinal Protozoa of Man.* Proc R Soc Med, 1920. **13**(Sect Hist Med): p. 1-15.
4. Butschli, O. and C. Schwager, *Protozoa: Abt. Sarkodina und Sporozoa*. 1887: C.F. Winter.
5. Farthing, M.J., *Giardiasis.* Gastroenterol Clin North Am, 1996. **25**(3): p. 493-515.
6. Adam, R.D., *Biology of Giardia lamblia.* Clin Microbiol Rev, 2001. **14**(3): p. 447-75.
7. Knight, J., *Giardia: not so special, after all?* Nature, 2004. **429**(6989): p. 236-7.
8. Rendtorff, R.C., *The experimental transmission of human intestinal protozoan parasites. II. Giardia lamblia cysts given in capsules.* Am J Hyg, 1954. **59**(2): p. 209-20.
9. Karapetyan, A., *In vitro cultivation of Giardia duodenalis.* J Parasitol, 1962. **48**: p. 337-40.
10. Lipoldová, M., *Giardia and Vilém Dušan Lambl.* PLoS Negl Trop Dis, 2014. **8**(5): p. e2686.
11. Sogin, M.L., et al., *Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia.* Science, 1989. **243**(4887): p. 75-7.
12. Saghaug, C.S., *Developing a flow cytometric method to characterize human T cell responses against the enteric protozoan parasite Giardia lamblia*, in *Department of Clinical Science*. 2014, University of Bergen: Bergen. p. 127.
13. Monis, P.T., S.M. Caccio, and R.C. Thompson, *Variation in Giardia: towards a taxonomic revision of the genus.* Trends Parasitol, 2009. **25**(2): p. 93-100.
14. Ankarklev, J., et al., *Behind the smile: cell biology and disease mechanisms of Giardia species.* Nat Rev Microbiol, 2010. **8**(6): p. 413-22.
15. Caccio, S.M. and U. Ryan, *Molecular epidemiology of giardiasis.* Mol Biochem Parasitol, 2008. **160**(2): p. 75-80.
16. Adam, R.D., *The Giardia lamblia genome.* Int J Parasitol, 2000. **30**(4): p. 475-84.
17. Feng, Y. and L. Xiao, *Zoonotic potential and molecular epidemiology of Giardia species and giardiasis.* Clin Microbiol Rev, 2011. **24**(1): p. 110-40.
18. Monis, P.T., et al., *Genetic diversity within the morphological species Giardia intestinalis and its relationship to host origin.* Infect Genet Evol, 2003. **3**(1): p. 29-38.
19. Cotton, J.A., J.K. Beatty, and A.G. Buret, *Host parasite interactions and pathophysiology in Giardia infections.* Int J Parasitol, 2011. **41**(9): p. 925-33.
20. Robertson, L.J., et al., *Giardiasis--why do the symptoms sometimes never stop?* Trends Parasitol, 2010. **26**(2): p. 75-82.
21. Zahedi, A., D. Field, and U. Ryan, *Molecular typing of Giardia duodenalis in humans in Queensland - first report of Assemblage E.* Parasitology, 2017. **144**(9): p. 1154-1161.
22. Sprong, H., S.M. Cacciò, and J.W. van der Giessen, *Identification of zoonotic genotypes of Giardia duodenalis.* PLoS Negl Trop Dis, 2009. **3**(12): p. e558.
23. Binz, N., et al., *Comparative studies on the growth dynamics of two genetically distinct isolates of Giardia duodenalis in vitro.* Int J Parasitol, 1992. **22**(2): p. 195-202.
24. Matey, E.J., et al., *Positive correlation of HIV infection with Giardia intestinalis assemblage B but not with assemblage A in asymptomatic Kenyan children.* Aids, 2016. **30**(15): p. 2385-7.
25. Bénéré, E., et al., *Intestinal growth and pathology of Giardia duodenalis assemblage subtype AI, AII, B and E in the gerbil model.* Parasitology, 2012. **139**(4): p. 424-33.

26. Adam, R.D., et al., *Genome sequencing of Giardia lamblia genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig).* Genome Biol Evol, 2013. **5**(12): p. 2498-511.
27. Jerlstrom-Hultqvist, J., et al., *Genome analysis and comparative genomics of a Giardia intestinalis assemblage E isolate.* BMC Genomics, 2010. **11**: p. 543.
28. Saghaug, C.S., et al., *Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical Giardia lamblia assemblage A and B isolates.* Infect Drug Resist, 2019. **12**: p. 1221-1235.
29. Franzen, O., et al., *Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species?* PLoS Pathog, 2009. **5**(8): p. e1000560.
30. Collaborators, G.C.o.D., *Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016.* Lancet, 2017. **390**(10100): p. 1151-1210.
31. Argüello-García, R., et al., *Chapter Six - Drug resistance in Giardia: Mechanisms and alternative treatments for Giardiasis*, in *Advances in Parasitology*, M.G. Ortega-Pierres, Editor. 2020, Academic Press. p. 201-282.
32. Lalle, M. and K. Hanevik, *Treatment-refractory giardiasis: challenges and solutions.* Infect Drug Resist, 2018. **11**: p. 1921-1933.
33. Roxström-Lindquist, K., et al., *Giardia immunity--an update.* Trends Parasitol, 2006. **22**(1): p. 26-31.
34. Lanata, C.F., et al., *Global causes of diarrheal disease mortality in children <5 years of age: a systematic review.* PLoS One, 2013. **8**(9): p. e72788.
35. Ryan, U., et al., *Giardia: an under-reported foodborne parasite.* Int J Parasitol, 2019. **49**(1): p. 1-11.
36. Craun, G.F., *Waterborne Outbreaks of Giardiasis*, in *Giardia and Giardiasis: Biology, Pathogenesis, and Epidemiology*, S.L. Erlandsen and E.A. Meyer, Editors. 1984, Springer US: Boston, MA. p. 243-261.
37. Wensaas, K.A., N. Langeland, and G. Rørtveit, *[Uncovering the giardiasis-outbreak in Bergen 2004].* Tidsskr Nor Laegeforen, 2007. **127**(17): p. 2222-5.
38. Broglia, A., et al., *Molecular typing of Giardia duodenalis isolates from German travellers.* Parasitol Res, 2013. **112**(10): p. 3449-56.
39. VKM, et al. *Giardia og Cryptosporidium i norsk drikkevann*. Vitenskapelig uttalelse fra faggruppe for hygiene og smittestoffer i Vitenskapskomiteen for mat og miljø 2020; 2020/06/19:[1-38]. Available from: https://vkm.no/download/18.1243fd9b172b80f602451d1a/1613375165832/Giardia%20og%20Cryptosporidium%20i%20norsk%20drikkevannn.pdf.
40. Leder, K. and P.F. Weller, *Giardiasis: Epidemiology, clinical manifestations, and diagnosis*, E.T. Ryan, Editor. 2021, UpToDate.
41. Cacciò, S.M., M. Lalle, and S.G. Svärd, *Host specificity in the Giardia duodenalis species complex.* Infect Genet Evol, 2018. **66**: p. 335-345.
42. Leonhard, S., et al., *The molecular characterisation of Giardia from dogs in southern Germany.* Vet Parasitol, 2007. **150**(1-2): p. 33-8.
43. Liu, H., et al., *Prevalence and multilocus genotyping of potentially zoonotic Giardia duodenalis in pigs in Shanghai, China.* Parasitology, 2019. **146**(9): p. 1199-1205.
44. Khan, S.M., et al., *Molecular evidence for zoonotic transmission of Giardia duodenalis among dairy farm workers in West Bengal, India.* Vet Parasitol, 2011. **178**(3-4): p. 342-5.
45. Tsui, C.K., et al., *Beaver Fever: Whole-Genome Characterization of Waterborne Outbreak and Sporadic Isolates To Study the Zoonotic Transmission of Giardiasis.* mSphere, 2018. **3**(2).
46. Ryan, U. and S.M. Cacciò, *Zoonotic potential of Giardia.* Int J Parasitol, 2013. **43**(12-13): p. 943-56.
47. Cooper, M.A., et al., *Molecular analysis of household transmission of Giardia lamblia in a region of high endemicity in Peru.* J Infect Dis, 2010. **202**(11): p. 1713-21.

48. Upcroft, J.A. and P. Upcroft, *Two distinct varieties of Giardia in a mixed infection from a single human patient.* J Eukaryot Microbiol, 1994. **41**(3): p. 189-94.
49. Gelanew, T., et al., *Molecular characterization of human isolates of Giardia duodenalis from Ethiopia.* Acta Trop, 2007. **102**(2): p. 92-9.
50. Maloney, J.G., A. Molokin, and M. Santin, *Assessment of next generation amplicon sequencing of the beta-giardin gene for the detection of Giardia duodenalis assemblages and mixed infections.* Food Waterborne Parasitol, 2020. **21**: p. e00098.
51. Nixon, J.E., et al., *Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to Giardia lamblia and Entamoeba histolytica.* Eukaryot Cell, 2002. **1**(2): p. 181-90.
52. Paget, T.A., et al., *The effects of oxygen on fermentation in Giardia lamblia.* Mol Biochem Parasitol, 1993. **57**(1): p. 65-71.
53. Atkinson, H., *Respiration in nematodes.* Nematodes as biological models, 1980. **2**: p. 101-138.
54. Allain, T., et al., *Interactions of Giardia sp. with the intestinal barrier: Epithelium, mucus, and microbiota.* Tissue Barriers, 2017. **5**(1): p. e1274354.
55. Chávez-Munguía, B., R. Cedillo-Rivera, and A. Martínez-Palomo, *The ultrastructure of the cyst wall of Giardia lamblia.* J Eukaryot Microbiol, 2004. **51**(2): p. 220-6.
56. Jarroll, E.L., A.K. Bingham, and E.A. Meyer, *Effect of chlorine on Giardia lamblia cyst viability.* Appl Environ Microbiol, 1981. **41**(2): p. 483-7.
57. Elmendorf, H.G., S.C. Dawson, and J.M. McCaffery, *The cytoskeleton of Giardia lamblia.* Int J Parasitol, 2003. **33**(1): p. 3-28.
58. Yu, L.Z., C.W. Birky, Jr., and R.D. Adam, *The two nuclei of Giardia each have complete copies of the genome and are partitioned equationally at cytokinesis.* Eukaryotic cell, 2002. **1**(2): p. 191-199.
59. Ayala, F.J., *Is sex better? Parasites say "no".* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3346-8.
60. Lasek-Nesselquist, E., et al., *Genetic exchange within and between assemblages of Giardia duodenalis.* J Eukaryot Microbiol, 2009. **56**(6): p. 504-18.
61. Ramesh, M.A., S.B. Malik, and J.M. Logsdon, Jr., *A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis.* Curr Biol, 2005. **15**(2): p. 185-91.
62. Mastronicola, D., et al., *Antioxidant defence systems in the protozoan pathogen Giardia intestinalis.* Mol Biochem Parasitol, 2016. **206**(1-2): p. 56-66.
63. Ma'ayeh, S.Y., L. Knorr, and S.G. Svard, *Transcriptional profiling of Giardia intestinalis in response to oxidative stress.* Int J Parasitol, 2015. **45**(14): p. 925-38.
64. Argüello-García, R., et al., *Drug resistance in Giardia: Mechanisms and alternative treatments for Giardiasis.* Adv Parasitol, 2020. **107**: p. 201-282.
65. Mastronicola, D., et al., *Flavohemoglobin and nitric oxide detoxification in the human protozoan parasite Giardia intestinalis.* Biochem Biophys Res Commun, 2010. **399**(4): p. 654-8.
66. Tekwani, B.L. and R.K. Mehlotra, *Molecular basis of defence against oxidative stress in Entamoeba histolytica and Giardia lamblia.* Microbes Infect, 1999. **1**(5): p. 385-94.
67. Brown, D.M., J.A. Upcroft, and P. Upcroft, *Cysteine is the major low-molecular weight thiol in Giardia duodenalis.* Molecular and Biochemical Parasitology, 1993. **61**(1): p. 155-158.
68. Leitsch, D., *A review on metronidazole: an old warhorse in antimicrobial chemotherapy.* Parasitology, 2019. **146**(9): p. 1167-1178.
69. Raj, D., et al., *Differential gene expression in Giardia lamblia under oxidative stress: significance in eukaryotic evolution.* Gene, 2014. **535**(2): p. 131-9.
70. Ansell, B.R., et al., *Drug resistance in Giardia duodenalis.* Biotechnol Adv, 2015. **33**(6 Pt 1): p. 888-901.
71. Rafferty, S., et al., *Giardia lamblia encodes a functional flavohemoglobin.* Biochem Biophys Res Commun, 2010. **399**(3): p. 347-51.

72. Gardner, P.R., et al., *Nitric oxide dioxygenase: an enzymic function for flavohemoglobin.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10378-83.
73. Gardner, P.R., *Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases.* J Inorg Biochem, 2005. **99**(1): p. 247-66.
74. Kooyman, F.N.J., J.A. Wagenaar, and A. Zomer, *Whole-genome sequencing of dog-specific assemblages C and D of Giardia duodenalis from single and pooled cysts indicates host-associated genes.* Microb Genom, 2019. **5**(12).
75. Ansell, B.R., et al., *Divergent Transcriptional Responses to Physiological and Xenobiotic Stress in Giardia duodenalis.* Antimicrob Agents Chemother, 2016. **60**(10): p. 6034-45.
76. Müller, J., et al., *Resistance formation to nitro drugs in Giardia lamblia: No common markers identified by comparative proteomics.* Int J Parasitol Drugs Drug Resist, 2019. **9**: p. 112-119.
77. Forrester, M.T. and M.W. Foster, *Protection from nitrosative stress: a central role for microbial flavohemoglobin.* Free Radic Biol Med, 2012. **52**(9): p. 1620-33.
78. Helmick, R.A., et al., *Imidazole antibiotics inhibit the nitric oxide dioxygenase function of microbial flavohemoglobin.* Antimicrob Agents Chemother, 2005. **49**(5): p. 1837-43.
79. Furuno, M., et al., *CDS annotation in full-length cDNA sequence.* Genome Res, 2003. **13**(6b): p. 1478-87.
80. Isaac-Renton, J.L., H. Shahriari, and W.R. Bowie, *Comparison of an in vitro method and an in vivo method of Giardia excystation.* Appl Environ Microbiol, 1992. **58**(5): p. 1530-3.
81. Benere, E., et al., *Infectivity of Giardia duodenalis Assemblages A and E for the gerbil and axenisation of duodenal trophozoites.* Parasitol Int, 2010. **59**(4): p. 634-7.
82. Hanevik, K., et al., *Whole genome sequencing of clinical isolates of Giardia lamblia.* Clin Microbiol Infect, 2015. **21**(2): p. 192.e1-3.
83. Morrison, H.G., et al., *Genomic minimalism in the early diverging intestinal parasite Giardia lamblia.* Science, 2007. **317**(5846): p. 1921-6.
84. Wielinga, C., et al., *Identification of polymorphic genes for use in assemblage B genotyping assays through comparative genomics of multiple assemblage B Giardia duodenalis isolates.* Mol Biochem Parasitol, 2015. **201**(1): p. 1-4.
85. Ankarklev, J., et al., *Comparative genomic analyses of freshly isolated Giardia intestinalis assemblage A isolates.* BMC Genomics, 2015. **16**: p. 697.
86. Xu, F., A. Jex, and S.G. Svard, *A chromosome-scale reference genome for Giardia intestinalis WB.* Sci Data, 2020. **7**(1): p. 38.
87. Wenger, A.M., et al., *Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome.* Nat Biotechnol, 2019. **37**(10): p. 1155-1162.
88. Rhoads, A. and K.F. Au, *PacBio sequencing and its applications.* Genomics, proteomics & bioinformatics, 2015. **13**(5): p. 278-289.
89. Wei, Z.G. and S.W. Zhang, *NPBSS: a new PacBio sequencing simulator for generating the continuous long reads with an empirical model.* BMC Bioinformatics, 2018. **19**(1): p. 177.
90. Lu, H., F. Giordano, and Z. Ning, *Oxford Nanopore MinION Sequencing and Genome Assembly.* Genomics Proteomics Bioinformatics, 2016. **14**(5): p. 265-279.
91. Pfeiffer, F., et al., *Systematic evaluation of error rates and causes in short samples in next-generation sequencing.* Sci Rep, 2018. **8**(1): p. 10950.
92. McNaughton, A.L., et al., *Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV).* Scientific Reports, 2019. **9**(1): p. 7081.
93. Troell, K., et al., *Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes.* BMC Genomics, 2016. **17**: p. 471.
94. Gao, C., M. Zhang, and L. Chen, *The Comparison of Two Single-cell Sequencing Platforms: BD Rhapsody and 10x Genomics Chromium.* Curr Genomics, 2020. **21**(8): p. 602-609.
95. Erlandsen, S.L. and E.M. Rasch, *The DNA content of trophozoites and cysts of Giardia lamblia by microdensitometric quantitation of Feulgen staining and examination by laser scanning confocal microscopy.* J Histochem Cytochem, 1994. **42**(11): p. 1413-6.

96. Tůmová, P., et al., *Unequal distribution of genes and chromosomes refers to nuclear diversification in the binucleated Giardia intestinalis.* Int J Parasitol, 2019. **49**(6): p. 463-470.
97. Bernander, R., J.E. Palm, and S.G. Svard, *Genome ploidy in different stages of the Giardia lamblia life cycle.* Cell Microbiol, 2001. **3**(1): p. 55-62.
98. Le Blancq, S.M. and R.D. Adam, *Structural basis of karyotype heterogeneity in Giardia lamblia.* Mol Biochem Parasitol, 1998. **97**(1-2): p. 199-208.
99. Tumova, P., et al., *Constitutive aneuploidy and genomic instability in the single-celled eukaryote Giardia intestinalis.* Microbiologyopen, 2016. **5**(4): p. 560-74.
100. Tůmová, P., et al., *Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of Giardia.* Chromosoma, 2007. **116**(1): p. 65-78.
101. Gamlem, A., *Genetic variation in the metronidazole metabolism genes nitroreductase-1 and 2*, in *Department of Clinical Science*. 2020, University of Bergen: Bergen. p. 78.
102. Auton, A., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
103. Kwok, P.Y. and X. Chen, *Detection of single nucleotide polymorphisms.* Curr Issues Mol Biol, 2003. **5**(2): p. 43-60.
104. Benhabiles, H., J. Jia, and F. Lejeune, *Chapter 3 - Strategies to Correct Nonsense Mutations*, in *Nonsense Mutation Correction in Human Diseases*, H. Benhabiles, J. Jia, and F. Lejeune, Editors. 2016, Academic Press: Boston. p. 107-165.
105. Choi, Y., et al., *Predicting the functional effect of amino acid substitutions and indels.* PLoS One, 2012. **7**(10): p. e46688.
106. *Scitable by Nature Education*. Definition: allele 2014 6th of August 2021]; Available from: https://www.nature.com/scitable/definition/allele-48/.
107. *Scitable by Nature Education*. haplotype/haplotypes 2014 6th of August 2014]; Available from: https://www.nature.com/scitable/definition/haplotype-142/.
108. Choy, S.H., et al., *Population expansion and gene flow in Giardia duodenalis as revealed by triosephosphate isomerase gene.* Parasit Vectors, 2015. **8**: p. 454.
109. Monis, P.T., et al., *Molecular systematics of the parasitic protozoan Giardia intestinalis.* Mol Biol Evol, 1999. **16**(9): p. 1135-44.
110. Hussein, A.I., et al., *Multiple-subgenotype infections of Giardia intestinalis detected in Palestinian clinical cases using a subcloning approach.* Parasitol Int, 2009. **58**(3): p. 258-62.
111. Siripattanapipong, S., et al., *Clonal diversity of the glutamate dehydrogenase gene in Giardia duodenalis from Thai isolates: evidence of genetic exchange or mixed infections?* BMC Microbiol, 2011. **11**: p. 206.
112. Kosuwin, R., et al., *Clonal diversity in Giardia duodenalis isolates from Thailand: evidences for intragenic recombination and purifying selection at the beta giardin locus.* Gene, 2010. **449**(1-2): p. 1-8.
113. Teodorovic, S., J.M. Braverman, and H.G. Elmendorf, *Unusually low levels of genetic variation among Giardia lamblia isolates.* Eukaryot Cell, 2007. **6**(8): p. 1421-30.
114. Mizuno, T., et al., *Extremely diversified haplotypes observed among assemblage B population of Giardia intestinalis in Kenya.* Parasitol Int, 2020. **75**: p. 102038.
115. Lecova, L., P. Tumova, and E. Nohynkova, *Clone-based haplotyping of Giardia intestinalis assemblage B human isolates.* Parasitol Res, 2019. **118**(1): p. 355-361.
116. Buret, A.G., *Mechanisms of epithelial dysfunction in giardiasis.* Gut, 2007. **56**(3): p. 316-7.
117. Ortega-Pierres, M.G., *Giardia and Giardiasis: Part A*. 2019: Elsevier Science.
118. Escobedo, A.A., et al., *Management of chronic Giardia infection.* Expert Rev Anti Infect Ther, 2014. **12**(9): p. 1143-57.
119. Einarsson, E., S. Ma'ayeh, and S.G. Svärd, *An up-date on Giardia and giardiasis.* Curr Opin Microbiol, 2016. **34**: p. 47-52.
120. Fink, M.Y. and S.M. Singer, *The Intersection of Immune Responses, Microbiota, and Pathogenesis in Giardiasis.* Trends Parasitol, 2017. **33**(11): p. 901-913.

121. Scott, K.G., L.C. Yu, and A.G. Buret, *Role of CD8+ and CD4+ T lymphocytes in jejunal mucosal injury during murine giardiasis.* Infect Immun, 2004. **72**(6): p. 3536-42.
122. Veenemans, J., et al., *Protection against diarrhea associated with Giardia intestinalis Is lost with multi-nutrient supplementation: a study in Tanzanian children.* PLoS Negl Trop Dis, 2011. **5**(6): p. e1158.
123. Bartelt, L.A. and R.B. Sartor, *Advances in understanding Giardia: determinants and mechanisms of chronic sequelae.* F1000Prime Rep, 2015. **7**: p. 62.
124. Hanevik, K., et al., *Persisting symptoms and duodenal inflammation related to Giardia duodenalis infection.* J Infect, 2007. **55**(6): p. 524-30.
125. Rogawski, E.T., et al., *Determinants and Impact of Giardia Infection in the First 2 Years of Life in the MAL-ED Birth Cohort.* J Pediatric Infect Dis Soc, 2017. **6**(2): p. 153-160.
126. Donowitz, J.R., et al., *A Prospective Longitudinal Cohort to Investigate the Effects of Early Life Giardiasis on Growth and All Cause Diarrhea.* Clin Infect Dis, 2016. **63**(6): p. 792-7.
127. Berkman, D.S., et al., *Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study.* Lancet, 2002. **359**(9306): p. 564-71.
128. Wensaas, K.A., et al., *Irritable bowel syndrome and chronic fatigue 3 years after acute giardiasis: historic cohort study.* Gut, 2012. **61**(2): p. 214-9.
129. Mørch, K., et al., *High rate of fatigue and abdominal symptoms 2 years after an outbreak of giardiasis.* Trans R Soc Trop Med Hyg, 2009. **103**(5): p. 530-2.
130. Mørch, K., et al., *Chronic fatigue syndrome 5 years after giardiasis: differential diagnoses, characteristics and natural course.* BMC Gastroenterol, 2013. **13**: p. 28.
131. Hooshyar, H., et al., *Giardia lamblia infection: review of current diagnostic strategies.* Gastroenterol Hepatol Bed Bench, 2019. **12**(1): p. 3-12.
132. Verweij, J.J., et al., *Simultaneous detection of Entamoeba histolytica, Giardia lamblia, and Cryptosporidium parvum in fecal samples by using multiplex real-time PCR.* J Clin Microbiol, 2004. **42**(3): p. 1220-3.
133. Minak, J., et al., *Evaluation of rapid antigen point-of-care tests for detection of Giardia and Cryptosporidium species in human fecal specimens.* J Clin Microbiol, 2012. **50**(1): p. 154-6.
134. Faubert, G., *Immune response to Giardia duodenalis.* Clin Microbiol Rev, 2000. **13**(1): p. 35-54, table of contents.
135. Istre, G.R., et al., *Waterborne giardiasis at a mountain resort: evidence for acquired immunity.* Am J Public Health, 1984. **74**(6): p. 602-4.
136. Isaac-Renton, J.L., et al., *A second community outbreak of waterborne giardiasis in Canada and serological investigation of patients.* Trans R Soc Trop Med Hyg, 1994. **88**(4): p. 395-9.
137. Hanevik, K., et al., *Human cellular immune response against Giardia lamblia 5 years after acute giardiasis.* J Infect Dis, 2011. **204**(11): p. 1779-86.
138. Singer, S.M. and T.E. Nash, *T-cell-dependent control of acute Giardia lamblia infections in mice.* Infect Immun, 2000. **68**(1): p. 170-5.
139. Ebert, E.C., *Giardia induces proliferation and interferon gamma production by intestinal lymphocytes.* Gut, 1999. **44**(3): p. 342-6.
140. Singer, S.M., M.Y. Fink, and V.V. Angelova, *Recent insights into innate and adaptive immune responses to Giardia.* Adv Parasitol, 2019. **106**: p. 171-208.
141. Touz, M.C., et al., *Arginine deiminase has multiple regulatory roles in the biology of Giardia lamblia.* J Cell Sci, 2008. **121**(Pt 17): p. 2930-8.
142. Solaymani-Mohammadi, S. and S.M. Singer, *Giardia duodenalis: the double-edged sword of immune responses in giardiasis.* Exp Parasitol, 2010. **126**(3): p. 292-7.
143. Halliez, M.C. and A.G. Buret, *Extra-intestinal and long term consequences of Giardia duodenalis infections.* World J Gastroenterol, 2013. **19**(47): p. 8974-85.
144. Leitsch, D., *Drug Resistance in the Microaerophilic Parasite Giardia lamblia.* Curr Trop Med Rep, 2015. **2**(3): p. 128-135.
145. Morch, K., et al., *Treatment-ladder and genetic characterisation of parasites in refractory giardiasis after an outbreak in Norway.* J Infect, 2008. **56**(4): p. 268-73.

146. Solaymani-Mohammadi, S., et al., *A meta-analysis of the effectiveness of albendazole compared with metronidazole as treatments for infections with Giardia duodenalis.* PLoS Negl Trop Dis, 2010. **4**(5): p. e682.
147. Powell, S.J., et al., *Metronidazole in amoebic dysentery and amoebic liver abscess.* Lancet, 1966. **2**(7477): p. 1329-31.
148. Hirschl, A.M., et al., *The efficacy of antimicrobial treatment in Campylobacter pylori-associated gastritis and duodenal ulcer.* Scand J Gastroenterol Suppl, 1988. **142**: p. 76-81.
149. Carter, E.R., et al., *Nitroimidazole-refractory giardiasis: a growing problem requiring rational solutions.* Clin Microbiol Infect, 2018. **24**(1): p. 37-42.
150. Murphy, P.A. and E. Jones, *Use of oral metronidazole in pregnancy. Risks, benefits, and practice guidelines.* J Nurse Midwifery, 1994. **39**(4): p. 214-20.
151. Koss, C.A., et al., *Investigation of metronidazole use during pregnancy and adverse birth outcomes.* Antimicrob Agents Chemother, 2012. **56**(9): p. 4800-5.
152. Ydsten, K.A., U. Hellgren, and H. Asgeirsson, *Quinacrine treatment of nitroimidazole-refractory giardiasis.* J Infect Dis, 2021.
153. Meltzer, E., T. Lachish, and E. Schwartz, *Treatment of giardiasis after nonresponse to nitroimidazole.* Emerg Infect Dis, 2014. **20**(10): p. 1742-4.
154. Rodríguez-García, R., L.M. Rodríguez-Guzmán, and A.H. Cruz del Castillo, *[Effectiveness and safety of mebendazole compared to nitazoxanide in the treatment of Giardia lamblia in children].* Rev Gastroenterol Mex, 1999. **64**(3): p. 122-6.
155. Ordóñez-Mena, J.M., N.D. McCarthy, and T.R. Fanshawe, *Comparative efficacy of drugs for treating giardiasis: a systematic update of the literature and network meta-analysis of randomized clinical trials.* J Antimicrob Chemother, 2018. **73**(3): p. 596-606.
156. Nabarro, L.E., et al., *Increased incidence of nitroimidazole-refractory giardiasis at the Hospital for Tropical Diseases, London: 2008-2013.* Clin Microbiol Infect, 2015. **21**(8): p. 791-6.
157. Requena-Mendez, A., et al., *The Use of Quinacrine in Nitroimidazole-resistant Giardia Duodenalis: An Old Drug for an Emerging Problem.* J Infect Dis, 2017. **215**(6): p. 946-953.
158. Lopez-Velez, R., et al., *Short course combination therapy for giardiasis after nitroimidazole failure.* Am J Trop Med Hyg, 2010. **83**(1): p. 171-3.
159. Abraham, R.J., et al., *Aminoguanidines: New leads for treatment of Giardia duodenalis infection.* Int J Parasitol Drugs Drug Resist, 2019. **10**: p. 38-44.
160. Leitsch, D., *A review on metronidazole: an old warhorse in antimicrobial chemotherapy.* Parasitology, 2017: p. 1-12.
161. Pal, D., et al., *Giardia, Entamoeba, and Trichomonas enzymes activate metronidazole (nitroreductases) and inactivate metronidazole (nitroimidazole reductases).* Antimicrob Agents Chemother, 2009. **53**(2): p. 458-64.
162. Leitsch, D., et al., *Pyruvate:ferredoxin oxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in Giardia lamblia.* J Antimicrob Chemother, 2011. **66**(8): p. 1756-65.
163. Leitsch, D., J. Muller, and N. Muller, *Evaluation of Giardia lamblia thioredoxin reductase as drug activating enzyme and as drug target.* Int J Parasitol Drugs Drug Resist, 2016. **6**(3): p. 148-153.
164. Dingsdag, S.A. and N. Hunter, *Metronidazole: an update on metabolism, structure-cytotoxicity and resistance mechanisms.* J Antimicrob Chemother, 2018. **73**(2): p. 265-279.
165. Vivancos, V., et al., *Giardiasis: Characteristics, Pathogenesis and New Insights About Treatment.* Curr Top Med Chem, 2018. **18**(15): p. 1287-1303.
166. Leitsch, D., et al., *Nitroimidazole drugs vary in their mode of action in the human parasite Giardia lamblia.* Int J Parasitol Drugs Drug Resist, 2012. **2**: p. 166-70.
167. Holmes, A.H., et al., *Understanding the mechanisms and drivers of antimicrobial resistance.* Lancet, 2016. **387**(10014): p. 176-87.

168. Müller, M. and T.E. Gorrell, *Metabolism and metronidazole uptake in Trichomonas vaginalis isolates with different metronidazole susceptibilities.* Antimicrob Agents Chemother, 1983. **24**(5): p. 667-73.
169. Szentmihályi, K., et al., *Metronidazole does not show direct antioxidant activity in in vitro global systems.* Heliyon, 2021. **7**(4): p. e06902.
170. Muller, J. and N. Muller, *Nitroreductases of bacterial origin in Giardia lamblia: Potential role in detoxification of xenobiotics.* Microbiologyopen, 2019. **8**(10): p. e904.
171. Muller, J., P. Schildknecht, and N. Muller, *Metabolism of nitro drugs metronidazole and nitazoxanide in Giardia lamblia: characterization of a novel nitroreductase (GlNR2).* J Antimicrob Chemother, 2013. **68**(8): p. 1781-9.
172. Müller, M., *Reductive activation of nitroimidazoles in anaerobic microorganisms.* Biochem Pharmacol, 1986. **35**(1): p. 37-41.
173. Muller, J., et al., *Comparative characterisation of two nitroreductases from Giardia lamblia as potential activators of nitro compounds.* Int J Parasitol Drugs Drug Resist, 2015. **5**(2): p. 37-43.
174. Nillius, D., J. Muller, and N. Muller, *Nitroreductase (GlNR1) increases susceptibility of Giardia lamblia and Escherichia coli to nitro drugs.* J Antimicrob Chemother, 2011. **66**(5): p. 1029-35.
175. Muller, J., et al., *A novel Giardia lamblia nitroreductase, GlNR1, interacts with nitazoxanide and other thiazolides.* Antimicrob Agents Chemother, 2007. **51**(6): p. 1979-86.
176. Dan, M., A.L. Wang, and C.C. Wang, *Inhibition of pyruvate-ferredoxin oxidoreductase gene expression in Giardia lamblia by a virus-mediated hammerhead ribozyme.* Mol Microbiol, 2000. **36**(2): p. 447-56.
177. Townson, S.M., J.A. Upcroft, and P. Upcroft, *Characterisation and purification of pyruvate:ferredoxin oxidoreductase from Giardia duodenalis.* Mol Biochem Parasitol, 1996. **79**(2): p. 183-93.
178. Chen, P.Y., et al., *Binding site for coenzyme A revealed in the structure of pyruvate:ferredoxin oxidoreductase from Moorella thermoacetica.* Proc Natl Acad Sci U S A, 2018. **115**(15): p. 3846-3851.
179. Watkins, R.R. and L. Eckmann, *Treatment of giardiasis: current status and future directions.* Curr Infect Dis Rep, 2014. **16**(2): p. 396.
180. Leitsch, D., et al., *Trichomonas vaginalis: metronidazole and other nitroimidazole drugs are reduced by the flavin enzyme thioredoxin reductase and disrupt the cellular redox system. Implications for nitroimidazole toxicity and resistance.* Mol Microbiol, 2009. **72**(2): p. 518-36.
181. Tejman-Yarden, N., et al., *A reprofiled drug, auranofin, is effective against metronidazole-resistant Giardia lamblia.* Antimicrob Agents Chemother, 2013. **57**(5): p. 2029-35.
182. WHO. *Antimicrobial resistance*. 2020 [cited 2021 July 27]; Available from: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.
183. Munoz Gutierrez, J., et al., *Refractory giardiasis in Spanish travellers.* Travel Med Infect Dis, 2013. **11**(2): p. 126-9.
184. Cañete, R., et al., *5-Nitroimidazole refractory giardiasis is common in Matanzas, Cuba and effectively treated by secnidazole plus high-dose mebendazole or quinacrine: a prospective observational cohort study.* Clin Microbiol Infect, 2020. **26**(8): p. 1092.e1-1092.e6.
185. Cruz, A., et al., *Isolation, excystation and axenization of Giardia lamblia isolates: in vitro susceptibility to metronidazole and albendazole.* J Antimicrob Chemother, 2003. **51**(4): p. 1017-20.
186. Ansell, B.R., et al., *Transcriptomics Indicates Active and Passive Metronidazole Resistance Mechanisms in Three Seminal Giardia Lines.* Front Microbiol, 2017. **8**: p. 398.
187. Ellis, J.E., et al., *Oxygen affinities of metronidazole-resistant and -sensitive stocks of Giardia intestinalis.* Int J Parasitol, 1993. **23**(1): p. 35-9.
188. Saghaug, C.S., et al., *Genetic Diversity of the Flavohemoprotein Gene of Giardia lamblia: Evidence for High Allelic Heterozygosity and Copy Number Variation.* Infect Drug Resist, 2020. **13**: p. 4531-4545.

189. Müller, J., A. Hemphill, and N. Müller, *Physiological aspects of nitro drug resistance in Giardia lamblia.* Int J Parasitol Drugs Drug Resist, 2018. **8**(2): p. 271-277.
190. Smith, N.C., C. Bryant, and P.F. Boreham, *Possible roles for pyruvate:ferredoxin oxidoreductase and thiol-dependent peroxidase and reductase activities in resistance to nitroheterocyclic drugs in Giardia intestinalis.* Int J Parasitol, 1988. **18**(7): p. 991-7.
191. Smith, P.D., et al., *Chronic giardiasis: studies on drug sensitivity, toxin production, and host immune response.* Gastroenterology, 1982. **83**(4): p. 797-803.
192. Goodwin, A., et al., *Metronidazole resistance in Helicobacter pylori is due to null mutations in a gene (rdxA) that encodes an oxygen-insensitive NADPH nitroreductase.* Mol Microbiol, 1998. **28**(2): p. 383-93.
193. Jenks, P.J. and D.I. Edwards, *Metronidazole resistance in Helicobacter pylori.* Int J Antimicrob Agents, 2002. **19**(1): p. 1-7.
194. Paulish-Miller, T.E., et al., *Trichomonas vaginalis metronidazole resistance is associated with single nucleotide polymorphisms in the nitroreductase genes ntr4Tv and ntr6Tv.* Antimicrob Agents Chemother, 2014. **58**(5): p. 2938-43.
195. Rice, E.W. and F.W. Schaefer, 3rd, *Improved in vitro excystation procedure for Giardia lamblia cysts.* J Clin Microbiol, 1981. **14**(6): p. 709-10.
196. Keister, D.B., *Axenic culture of Giardia lamblia in TYI-S-33 medium supplemented with bile.* Trans R Soc Trop Med Hyg, 1983. **77**(4): p. 487-8.
197. Ogren, J., et al., *Prevalence of Dientamoeba fragilis, Giardia duodenalis, Entamoeba histolytica/dispar, and Cryptosporidium spp in Da Nang, Vietnam, detected by a multiplex real-time PCR.* Apmis, 2016. **124**(6): p. 529-33.
198. Robertson, L.J., et al., *Application of Genotyping during an Extensive Outbreak of Waterborne Giardiasis in Bergen, Norway, during Autumn and Winter 2004.* Applied and Environmental Microbiology, 2006. **72**(3): p. 2212-2217.
199. Lorenz, T.C., *Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies.* J Vis Exp, 2012(63): p. e3998.
200. Monteiro, L., et al., *Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model.* J Clin Microbiol, 1997. **35**(4): p. 995-8.
201. Dey, P., *Polymerase Chain Reaction: Principle, Technique and Applications in Pathology*, in *Basic and Advanced Laboratory Techniques in Histopathology and Cytology*. 2018, Springer Singapore: Singapore. p. 201-211.
202. Lefever, S., et al., *Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays.* Clin Chem, 2013. **59**(10): p. 1470-80.
203. Pelt-verkuil, E., A. van Belkum, and J. Hays, *Principles and Technical Aspects of PCR Amplification*. 2008.
204. Yang, R., et al., *Development of a quantitative PCR (qPCR) for Giardia and analysis of the prevalence, cyst shedding and genotypes of Giardia present in sheep across four states in Australia.* Exp Parasitol, 2014. **137**: p. 46-52.
205. Verweij, J.J., et al., *Real-time PCR for the detection of Giardia lamblia.* Mol Cell Probes, 2003. **17**(5): p. 223-5.
206. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
207. Aurrecoechea, C., et al., *GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis.* Nucleic Acids Res, 2009. **37**(Database issue): p. D526-30.
208. Lee, P.Y., et al., *Agarose gel electrophoresis for the separation of DNA fragments.* J Vis Exp, 2012(62).
209. Lodish, H., A. Berk, and S. Zipursky, *Molecular Cell Biology*, in *Section 7.1, DNA Cloning with Plasmid Vectors*. 2000, W. H. Freeman: New York.
210. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.

211. Biolabs, N.E., *NEB® 5-alpha Competent E. coli (High Efficiency)*. 2020.
212. Rozas, J., et al., *DnaSP, DNA polymorphism analyses by the coalescent and other methods.* Bioinformatics, 2003. **19**(18): p. 2496-7.
213. Rozas, J., et al., *DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.* Mol Biol Evol, 2017. **34**(12): p. 3299-3302.
214. Nei, M. and W.H. Li, *Mathematical model for studying genetic variation in terms of restriction endonucleases.* Proc Natl Acad Sci U S A, 1979. **76**(10): p. 5269-73.
215. Haas, B.J., et al., *Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.* Genome Res, 2011. **21**(3): p. 494-504.
216. Huber, T., G. Faulkner, and P. Hugenholtz, *Bellerophon: a program to detect chimeric sequences in multiple sequence alignments.* Bioinformatics, 2004. **20**(14): p. 2317-9.
217. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat Protoc, 2015. **10**(6): p. 845-58.
218. Ilari, A., et al., *The X-ray structure of ferric Escherichia coli flavohemoglobin reveals an unexpected geometry of the distal heme pocket.* J Biol Chem, 2002. **277**(26): p. 23725-32.
219. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
220. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
221. Nakazato, T., T. Ohta, and H. Bono, *Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive.* PLoS One, 2013. **8**(10): p. e77910.
222. Levy, S.E. and B.E. Boone, *Next-Generation Sequencing Strategies.* Cold Spring Harb Perspect Med, 2019. **9**(7).
223. Fox, E.J., et al., *Accuracy of Next Generation Sequencing Platforms.* Next Gener Seq Appl, 2014. **1**.
224. Illumina, I., *Calling Sequencing SNPs*, Illumina, Editor. 2010, Illumina: 9885 Towne Centre Drive, San Diego, CA, USA.
225. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1135-45.
226. Saghaug, C.S., et al., *Genetic diversity in the metronidazole metabolism genes nitroreductase 1 and 2 in susceptible and refractory clinical samples of Giardia lamblia*. 2021.
227. Thompson, A.S., et al., *Accurate and Efficient One-Pot Reverse Transcription and Amplification of 2'-Fluoro-Modified Nucleic Acids by Commercial DNA Polymerases.* Biochemistry, 2020. **59**(31): p. 2833-2841.
228. Potapov, V. and J.L. Ong, *Examining Sources of Error in PCR by Single-Molecule Sequencing.* PLoS One, 2017. **12**(1): p. e0169774.
229. Pezza, J.A. and L. Sun, *Polymerase Fidelity: What is it, and what does it mean for your PCR?*, N.E. Biolabs, Editor. 2012.
230. Aguiar, J.M., et al., *Evidence of heterozygosity and recombinant alleles in single cysts of Giardia duodenalis.* Rev Bras Parasitol Vet, 2016. **25**(2): p. 187-95.
231. Boutte, J., et al., *Haplotype Detection from Next-Generation Sequencing in High-Ploidy-Level Species: 45S rDNA Gene Copies in the Hexaploid Spartina maritima.* G3 (Bethesda), 2015. **6**(1): p. 29-40.
232. Wang, G.C.Y. and Y. Wang, *The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species.* Microbiology (Reading), 1996. **142 (Pt 5)**: p. 1107-1114.
233. Lasken, R.S. and T.B. Stockwell, *Mechanism of chimera formation during the Multiple Displacement Amplification reaction.* BMC Biotechnol, 2007. **7**: p. 19.
234. Arroyo Mühr, L.S., et al., *De novo sequence assembly requires bioinformatic checking of chimeric sequences.* PLoS One, 2020. **15**(8): p. e0237455.

235. Thompson, A., *Human giardiasis: genotype-linked differences in clinical symptomatology.* Trends Parasitol, 2001. **17**(10): p. 465.
236. Emery, S.J., et al., *Differential protein expression and post-translational modifications in metronidazole-resistant Giardia duodenalis.* Gigascience, 2018. **7**(4).
237. Sisson, G., et al., *Metronidazole activation is mutagenic and causes DNA fragmentation in Helicobacter pylori and in Escherichia coli containing a cloned H. pylori RdxA(+) (Nitroreductase) gene.* J Bacteriol, 2000. **182**(18): p. 5091-6.
238. Galeh, T.M., et al., *Introducing nitazoxanide as a promising alternative treatment for symptomatic to metronidazole-resistant giardiasis in clinical isolates.* Asian Pac J Trop Med, 2016. **9**(9): p. 887-92.
239. Begaydarova, R., et al., *Determination of PFOR gene expression in strains of G. intestinalis with different inhibitory concentrations of metronidazole.* J Infect Dev Ctries, 2015. **9**(5): p. 519-23.
240. Testa, F., et al., *The superoxide reductase from the early diverging eukaryote Giardia intestinalis.* Free Radic Biol Med, 2011. **51**(8): p. 1567-74.
241. Johnson, D.C., et al., *Structure, function, and formation of biological iron-sulfur clusters.* Annu Rev Biochem, 2005. **74**: p. 247-81.
242. Honda, M., et al., *Determination of gene copy number and genotype of transgenic Arabidopsis thaliana by competitive PCR.* Journal of Experimental Botany, 2002. **53**(373): p. 1515-1520.
243. Lalle, M., et al., *Genetic heterogeneity at the beta-giardin locus among human and animal isolates of Giardiaduodenalis and identification of potentially zoonotic subgenotypes.* Int J Parasitol, 2005. **35**(2): p. 207-13.
244. Waldram, A., et al., *Prevalence of Giardia infection in households of Giardia cases and risk factors for household transmission.* BMC Infect Dis, 2017. **17**(1): p. 486.
245. Bussotti, G., et al., *Leishmania Genome Dynamics during Environmental Adaptation Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification.* mBio, 2018. **9**(6).
246. Imamura, H., et al., *Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent.* Elife, 2016. **5**.
247. Domagalska, M.A., et al., *Genomes of Leishmania parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent.* PLoS Negl Trop Dis, 2019. **13**(12): p. e0007900.
248. Achaz, G., *Frequency spectrum neutrality tests: one for all and all for one.* Genetics, 2009. **183**(1): p. 249-58.
249. Nei, M., T. Maruyama, and R. Chakraborty, *THE BOTTLENECK EFFECT AND GENETIC VARIABILITY IN POPULATIONS.* Evolution, 1975. **29**(1): p. 1-10.
250. Pandey, J., et al., *Genetic diversity and population structure of advanced clones selected over forty years by a potato breeding program in the USA.* Scientific Reports, 2021. **11**(1): p. 8344.
251. Glinka, S., et al., *Demography and natural selection have shaped genetic variation in Drosophila melanogaster: a multi-locus approach.* Genetics, 2003. **165**(3): p. 1269-78.
252. Savage, S.A., et al., *Nucleotide diversity and population differentiation of the Melanocortin 1 Receptor gene, MC1R.* BMC Genetics, 2008. **9**(1): p. 31.
253. Nash, T.E., et al., *Treatment of Patients with Refractory Giardiasis.* Clinical Infectious Diseases, 2001. **33**(1): p. 22-28.
254. Ament, M.E. and C.E. Rubin, *Relation of giardiasis to abnormal intestinal structure and function in gastrointestinal immunodeficiency syndromes.* Gastroenterology, 1972. **62**(2): p. 216-26.
255. Hermans, P.E., J.A. Diaz-Buxo, and J.D. Stobo, *Idiopathic late-onset immunoglobulin deficiency. Clinical observations in 50 patients.* Am J Med, 1976. **61**(2): p. 221-37.
256. Ward, H., et al., *Small intestinal nodular lymphoid hyperplasia in patients with giardiasis and normal serum immunoglobulins.* Gut, 1983. **24**(2): p. 120-6.

257. Faria, C.P., et al., *Associations of Giardia lamblia assemblages with HIV infections and symptomatology: HIV virus and assemblage B were they born to each other?* Acta Tropica, 2017. **172**: p. 80-85.
258. Squire, S.A. and U. Ryan, *Cryptosporidium and Giardia in Africa: current and future challenges.* Parasit Vectors, 2017. **10**(1): p. 195.
259. Yadav, P., et al., *Refractory giardiasis: a molecular appraisal from a tertiary care centre in India.* Indian J Med Microbiol, 2014. **32**(4): p. 378-82.
260. Mørch, K. and K. Hanevik, *Giardiasis treatment: an update with a focus on refractory disease.* Curr Opin Infect Dis, 2020. **33**(5): p. 355-364.
261. Neumayr, A., et al., *Efficacy and tolerability of quinacrine monotherapy and albendazole plus chloroquine combination therapy in nitroimidazole-refractory giardiasis: a TropNet study.* Clin Infect Dis, 2021.
262. Townson, S.M., et al., *Induction of metronidazole and furazolidone resistance in Giardia.* Trans R Soc Trop Med Hyg, 1992. **86**(5): p. 521-2.
263. Muller, J., et al., *Identification of differentially expressed genes in a Giardia lamblia WB C6 clone resistant to nitazoxanide and metronidazole.* J Antimicrob Chemother, 2008. **62**(1): p. 72-82.
264. Loderstädt, U. and H. Frickmann, *Antimicrobial resistance of the enteric protozoon Giardia duodenalis - A narrative review.* Eur J Microbiol Immunol (Bp), 2021. **11**(2): p. 29-43.
265. Pearce, D.A., J.A. Reynoldson, and R.C. Thompson, *A comparison of two methods for assessing drug sensitivity in Giardia duodenalis.* Appl Parasitol, 1996. **37**(2): p. 111-6.
266. Leitsch, D., *Drug susceptibility testing in microaerophilic parasites: Cysteine strongly affects the effectivities of metronidazole and auranofin, a novel and promising antimicrobial.* Int J Parasitol Drugs Drug Resist, 2017. **7**(3): p. 321-327.
267. Dyson, P., *Bacterial and Bacteriophage Genetics (4th edn).* Heredity, 2001. **86**(1): p. 117-117.
268. Mora, A., et al., *In the loop: promoter-enhancer interactions and bioinformatics.* Brief Bioinform, 2016. **17**(6): p. 980-995.
269. Elmendorf, H.G., et al., *Initiator and upstream elements in the alpha2-tubulin promoter of Giardia lamblia.* Mol Biochem Parasitol, 2001. **113**(1): p. 157-69.
270. Teodorovic, S., C.D. Walls, and H.G. Elmendorf, *Bidirectional transcription is an inherent feature of Giardia lamblia promoters and contributes to an abundance of sterile antisense transcripts throughout the genome.* Nucleic Acids Res, 2007. **35**(8): p. 2544-53.
271. Vanacova, S., et al., *Molecular biology of the amitochondriate parasites, Giardia intestinalis, Entamoeba histolytica and Trichomonas vaginalis.* Int J Parasitol, 2003. **33**(3): p. 235-55.
272. Yee, J., et al., *Transcriptional analysis of the glutamate dehydrogenase gene in the primitive eukaryote, Giardia lamblia. Identification of a primordial gene promoter.* J Biol Chem, 2000. **275**(15): p. 11432-9.
273. Jex, A.R., et al., *Recent advances in functional research in Giardia intestinalis.* Adv Parasitol, 2020. **107**: p. 97-137.
274. Poole, R.K., *Flavohaemoglobin: the pre-eminent nitric oxide-detoxifying machine of microorganisms.* F1000Res, 2020. **9**.

I

Infection and Drug Resistance

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical *Giardia lamblia* assemblage A and B isolates

This article was published in the following Dove Press journal:
Infection and Drug Resistance

Christina S Saghaug[1,2]
Christian Klotz[3]
Juha P Kallio[4]
Hans-Richard Brattbakk[1,5]
Tomasz Stokowy[1,5]
Toni Aebischer[3]
Inari Kursula[4,6]
Nina Langeland[1,2,7]
Kurt Hanevik[1,2]

[1]Department of Clinical Science, University of Bergen, Bergen, Hordaland, Norway; [2]Norwegian National Advisory Unit on Tropical Infectious Diseases, Department of Medicine, Haukeland University Hospital, Bergen, Hordaland, Norway; [3]Department of Infectious Diseases, Unit 16 Mycotic and Parasitic Agents and Mycobacteria, Robert Koch-Institute, Berlin, Germany; [4]Department of Biomedicine, University of Bergen, Bergen, Hordaland, Norway; [5]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Hordaland, Norway; [6]Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland; [7]Department of Medicine, Haraldsplass Deaconess Hospital, Bergen, Hordaland, Norway

Correspondence: Christina S Saghaug
Department of Clinical Science, University of Bergen, 8th Floor, Lab-Building, N-5021, Bergen, Norway
Tel +479 013 2414
Email christina.saghaug@uib.no

Purpose: Treatment-refractory *Giardia* cases have increased rapidly within the last decade. No markers of resistance nor a standardized susceptibility test have been established yet, but several enzymes and their pathways have been associated with metronidazole (MTZ) resistant *Giardia*. Very limited data are available regarding genetic variation in these pathways. We aimed to investigate genetic variation in metabolic pathway genes proposed to be involved in MTZ resistance in recently acquired, cultured clinical isolates.

Methods: Whole genome sequencing of 12 assemblage A2 and 8 assemblage B isolates was done, to decipher genomic variation in *Giardia*. Twenty-nine genes were identified in a literature search and investigated for their single nucleotide variants (SNVs) in the coding/non-coding regions of the genes, either as amino acid changing (non-synonymous SNVs) or non-changing SNVs (synonymous).

Results: In *Giardia* assemblage B, several genes involved in MTZ activation or oxidative stress management were found to have higher numbers of non-synonymous SNVs (thioredoxin peroxidase, nitroreductase 1, ferredoxin 2, NADH oxidase, nitroreductase 2, alcohol dehydrogenase, ferredoxin 4 and ferredoxin 1) than the average variation. For *Giardia* assemblage A2, the highest genetic variability was found in the ferredoxin 2, ferredoxin 6 and in nicotinamide adenine dinucleotide phosphate (NADPH) oxidoreductase putative genes. SNVs found in the ferredoxins and nitroreductases were analyzed further by alignment and homology modeling. SNVs close to the iron-sulfur cluster binding sites in nitroreductase-1 and 2 and ferredoxin 2 and 4 could potentially affect protein function. Flavohemoprotein seems to be a variable-copy gene, due to higher, but variable coverage compared to other genes investigated.

Conclusion: In clinical *Giardia* isolates, genetic variability is common in important genes in the MTZ metabolizing pathway and in the management of oxidative and nitrosative stress and includes high numbers of non-synonymous SNVs. Some of the identified amino acid changes could potentially affect the respective proteins important in the MTZ metabolism.

Keywords: drug metabolism, resistance, genetic analysis, metronidazole genes, ferredoxin, genetic diversity

Introduction

Giardia lamblia is a microaerophilic, eukaryotic and tetraploid parasite, annually responsible for up to ~280 million human infections worldwide.[1] *Giardia* is endemic in low- and middle-income countries (LMIC), and infection with the parasite has been shown to affect growth and cognitive functions in children in resource-poor regions.[2,3] In high-income countries, *Giardia* is usually transmitted by waterborne outbreaks or

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/IDR.S177997

found in returning travelers from LMIC.[4,5] Giardiasis is clinically associated with prolonged diarrhea and stomach pain, and it may lead to post-infectious sequela, eg, irritable bowel syndrome and chronic fatigue.[6,7]

The antibiotic metronidazole (MTZ) is a prodrug belonging to the 5-nitromidazole chemical group, and a common first-line treatment for giardiasis.[8] With its activity against microaerophilic and anaerobic pathogens, MTZ has a treatment efficacy of 73–100%.[9] MTZ enters the parasite through passive diffusion, it is activated through enzymatic reduction and releases toxic intermediates. This reaction can only occur in an anaerobic or microaerophilic environment.[10,11] Effects of activated MTZ on susceptible *Giardia* include protein malfunction, multiplication disruption, DNA damage and oxidative stress which will lead to cell death.[12–15] MTZ can be activated by four main enzymes in *Giardia*: pyruvate-flavodoxin oxidoreductases (PFOR-1 and 2), the thiol-cycling associated enzyme thioredoxin reductase (TrxR)[16] and nitroreductase (NR)-1. The iron-sulfur (Fe-S) cluster binding proteins, ferredoxins (Fd) have also been shown to participate in activation of MTZ in *Giardia*.[16] NR-2, a paralog of NR-1, is hypothesized to detoxify MTZ by fully reducing the drug without making toxic intermediates.[17–19] While *Giardia* is exposed to MTZ, it also has to constantly detoxify oxygen, nitric oxide (NO) and other harmful chemicals that may affect growth and viability.[20]

The majority of proposed resistance mechanisms consist of either down-regulation of enzymes that render MTZ cytotoxic (NR-1, PFORs and TrxR), or up-regulation of enzymes that can make MTZ an inert molecule (NR-2).[19] Most work regarding drug resistance mechanisms in *Giardia* has been done using laboratory induced resistant assemblage A1 isolates, which rarely infects humans, compared to infections with *Giardia* assemblage A2 and B, which are common. MTZ is probably activated through several pathways and exhibit a pleiotropic mode of action, and resistance to MTZ has generally taken longer time to develop compared to resistance against most other antibiotics.[21]

MTZ resistance in *Giardia* is a growing problem. Clinical experience and recent studies of returning travelers have shown that MTZ-refractory *Giardia* cases have a prevalence of ~20%.[22,23] The prevalence has dramatically increased over the last few years, reaching up to 40% in a referral outpatient clinic in London.[24] The highest prevalence of MTZ refractory cases have been found in returning travelers from Asia, especially India, or the south and east Mediterranean areas.[22,24–26]

No standardized susceptibility assays for *Giardia* isolates are available,[27] and susceptibility testing in the laboratory is challenging due to genomic complexity of the parasite. Moreover, isolates are difficult to culture, and growth rates may differ between isolates and assemblages. The microaerophilic environment in which *Giardia* thrives is difficult to mimic in vitro for drug susceptibility assays. Experiments may be performed in unfavorable high levels of O_2, affecting growth rates and MTZ metabolism of *Giardia*. Thus, a reliable evaluation of MTZ's activation and tolerance can be difficult to obtain.[28] Studies of MTZ metabolism pathway genes in *Giardia* are mostly based on gene expression analysis in isogenic laboratory-derived resistant *Giardia* of the assemblage AI strains, rarely found in humans.[13,18,29–34] So far, no specific markers for resistance in *Giardia* have been discovered, and resistance is likely complex and based on several factors. These factors could include amino acid variations in proteins and altered function, post- or pre-translational variants, changes in expression of activating or protective proteins, modifications of regulatory pathways and epigenetic modifications,[16,35–39] and a recent study of resistant *Giardia* strains showed that MTZ is capable of inducing significant changes in proteins found in the antioxidant, electron transport, and pyruvate catabolism networks. In addition, a correlation between acetylation of these proteins and MTZ resistance was found.[35]

Begaydarova et al 2015 investigated clinical isolates and linked PFOR gene expression to MTZ resistance, but PFOR was not found to be a good resistance marker.[40] Another study by Galeh et al 2016[41] linked non-synonymous single nucleotide variants (nsSNVs) in PFOR and NR genes to MTZ resistance and possibly through reduced translation and thereby reduced MTZ activation. Genetic variation in *Giardia* genome is not well known, except for a few genotyping genes, ie, *triose phosphate isomerase* (*tpi), glutamate dehydrogenase* (*gdh*) and *beta-giardin* (*bg*).[42] The genetic variation among *Giardia* isolates has in one study been found to be low for some assemblage A genes, ie, Fd3 (DHA2_154390) with more SNPs in assemblage B.[43] However, the background genetic variability of the parasites' MTZ metabolizing pathways and management of the oxidative and nitrosative stress has not been investigated in detail and constitutes an important basis for further research into potential markers of MTZ resistance.

Laboratory-induced MTZ resistance may be lost during one en/excystation cycle,[44] indicating that metabolic

adaptations are likely to contribute to resistance. Still, the recent emergence of rapidly increasing clinical MTZ resistance raises the suspicion that strains with a genetically inheritable trait for resistance are now circulating. Indeed, one laboratory-induced resistance line has been shown to possess a non-sense mutation in one of its NR-1 alleles.[16] However, this challenge may best be studied by examining current clinical isolates rather than laboratory strains obtained decades ago. The prospect for whole genome sequencing of purified cysts from clinical stool samples makes it possible to perform such studies.[45] The aim of our study was to investigate the general genetic variation in proteins involved in metabolic pathways associated with MTZ resistance or management of oxidative stress substances.

Material and methods

Genes of interest

A PubMed literature search was performed using the word *Giardia* in combinations with one or two of the words: MTZ, refractory, resistance and oxidative stress, to identify genes in the MTZ metabolism pathway and oxidative and nitrosative stress management in *Giardia*. The search aimed to identify genes that were upregulated or downregulated during exposure to MTZ and/or free radicals or associated with the activation or inactivation of MTZ and genes with antioxidant properties. These publications and their references were used to establish a list of 29 candidate genes (see Table 1). In order to identify pairwise shared genes and the correct gene ID for *Giardia* assemblage A2 and B, supplementary data from Adam et al was used.[46] Finally, the *Giardia* DB (http://giardiadb.org) was used to find gene orthologs and paralogs. For the NR genes analyzed in this study, the annotations NR-1 and NR-2 have been used for clarity. NR-2 is also known as GlNR1/Fd-NR2 (gene ID: Gl50803_22677), and NR-1 as GlNR2/Fd-NR1 (gene ID: Gl50803_6175). For the two genes PFOR-1 and 2, the literature annotations were contradicting, and no more specifications could be found in the *Giardia* DB. The annotations used for the PFORs in this study are according to GenBank. The annotations of Fds used in our study is based on Müller et al 2008 and Ansell et al 2017.[16,47] One hypothetical Fd presented by Nixon et al 2002[48] and Ansell et al 2017,[16] Fd5, could not be found in the reference genomes of A2 and B, nor by GenBank blast. Another hypothetical Fd (DHA2_153401 and GSB_151614) was named Fd6 in tables and figures. Some of the putative Fds presented in this study may actually be domains in a larger protein but are presented as single Fds until more is known.

Giardia isolates

Twelve isolates of *Giardia* assemblage A2 and eight isolates of assemblage B were cultured from human clinical stool samples between 2011 and 2015 at the Robert Koch Institute, Berlin, Germany. The stool samples were collected by the Institute of Tropical Medicine and International Health, Charité, Berlin, Germany. One of the isolates, P324, was obtained from the Tropenmedizinische Ambulanz, University hospital Düsseldorf, Germany. No clinical data for the samples have been collected. *Giardia* cysts from the stool samples were purified using a two-step sucrose gradient centrifugation. The cysts were excysted using a standard two-step excystation method,[49] in order to acquire trophozoites. The trophozoites were grown in Diamond's TY-S-33 medium supplemented with bile at 37°C according to.[50] Collection of the trophozoites was done from confluent cultures in 11 mL tubes, and then the DNA was extracted using Maxwell® 16 FFPE Plus LEV DNA Purification Kit (Promega BioTech AB, Sweden, cat AS1135) together with Maxwell® 16 instrument (AS1000). The quality and amount of DNA was validated with NanoDrop and the concentration of the DNA was verified with Qubit 2.0 (Life Technologies, Grand Island, NY).

Genome sequencing

Fragmented *Giardia* DNA (2 µg, 550 bp insert size) from the trophozoites was first prepared and indexed using Illumina TruSeq DNA PCR-Free Sample Prep Kit (Cat. Nr 20015962 and 20015960S). Library size distribution was validated using the Agilent Technologies 2100 BioAnalyzer, High Sensitivity DNA chip, and quantification was done using the KAPA Library Quantification Kit for Illumina sequencing platforms. Normalized whole genome *Giardia* DNA-libraries were sequenced using the Illumina MiSeq paired-end technology (2×300 bp). FASTQ-files from the sequencing were assembled and aligned to the corresponding reference genomes of *Giardia* assemblage A2 and B (v.26) using Bowtie 2-2.2.3.[51] The reference genomes were obtained from *Giardia* DB in January 2016 versions 2013-11-25.[52]

Table 1 Genes included in the SNV analysis for metronidazole metabolism, oxidative- and nitrosative-stress management. The ferredoxins have been named according to Ansell et al 2017,[16] and another Fd, DHA2_153401/GSB_151614, earlier presented by Nixon et al[48], has been given the name Fd6 in this table. The other genes of interest have been named according to annotations from *Giardia* DB

Included candidate genes	Gene ID		Protein function.	References
	Assemblage A2	Assemblage B		
Nitroreductase family protein fused to ferredoxin domain Fd-NR1	DHA2_153380	GSB_153178	MTZ activation.	17,18,37,48
Nitroreductase Fd-NR-2 (NR-2)	DHA2_22677	GSB_22677	Detoxification of MTZ.	34,48
Nitroreductase family protein (NTR-1) -Ferredoxin (Fd1) -Ferredoxin, 4Fe-4S (Fd2) -2Fe-2S ferredoxin (Fd3) -Putative oxidoreductase 4Fe-4S (Fd4) -4Fe-4S binding domain family protein (Fd6)	DHA2_15307 DHA2_9662 DHA2_10329 DHA2_154390 DHA2_151386 DHA2_153401	GSB_15307 GSB_9662 GSB_10329 GSB_150173 GSB_153527 GSB_151614	Upreg. during MTZ exposure. No ferredoxin domain. Suspected co-factors for PFOR-1 and 2 for activation of MTZ. Fd3 is a hypothetical protein related to the ferredoxins.	15,48,77 16,32,37,43,48,54,78
A-type flavoprotein lateral transfer candidate	DHA2_10358	GSB_10358	O_2-scavening enzyme in redox system.	20,28,79
Thioredoxin reductase (TrxR)	DHA2_9827	GSB_9827	Reduce flavins and activate MTZ.	13,20,32,36,37,39,80,81
Thioredoxin-like protein (Trx) -Thioredoxin peroxidase (prx 1) -Thioredoxin peroxidase (prx 1) -Peroxiredoxin 1 -Thioredoxin peroxidase	DHA2_9355 DHA2_14521 DHA2_15383 DHA2_153915 DHA2_152385	GSB_9355 GSB_14521 GSB_15383 GSB_153801 GSB_151294	Endogenous thioredoxin for TrxR. Antioxidant system; proposed to detoxify peroxynitrite to prevent hydroxyl radical and detoxification of H_2O_2 to H_2O.	20,32,37,47,81 12,20,37,76
-Pyruvate-flavodoxin oxidoreductase (PFOR-1) -Pyruvate-flavodoxin oxidoreductase (PFOR-2)	DHA2_114609 DHA2_17063	GSB_114609 GSB_17063	Decarboxylate pyruvate to acetyl-CoA and send excess electrons to ferredoxin via iron-sulfur clusters. Can also activate MTZ by partial reduction. O_2 sensitive.	15,16,17,30,32,33,37,47,81
Histone H2A	DHA2_152990	GSB_151412	Exposure to MTZ causes phosphorylation of histone H2A. Upregulated in resistant *Giardia*.	13,47,82
Flavohemoprotein	DHA2_154000	GSB_151570	Detoxification of NO through conversion to nitrate (O_2-dependent).	16,20,36,37,73–76
Desulfoferredoxin (SOR)	DHA2_152891	GSB_153135	Iron-dependent enzyme. In presence of electron donor, it degrades superoxide anion to H_2O_2.	20,36,76
-Alcohol dehydrogenase -Alcohol dehydrogenase E	DHA2_13350 DHA2_93358	GSB_13350 GSB_93358	Both an alcohol- and acetaldehyde dehydrogenase. Down-reg. during H_2O_2 and MTZ exposure and can regenerate NAD under anaerobic conditions.	12,15,36,37,39,48,81,83

(*Continued*)

Table 1 (Continued).

Included candidate genes	Gene ID		Protein function.	References
	Assemblage A2	Assemblage B		
CoA-disulfide reductase NAD(P)H	DHA2_33769	GSB_33769	Reduces formation of ROS/protects O_2-labile proteins.	15,20,37,84
NADH oxidase	DHA2_9719	GSB_9719	Water-forming activity from H_2O_2. Major contributor in the electron pathway. It protects O_2 sensitive proteins.	15,16,20,37,39,48,81,84
NADPH oxidoreductase putative	DHA2_17151	GSB_17151	Upreg. during MTZ exposure and H_2O_2 exposure, but down-reg. under MTZ induced stress.	15,17,47
NADPH oxidoreductase putative	DHA2_17150	GSB_17150	Upreg. during MTZ/H_2O_2 exposure, but down-reg. under MTZ induced stress. May reduce O_2 to superoxide and H_2O_2 and contributor with ROS.	16,17,47,85
Acetyl-CoA synthetase	DHA2_13608	GSB_13608	Acetyl-CoA is used as substrate to make ATP in a microaerophilic environment.	37,48,83
Malate dehydrogenase	DHA2_3331	GSB_3331	Upregulated during H_2O_2 and MTZ exposure.	15
NADP-specific glutamate dehydrogenase (GDH)	DHA2_21942	GSB_21942	Antioxidant enzyme in the NADPH metabolism. Down-reg. in resistant lines.	34,36,86

Variant analysis

Aligned data were analyzed using Geneious v.10.2.4. Variant calling was done to identify SNVs using the parameters minimum variant frequency 0.1 and a minimum coverage of 10 nucleotides. The coding regions (CDS) of 29 candidate genes (including pre, 150 bp, - and post-coding regions, 50 bp) were extracted from *Giardia* assemblage A2 and B in Geneious. For genes with introns (Fd3 in both assemblages), genetic analysis was done in the CDS and SNVs in introns are therefore not presented. To obtain all the SNVs in one alignment, the 29 genes from each isolate were concatenated, and then compared against each other assemblage-wise. Comparisons of the 29 genes were also performed in Geneious using a global alignment with free end gaps. Alignments of ferredoxin domains are based on findings from Nixon et al 2002,[48] and for more detailed information see the Supplementary material. For each gene, the total number of possible SNV positions has been used. The SNVs were divided into nsSNVs or synonymous SNVs (sSNVs) in the coding regions or as SNVs in the non-coding regions. Ratios for sSNV and nsSNVs were calculated for each gene.

Homology modeling of the nitroreductases and ferredoxins

Homology modeling of the NRs and ferredoxins 1–6 were done based on the amino acid sequences used in the gene analysis. Models were created using Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) server with intensive mode.[53] Models were visualized and superimposed in PyMol (The PyMOL Molecular Graphics System, Version 2.1.0 Schrödinger, LLC).

Storage of data

All genetic data is kept on a secure server; Norwegian e-infrastructure for Life Sciences.

Results

Whole genome sequencing of 20 cultured *Giardia* assemblage A2 and B isolates was successfully carried out with an average coverage of 31.1 (110.3) (see Table S1) for the 29 genes of interest. Flavohemoprotein (DHA2_154000 and GSB_151570), was found to have a substantially higher coverage (87.6 (54.2)), where a total of 17 of the 20 *Giardia* assemblage A2 and B isolates had higher than average coverage.

Based on the criteria in the literature search, 29 candidate genes were selected and are presented in Table 1.

For the 29 candidate genes, the genetic variation was analyzed across the 12 assemblage A2 and eight assemblage B clinical isolates. The average number of nsSNVs per gene length was 0.3 % for assemblage A2 and 1.2% for assemblage B. Averages for SNVs for each gene are presented in Table S2. One of the genes in this study, Fd3, was found to have a 35 bp intron in assemblage A2 and a 36 bp intron in assemblage B, which supports earlier findings.[43,54]

Genes involved in the MTZ and oxidative stress and nitrosative stress management

Considerably more nsSNVs could be found in the *Giardia* assemblage B isolates than in the assemblage A2 isolates. A color-graded heatmap of the nsSNVs causing amino acid changes in the 29 proteins involved in MTZ metabolism, oxidative and nitrosative stress management is therefore only shown for *Giardia* assemblage B (Figure 1). The genetic variation in the flavohemoprotein (GSB_151570) is not illustrated in Figure 1 due to high frequency of SNVs and variably higher coverage than the average.

For two of the genes with highest frequency of genetic variability in assemblage B (GSB_151294 - thioredoxin peroxidase and GSB_153178 – NR-1), the nsSNVs constituted over 4% of the nucleotides in the gene, and only two genes were found to be without any nsSNVs (histone H2A; GSB_151412 and Fd6; GSB_151614). The average percentage of any SNV per CDS length was 5.4% in *Giardia* assemblage B and 1.1% in assemblage A2.

The majority of the genes was shown to have at least one sSNV, except for the four genes Fd3; DHA2_154390, NR-1; DHA2_153380, Fd1; DHA2_9662 and NADPH oxidoreductase putative; DHA2_17151 in assemblage A2. sSNVs were present in all genes in assemblage B. SNVs in the pre-CDS regions were shown to be higher in assemblage B (average of 10.1 SNVs in 150 bp) than in assemblage A2, where the average was found to be 1.1 SNVs per 150 bps. Post-CDS regions had generally lower variability (0.3 in A2 and 4.2 in B for 50 bp).

Nitroreductases

Three different NRs were investigated for the presence of SNVs in this study, the NR-1, NR-2 and the oxygen-insensitive NTR-1. NR-1, had one of the highest numbers of nsSNVs (4.1%) for *Giardia* assemblage B, with more nsSNVs than sSNVs in the CDS. However, there were lower than average numbers of SNVs in the non-coding regions of the NR-1 in assemblage B. The NR-1 gene in assemblage A2 was found to be fully conserved without any SNVs in the coding or non-coding regions. The NR-2 gene sequence also had high numbers of nsSNVs in *Giardia* assemblage B (2.1%), and SNVs in the pre-CDS regions were higher than the average (21 vs 10.1 out of 150 bp). The NR-2 in *Giardia* assemblage A2 had low numbers of SNVs, where only one was a nsSNV. The NTR-1 gene was fully conserved in assemblage A2, and had lower than average numbers on non-syn SNVs in B.

Ferredoxins

The Fds are proposed to be co-factors for the MTZ activating proteins, PFOR-1 and 2, and changes in these co-factors could potentially affect their function, binding and role. We found that three of the ferredoxins in the *Giardia* assemblage B isolates, Fd2; GSB_10329, Fd4; GSB_153527 and Fd1; GSB_9662, had higher numbers of nsSNVs than the average of 1.2% (respectively 3.1%, 1.7% and 1.7%). Interestingly the Fd2 was one of the proteins with the most SNVs in the post- and pre-CDS regions of all the genes investigated in assemblage B, and it also had a high number of nsSNVs in the coding region.

For *Giardia* assemblage A2 two of the ferredoxins had higher numbers of nsSNVs than the average of 0.3%, respectively Fd2; DHA2_10329 (2.1%) and Fd6; DHA2_153401 (1.1%). One fascinating finding was that the Fd6 was fully conserved in *Giardia* assemblage B while assemblage A2 had some SNVs present.

Antioxidant and metabolic proteins

Two of the free radical protective enzymes, prx (GSB_151294) and NADH oxidase (GSB_9719) and the metabolic protein alcohol dehydrogenase (GSB_13350) was demonstrated to have higher numbers of nsSNVs in assemblage B than the average. Post-CDS analysis revealed that TrxR; GSB_9827 was one of the proteins with the highest number of SNVs, while NADPH oxidoreductase putative; GSB_17150, the two prxs, GSB_153801 and GSB_14521, malate dehydrogenase; GSB_3331, flavoprotein; GSB_10358 and alcohol dehydrogenase; GSB_13350, all had higher numbers of post-CDS SNVs than the average.

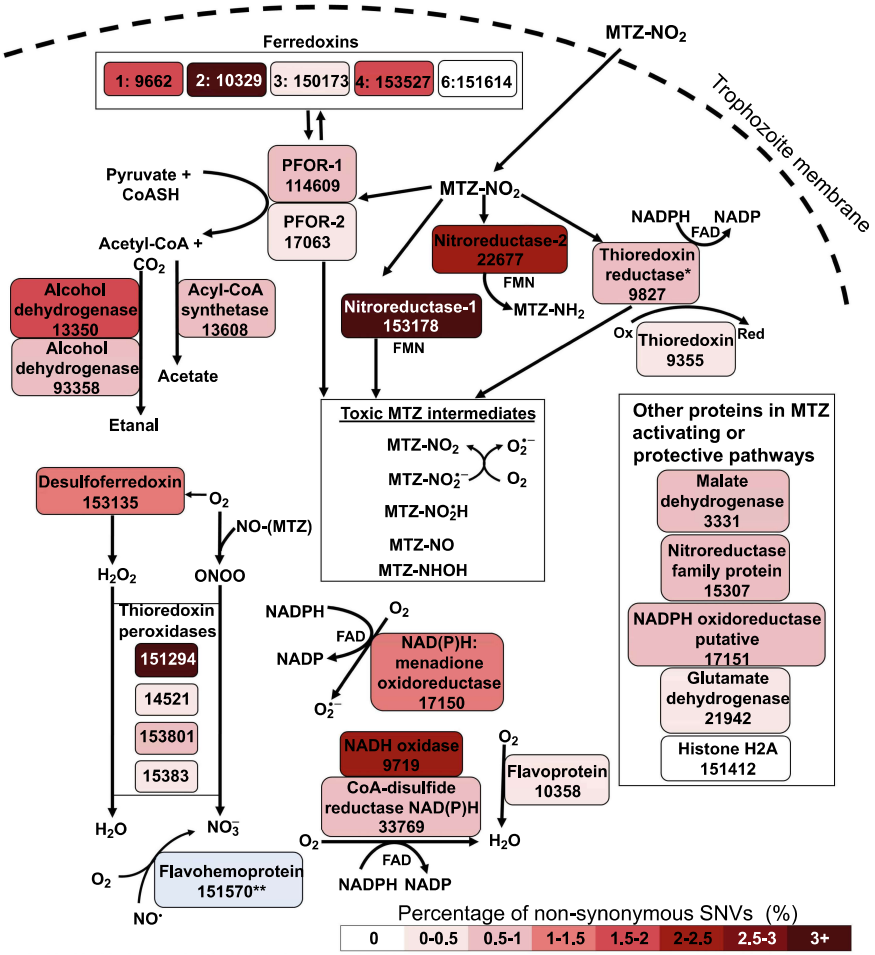


Figure 1 Heat map of proteins in the metabolism of metronidazole and oxidative/nitrosative stress management in *Giardia* assemblage B. Metronidazole (MTZ) passively diffuses through the trophozoite membrane and needs to be activated into toxic intermediates to execute its function as an antibiotic. Several theoretical intermediates exist for MTZ. MTZ can be activated either through the enzymes pyruvate: ferredoxin oxidoreductase (PFOR)-1 and/or 2 with ferredoxin (fd) as a co-factor, by nitroreductase-1 with flavin mononucleotide (FMN) as a co-factor or by thioredoxin reductase with the redox cofactor flavin adenine dinucleotide (FAD). MTZ may also be converted into an inert metabolite through the enzyme nitroreductase-2. The MTZ-NO formed during activation may react with $O_2^{\bullet-}$ and create the reactive free radical molecule peroxynitrite ($ONOO^-$). In order to remove toxic free radicals many enzymes exhibit protective functions, ie, thioredoxin peroxidase enzymes may convert the peroxynitrite to the harmless molecule nitrate (NO_3^-). The free radical form of $NO^{\bullet}$ may be converted to nitrate by flavohemoprotein (O_2 dependent reaction). In the microaerophilic environment, *Giardia* is repeatedly exposed to the harmful $O_2/O_2^{\bullet-}$. The O_2 may be metabolized to H_2O_2 through desulfoferredoxin (SOR) and further H_2O_2 is converted to H_2O by the thioredoxin peroxidase enzymes. O_2 may also be converted to H_2O by NADH oxidase enzymes or the O_2 scavenging enzyme flavoprotein. Other enzymes may cause free radicals in *Giardia*, ie, NADPH oxidoreductase putative is areactive oxygen species (ROS) contributor due to conversion of O_2 to a more reactive free radical form ($O_2^{\bullet-}$). Another contributor to ROS is the reduced version of thioredoxin and it may initiate protein misfolding. The color gradation of the proteins represents the number of nsSNVs positions per length of each gene. *Thioredoxin reductase is active in both MTZ metabolism and reduction of the protein thioredoxin. **Flavohemoprotein is a suspected multicopy gene and the number of SNVs has not been defined and is not represented by color in the heatmap.

Alignment of the ferredoxins and the nitroreductases

Some of the ferredoxins and NRs were found to have high numbers of nsSNVs in many of the *Giardia* assemblage A and B isolates. The amino acid changing SNVs and their positions in the proteins were therefore analyzed further, and an alignment was made to compare the cysteine domains in the five different ferredoxin proteins included in this analysis in addition to NR-1 and NR-2 and a hypothetical Fd

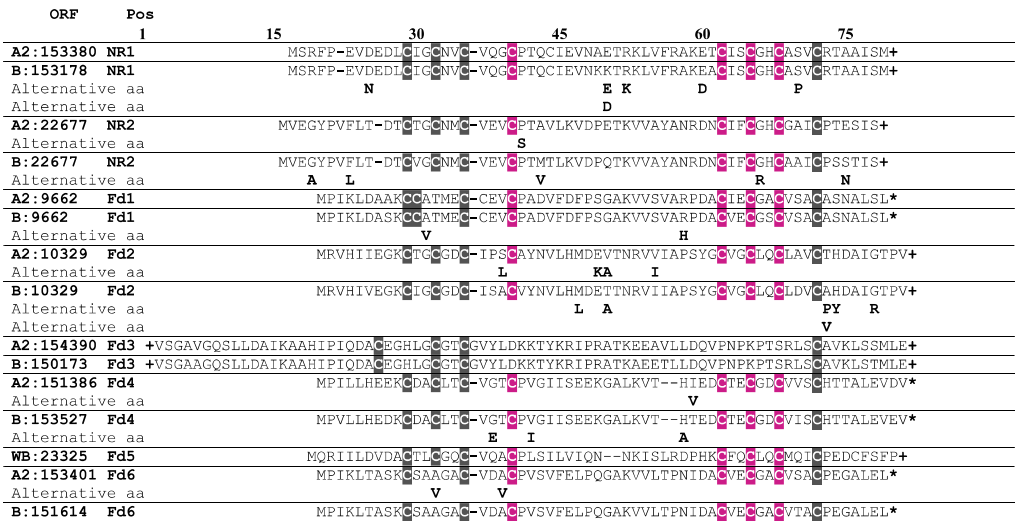


Figure 2 Alignment of the ferredoxins (Fd1-Fd6) and nitroreductase 1 and 2. The genetic variability is represented by alternative amino acids under the original protein sequences. Ferredoxin 5 was found to be an uncharacterized protein and is not annotated in the DHA2 and GSB genomes and is thus presented by the hypothetical sequence from *Giardia* WB. Only the ferredoxin domains in the N-terminus of the proteins are represented, and the nitro-flavin mononucleotide (FMN) reductase domain in the C-terminus of the NRs are not shown. The cysteine residues responsible for forming the iron-sulfur clusters are shaded in either black or magenta to represent how the residues connect. *end of the protein, +the protein has been cut and only the ferredoxin domain is shown.

previously described in WB (Fd5)[48] The proteins NR1, Fd1, and Fd3 in assemblage A2 and Fd3 and Fd6 in assemblage B had no nsSNVs in their ferredoxin domains. The nsSNVs positioned around the cysteine residues in the ferredoxin domains responsible for forming the Fe-S clusters binding sites (electron transport sites) are shown in Figure 2.

Homology modeling of the ferredoxin domains

In order to be able to analyze the effect of the amino acid changes, homology modeling was used. The ferredoxin domain sequence alignment presented in Figure 2 and the putative cysteine residues for binding Fe-S clusters are coded with the same color codes as shown in the homology models of the domains in Figure 3. Homology modeling indicates various differences within the ferredoxin domains.

Fd1, Fd2 and Fd4 can be superimposed on ferredoxin domains of NR-1 and 2. All of them clearly have two cysteine motifs, with four cysteines each, able to bind Fe-S clusters, probably of the type 4Fe-4S.[55,56] The rest of the ferredoxins differ from the others by shape and ability to bind Fe-S clusters. The Fd6 has two cysteine motifs but one of the motifs only has three coordinating cysteines instead of four, which could indicate that the Fd6 only can bind one 4Fe-4S cluster or potentially one 4Fe-4S and an additional 3Fe-4S cluster (bound to 3-cystein motif). The Fd3 has only one clear motif for binding 4Fe-4S cluster and in addition it has a proposed elongated C-terminus. Fd5 has both an elongated N- and a C-terminus.

Some of the amino acid substitutions were positioned close to the ferredoxin domain with Fe-S cluster binding cysteines and sufficiently different in charge or hydrophobicity to be able to potentially cause a change significant enough to affect protein function. Some of the amino acids in the ferredoxin domains are surface exposed and changes here could potentially affect the proteins' ability to interact with other proteins in possible intermolecular electron transfers. These changes include changes from a positively charged amino acid such as K to a negatively charged E or D seen in NR-1 assemblage B, or a change from E to K in Fd2 assemblage A2 and in Fd2 from assemblage B where G is changed to R.

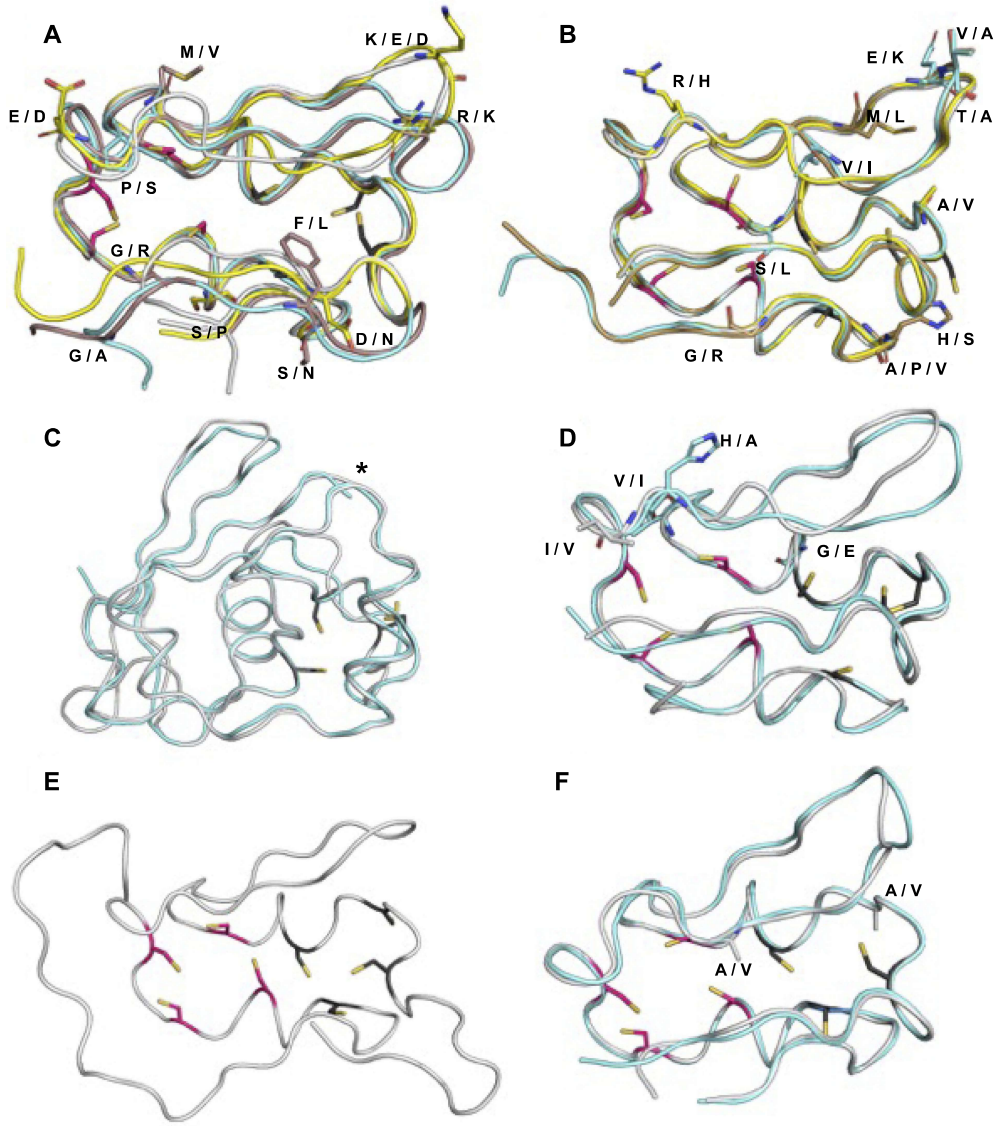


Figure 3 Cartoon tube representation of homology models of Nitroreductase (NR) ferredoxin domains and ferredoxins (Fd) 1–6. Cysteine residues responsible for the binding of [Fe4-S4] iron–sulfur cluster are represented with sticks colored with dark gray and red for clusters 1 and 2, respectively. For the mutations caused by SNVs, the original amino acid residue is shown in stick and possible mutations are marked with one letter code. (**A**) NR1_A2 (grey), NR2_A2 (cyan), NR1_B (yellow) and NR2_B (brown) (**B**) Fd1_A2 (grey), Fd2_A2 (cyan), Fd1_B (yellow) and Fd2_B (brown) (**C**) Fd3_A2 (grey), Fd3_B (cyan). The extended C-terminus has been truncated from the figure and is marked with an asterisk. (**D**) Fd4_A2 (grey), Fd4_B (cyan). (**E**) Fd5_WB (grey) (**F**) Fd6_A2 (grey), Fd6_B (cyan).

In NR-1 from assemblage B, S can be changed to P (pos. 70 Figure 2), and in NR-2 from A2, P can change to S (pos. 41 in Figure 2) which might change the conformation of the protein backbone, as P is structurally more rigid than other amino acids. These SNVs are located next to or between two cysteines coordinating Fe-S cluster binding in NR-1 and NR-2, respectively. SNVs that could potentially have an effect to electrostatic properties of Fe-S clusters are G to R (pos 66)

change in NR-2 from assemblage B and G to E (pos 38) change in Fd4 from assemblage B.

Discussion

The genome of *Giardia* trophozoites is tetraploid, consisting of two diploid nuclei. The variability among and within the assemblages can therefore potentially be high. In our study we show that the nsSNVs are common in the investigated genes in culturable clinical isolates of *Giardia* assemblage A2 and B. Generally, *Giardia* assemblage B had higher numbers of SNVs in the 29 genes investigated, and relates well to the earlier finding that the allelic sequence heterozygosity (ASH) in assemblage B is up to 0.53%,[57] whereas the ASH of two A2 isolates, AS-98 and AS-175, has been determined to be 0.25–0.35%.[46] In *Giardia* assemblage B, particularly high numbers of nsSNVs were found in gene sequences encoding MTZ activating enzyme NR-1, the PFOR co-factors Fd2, Fd1 and Fd4, the protective enzymes prx and NADH oxidase, the metabolic enzyme alcohol dehydrogenase, in addition to the MTZ inactivating enzyme NR-2. In assemblage A2, gene sequences of Fd2, Fd6 and the ROS-contributing enzyme NADPH oxidoreductase putative had the greatest number of nsSNVs.

Homology modeling shows that some of the amino acid changing SNVs are in close proximity to the cysteine residues responsible for forming Fe-S clusters and could potentially affect binding affinities of the proteins NR-1, NR-2, Fd2 and Fd4, or how they may exert their roles as electron transferring proteins.

Genetic variation as potential resistance markers

High numbers of nsSNVs found in the *Giardia* assemblage B isolates may indicate that assemblage B mutates faster than assemblage A2 isolates. The higher variation frequencies found in the B isolates may be associated with greater functional variation, higher prevalence of disease in humans and higher rates of treatment-refractory cases.[58,59]

A recent study,[41] using clinical *Giardia* isolates, suggested that nsSNVs in NR-2 and PFOR genes could be linked to resistant *Giardia* isolates, but the study was based on only two treatment-refractory isolates. In addition, this study gives potential evidence that different haplotypes of these two genes may exist within the genome. This finding correlates with our observation that the numbers of nsSNVs in the NR-2 protein in *Giardia* assemblage B was higher than average. The PFORs in our study were found to have lower numbers of nsSNVs than average. However, the co-factors for the PFORs, Fds, were found to be highly variable within three of the five Fds in assemblage B and two of the Fds in *Giardia* assemblage A2.

The role of the MTZ activating enzyme, PFOR, is thought to be the transfer of electrons, through its bound Fe-S clusters, to soluble Fds.[38] Lower levels of PFOR and ferredoxin enzyme activities have been characterized in resistant *Giardia* cultures and were also connected to MTZ sensitivity in clinical isolates.[16,32,60,61] PFOR gene expression was, however, not found to be of significant difference in patient isolates with different susceptibility to MTZ in vitro.[40]

In other anaerobe organisms, genetic markers of resistance have been identified. The enzyme known as oxygen-insensitive nitroreductase, has been shown to be an MTZ activating enzyme in *Helicobacter pylori* and *Escherichia coli*, and bacteria with inactivated NTR or nsSNVs in this gene, may be resistant.[62–64] Additionally, resistance in the parasite *Trichomonas vaginalis* has been associated to SNVs in the two nitroreductase genes ntr4Tv and ntr6Tv.[65] SNVs found in intergenic regions of 13 genes in *T. vaginalis* have been linked to resistance as well, and the study suggests that a panel of genetic markers could be used as a diagnostic tool of resistance.[66] Intergenic analysis should be done in *Giardia* as well, in order to predict whether SNVs here may play a role in MTZ resistance.

An enzyme in *Giardia* known as oxygen-insensitive nitroreductase (NTR-1), is similar to the NTR in bacteria, and it might have been acquired through lateral transfer from bacteria.[48] The NTR-1 is different from the NR-1 and NR-2 due to missing ferredoxin domains in the protein and may not play a key role in the Fe-S cluster electron transportation. We found the presence of nsSNVs in NTR-1 to be low in our analysis. Whether this enzyme has a role in MTZ activation in *Giardia* is not known, but it may have a fundamental role in protection against oxidative- and nitrosative stress, as it is O_2 insensitive, and has been shown to be upregulated during MTZ exposure.[15]

Giardia could also potentially down-regulate enzymes that detoxify oxygen to protect itself from the activated form of MTZ. This down-regulation is favorable due to the fact that MTZ may only be reduced in microaerophilic or anaerobic environments. Presence of oxygen helps to deactivate MTZ by futile cycling, creating ROS.[60] The ROS will then require extensive detoxification by enzymes in the antioxidant system to avoid harmful cell damage.[67]

In *Giardia* assemblage B, high variability was found in the two oxygen detoxifying proteins NADH oxidase (GSB_9719) while the ROS contributing enzyme NADPH oxidoreductase putative (GSB_17150) was found to have higher variability than the average for both assemblages. Oxidoreductases have previously been shown to be upregulated when *Giardia* is challenged with H_2O_2.[37] The functions and characteristics of the NADPH oxidoreductases and their potential paralogs should be investigated in future studies, to understand whether they have a role MTZ resistance, as oxidoreductases with null mutations, causing non-functional proteins, have been associated with MTZ resistance in *H. pylori*.[68]

Some of the genes investigated in the present study were found to have few, or non-existent, nsSNVs (Fd6 and Histone H2A in assemblage B, in addition to 10 genes in assemblage A2, see Table S2). Changes in these conserved genes could potentially have more serious consequences and affect biological functions such as survival in culture, clinical infectivity and tolerance to free radicals, and could be something to address in future studies of culturable and non-culturable isolates.

Resistance induced in the laboratory may have different manifestations than clinical resistance, and the use of *Giardia* isolates directly from patients, without undergoing laboratory culturing are likely to supplement studies with isogenic laboratory strains.[69,70]

Non-coding regions

SNVs found in non-CDS regions of genes could affect regulatory elements and promoters which may further potentially affect gene transcription, mRNA translation and next protein expression.[35,71] The genes, NR-1 and the PFOR co-factor Fd2, were found to have especially high numbers of pre-CDS SNVs in *Giardia* assemblage B in our study. The sequence preceding the CDS, harbor transcription promoters, and based on scarce knowledge of these in *Giardia*, they are short, around 50 bp long, initiator-like AT-rich sequences approximately −65 to −29 base pairs upstream the ATG start codon.[57,72] Low conservation of the promoters in *Giardia* has previously been reported to be common, and may be related to our findings.[72] How the presence of SNVs in the promoter regions of genes in *Giardia* may affect transcription of the genes is however not known yet. The interpretation of the SNV findings in these regions are therefore limited, and in this study, we decided to present them, but not to evaluate their potential regulatory roles.

Ferredoxin domains and SNVs

The alignment of the NRs and Fds in this study, only include the ferredoxin domains, responsible for binding the Fe-S clusters (see Figure 2). This region has earlier been associated with drug activation, as the electrons are thought to be transferred from the ferredoxin domain to the active reduction center in order to reduce MTZ.[18] The active reducing center in the NRs has earlier been characterized as a nitroflavin mononucleotide (FMN) reductase domain.[18]

Most of the nsSNVs present in the ferredoxin domain alignment in Figure 2 will probably not alter the function of the proteins as the changes do not alter the basic properties, eg, polarity or charge, of the amino acid residues.

Some other changes are located close to the cysteine part of the 4Fe-4S binding motif and might, therefore, affect the geometry or the electron transfer properties of the Fe-S cluster. Activated MTZ has previously been speculated to be able to react with nucleotides and cysteines (free thiol groups) and form adducts.[38] Whether MTZ can form adducts with the proteins investigated in the present study and whether amino acid substitutions close to the cysteine domains could affect formation must be investigated further using structural- and molecular biology studies.

Other proteins that also have distinct Fd domains include the two large proteins PFOR-1 and PFOR-2. They have been hypothesized to have three motifs to bind 4FeS4 cluster. Even if the PFORs have their own Fd domains, they are still thought to use Fds in the cytosol for electron transfer and MTZ activation.[38]

Coverage and multicopy genes

Several of the genes in this study were found to have paralogs annotated in the genomes of *Giardia* assemblage A2 and B, and we aimed to include genes present in both of the *Giardia* assemblages. A total of 12 paralogs were found for the NADPH oxidoreductase, and an alignment was done to compare their similarity (data not shown). The alignment showed they were only 78.3% similar to one another, and probably not multicopy genes.

We were not able to investigate the true number of sSNVs and nsSNVs of the gene sequence of flavohemoprotein due to its variable higher coverage than the other genes, indicating that it is present in a variable number of copies in different isolates (Table S1). The SNVs could possibly be a combination of SNVs of all the alleles present of the flavohemoprotein gene and not true SNVs of one gene.

Two shorter paralogs of flavohemoprotein, DHA2_153759 and DHA2_152971 were identified, and had high identity to the longer flavohemoprotein DHA2_154000 (data not shown). Functions of these two proteins should be investigated in detail in the future to know if they have similar properties as the DHA2_154000, or if they potentially could be wrongly annotated or fragmented versions of DHA2_154000. The shorter flavohemoproteins could have affected the coverage of the *Giardia* assemblage A2 isolates, as they each would align to the 5′ or 3′ end of the DHA2_154000 gene. They do not overlap, leaving approx. 500 bp in the middle of the gene. The coverage was nevertheless even, and the two paralogs in assemblage A2 would also not explain the higher coverage of the assemblage B gene GSB_151570. The potential copy-number variation in this gene, coupled with amino acid changing SNVs could be important for understanding *Giardia*'s capability of handling oxidative stress, as flavohemoprotein is an important NO and O_2 scavenging enzyme.[16,20,36,37,73–76]

We also noted that histone H2A had 24 times lower coverage than the average in one of the assemblage A2 isolates. Possibly this gene could be missing, or found in another contig in this isolate, or a sequencing error may have occurred.

WGS limitations

The Illumina MiSeq paired-end technology sequencing used in our study gave a low coverage for several of the isolates and their genes, limiting the SNV analysis. For two assemblage B and one assemblage, A2 isolates the average coverage for the genes was only 11.3–14.0 (Table S1). Low coverage decreases the possibility of finding low prevalence SNVs, as the cutoff value was set to >10 in coverage. In addition, short reads of 2×300 bp prevented identification of full-length gene haplotypes in our study. The relatively crude reference genomes used in this analysis can also potentially hamper the SNV analysis, as the *Giardia* isolates used for obtaining the reference genomes have been grown in culture for decades and still consist of several hundred contigs, instead of the predicted five chromosomes.

Conclusion

Genetic variation in the form of SNVs is common in both of the human-infecting *Giardia* assemblages A2 and B. Assemblage B showed higher variation than assemblage A2, and presence of amino acid changing SNVs were especially high in genes associated with MTZ metabolism (ferredoxins, nitroreductase-1 and 2) and for some of the genes important in the detoxification of free radicals (thioredoxin peroxidase and NADH oxidase). Some of the discovered SNVs in this study could potentially affect the electron transfer properties of Fe-S clusters in some of the NRs and Fds investigated. Further studies are needed to address their functional properties, but study design can be guided by the presented findings. Analysis of genetic variation in circulating MTZ susceptible and treatment-refractory clinical *Giardia* isolates will be important to address the recent increase in MTZ resistant *Giardia* cases. The present study can be regarded as a basis for further studies into how genes in the metabolizing pathways may differ between resistant and susceptible *Giardia* isolates.

Ethical aspects

No patient data were used in the present study and any link between individual parasite data and patient information had been removed before WGS of trophozoites was carried out.

Data availability

The dataset used for this analysis and supporting the conclusions of this article can be found in GenBank with the accession numbers: MK043361 - MK043940.

Acknowledgments

We would like to thank the personnel at the Genomics Core Facility at Oslo University Hospital (oslo.genomics.no) for whole genome sequencing of the *Giardia* DNA and the personnel at the Department of Clinical Science, University of Bergen, for technical assistance with the sequencing. We acknowledge Prof Dr Ralf Ignatius from the Institute of Tropical Medicine and International Health, Charité, Berlin, Germany for kindly providing biological samples. We thank Petra Gosten-Heinrich, Robert Koch-Institute Berlin, for excellent technical assistance. This study was supported by the University of Bergen, the Norwegian National Advisory Unit on Tropical Infectious Diseases, Department of Medicine, Haukeland University Hospital, Helse Vest and Norsk overvaakningssystem for antibiotikaresistens hos mikrober (NORM). All the experiments, data analysis and evaluation were performed independently by the authors, without any interference from the funding institutions.

Author contributions

Culture of *Giardia* isolates and DNA extraction was performed by CK. Laboratory work related to whole genome sequencing and mapping of whole genome sequences was performed by HRB. SNV identification and alignment of the genes was performed by CSS. Homology modeling and amino acid change effects analysis for protein function was done by JPK and IK. CSS drafted the manuscript and KH supervised all parts of the study. All authors contributed toward data analysis, drafting and critically revising the paper, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Lane S, Lloyd D. Current trends in research into the waterborne parasite Giardia. *Crit Rev Microbiol*. 2002;28(2):123–147. doi:10.1080/1040-840291046713
2. Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. *Lancet (London, England)*. 2002;359(9306):564–571. doi:10.1016/S0140-6736(02)07744-9
3. Rogawski ET, Platts-Mills JA, Seidman JC, et al. Early antibiotic exposure in low-resource settings is associated with increased weight in the first two years of life. *J Pediatr Gastroenterol Nutr*. 2017;65(3):350–356.
4. Craun GF, Calderon RL, Craun MF. Outbreaks associated with recreational water in the United States. *Int J Environ Health Res*. 2005;15(4):243–262.
5. Farthing MJ. Giardiasis. *Gastroenterol Clin North Am*. 1996;25(3):493–515.
6. Hanevik K, Dizdar V, Langeland N, Hausken T. Development of functional gastrointestinal disorders after Giardia lamblia infection. *BMC Gastroenterol*. 2009;9:27.
7. Persson R, Wensaas KA, Hanevik K, Eide GE, Langeland N, Rortveit G. The relationship between irritable bowel syndrome, functional dyspepsia, chronic fatigue and overactive bladder syndrome: a controlled study 6 years after acute gastrointestinal infection. *BMC Gastroenterol*. 2015;15:66.
8. Lalle M. Giardiasis in the post genomic era: treatment, drug resistance and novel therapeutic perspectives. *Infect Disord Drug Targets*. 2010;10(4):283–294.
9. Solaymani-Mohammadi S, Genkinger JM, Loffredo CA, Singer SM. A meta-analysis of the effectiveness of albendazole compared with metronidazole as treatments for infections with Giardia duodenalis. *PLoS Negl Trop Dis*. 2010;4(5):e682.
10. Goldman P, Koch RL, Yeung TC, et al. Comparing the reduction of nitroimidazoles in bacteria and mammalian tissues and relating it to biological activity. *Biochem Pharmacol*. 1986;35(1):43–51.
11. Tian HF, Chen B, Wen JF. Giardiasis, drug resistance, and new target discovery. *Infect Disord Drug Targets*. 2010;10(4):295–302.
12. Leitsch D, Schlosser S, Burgess A, Duchene M. Nitroimidazole drugs vary in their mode of action in the human parasite Giardia lamblia. *Int J Parasitol Drugs Drug Resist*. 2012;2:166–170. doi:10.1016/j.ijpddr.2012.04.002
13. Uzlikova M, Nohynkova E. The effect of metronidazole on the cell cycle and DNA in metronidazole-susceptible and -resistant Giardia cell lines. *Mol Biochem Parasitol*. 2014;198(2):75–81. doi:10.1016/j.molbiopara.2015.01.005
14. Upcroft P, Upcroft JA. Drug targets and mechanisms of resistance in the anaerobic protozoa. *Clin Microbiol Rev*. 2001;14(1):150–164. doi:10.1128/CMR.14.1.150-164.2001
15. Raj D, Ghosh E, Mukherjee AK, Nozaki T, Ganguly S. Differential gene expression in Giardia lamblia under oxidative stress: significance in eukaryotic evolution. *Gene*. 2014;535(2):131–139. doi:10.1016/j.gene.2013.11.048
16. Ansell BR, Baker L, Emery SJ, et al. Transcriptomics indicates active and passive metronidazole resistance mechanisms in three seminal Giardia lines. *Front Microbiol*. 2017;8:398. doi:10.3389/fmicb.2017.00398
17. Muller J, Rout S, Leitsch D, Vaithilingam J, Hehl A, Muller N. Comparative characterisation of two nitroreductases from Giardia lamblia as potential activators of nitro compounds. *Int J Parasitol Drugs Drug Resist*. 2015;5(2):37–43. doi:10.1016/j.ijpddr.2015.03.001
18. Muller J, Schildknecht P, Muller N. Metabolism of nitro drugs metronidazole and nitazoxanide in Giardia lamblia: characterization of a novel nitroreductase (GlNR2). *J Antimicrob Chemother*. 2013;68(8):1781–1789. doi:10.1093/jac/dkt106
19. Nillius D, Muller J, Muller N. Nitroreductase (GlNR1) increases susceptibility of Giardia lamblia and Escherichia coli to nitro drugs. *J Antimicrob Chemother*. 2011;66(5):1029–1035. doi:10.1093/jac/dkr029
20. Mastronicola D, Falabella M, Forte E, Testa F, Sarti P, Giuffre A. Antioxidant defence systems in the protozoan pathogen Giardia intestinalis. *Mol Biochem Parasitol*. 2016;206(1–2):56–66. doi:10.1016/j.molbiopara.2015.12.002
21. Holmes AH, Moore LS, Sundsfjord A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet (London, England)*. 2016;387(10014):176–187. doi:10.1016/S0140-6736(15)00473-0
22. Munoz Gutierrez J, Aldasoro E, Requena A, et al. Refractory giardiasis in Spanish travellers. *Travel Med Infect Dis*. 2013;11(2):126–129. doi:10.1016/j.tmaid.2012.10.004
23. Requena-Mendez A, Goni P, Rubio E, et al. The use of quinacrine in nitroimidazole-resistant Giardia duodenalis: an old drug for an emerging problem. *J Infect Dis*. 2017;215(6):946–953. doi:10.1093/infdis/jix066
24. Nabarro LE, Lever RA, Armstrong M, Chiodini PL. Increased incidence of nitroimidazole-refractory giardiasis at the Hospital for Tropical Diseases, London: 2008–2013. *Clin Microbiol Infect*. 2015;21(8):791–796. doi:10.1016/j.cmi.2015.04.019
25. Meltzer E, Lachish T, Schwartz E. Treatment of giardiasis after nonresponse to nitroimidazole. *Emerg Infect Dis*. 2014;20(10):1742–1744. doi:10.3201/eid2010.140073
26. Requena-Mendez A, Goni P, Lobez S, et al. A family cluster of giardiasis with variable treatment responses: refractory giardiasis in a family after a trip to India. *Clin Microbiol Infect*. 2014;20(2):O135–O138. doi:10.1111/1469-0691.12327
27. Cruz A, Sousa MI, Azeredo Z, Leite E, Figueiredo de Sousa JC, Cabral M. Isolation, excystation and axenization of Giardia lamblia isolates: in vitro susceptibility to metronidazole and albendazole. *J Antimicrob Chemother*. 2003;51(4):1017–1020. doi:10.1093/jac/dkg150
28. Mastronicola D, Giuffre A, Testa F, et al. Giardia intestinalis escapes oxidative stress by colonizing the small intestine: a molecular hypothesis. *IUBMB Life*. 2011;63(1):21–25. doi:10.1002/iub.409
29. Lebbad M, Petersson I, Karlsson L, et al. Multilocus genotyping of human Giardia isolates suggests limited zoonotic transmission and association between assemblage B and flatulence in children. *PLoS Negl Trop Dis*. 2011;5(8):e1262. doi:10.1371/journal.pntd.0001370

30. Townson SM, Upcroft JA, Upcroft P. Characterisation and purification of pyruvate: ferredoxinoxidoreductase from Giardia duodenalis. *Mol Biochem Parasitol*. 1996;79(2):183–193.
31. Tejman-Yarden N, Millman M, Lauwaet T, et al. Impaired parasite attachment as fitness cost of metronidazole resistance in Giardia lamblia. *Antimicrob Agents Chemother*. 2011;55(10):4643–4651. doi:10.1128/AAC.00384-11
32. Leitsch D, Burgess AG, Dunn LA, et al. Pyruvate:ferredoxinoxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in Giardia lamblia. *J Antimicrob Chemother*. 2011;66(8):1756–1765. doi:10.1093/jac/dkr192
33. Dan M, Wang AL, Wang CC. Inhibition of pyruvate-ferredoxin oxidoreductase gene expression in Giardia lamblia by a virus-mediated hammerhead ribozyme. *Mol Microbiol*. 2000;36(2):447–456.
34. Muller J, Wastling J, Sanderson S, Muller N, Hemphill A. A novel Giardia lamblia nitroreductase, GlNR1, interacts with nitazoxanide and other thiazolides. *Antimicrob Agents Chemother*. 2007;51(6):1979–1986. doi:10.1128/AAC.01548-06
35. Emery SJ, Baker L, Ansell BRE, et al. Differential protein expression and post-translational modifications in metronidazole-resistant Giardia duodenalis. *GigaScience*. 2018;7:4. doi:10.1093/gigascience/giy024
36. Ansell BR, McConville MJ, Ma'ayeh SY, et al. Drug resistance in Giardia duodenalis. *Biotechnol Adv*. 2015;33(6 Pt 1):888–901. doi:10.1016/j.biotechadv.2015.04.009
37. Ma'ayeh SY, Knorr L, Svard SG. Transcriptional profiling of Giardia intestinalis in response to oxidative stress. *Int J Parasitol*. 2015;45(14):925–938. doi:10.1016/j.ijpara.2015.07.005
38. Leitsch D. A review on metronidazole: an old warhorse in antimicrobial chemotherapy. *Parasitology*. 2017;1–12. doi:10.1017/S0031182017002025
39. Leitsch D. Drug resistance in the microaerophilic parasite Giardia lamblia. *Curr Trop Med Rep*. 2015;2(3):128–135. doi:10.1007/s40475-015-0051-1
40. Begaydarova R, Yukhnevich Y, Babenko D, et al. Determination of PFOR gene expression in strains of G. intestinalis with different inhibitory concentrations of metronidazole. *J Infect Dev Ctries*. 2015;9(5):519–523. doi:10.3855/jidc.5768
41. Galeh TM, Kazemi A, Mahami-Oskouei M, et al. Introducing nitazoxanide as a promising alternative treatment for symptomatic to metronidazole-resistant giardiasis in clinical isolates. *Asian Pac J Trop Med*. 2016;9(9):887–892. doi:10.1016/j.apjtm.2016.07.013
42. Gabin-Garcia LB, Bartolome C, Abal-Fabeiro JL, Mendez S, Llovo J, Maside X. Strong genetic structure revealed by multilocus patterns of variation in Giardia duodenalis isolates of patients from Galicia (NW-Iberian Peninsula). *Infect. Genet. Evol*. 2016;48:131–141. doi:10.3855/jidc.5768
43. Teodorovic S, Braverman JM, Elmendorf HG. Unusually low levels of genetic variation among Giardia lamblia isolates. *Eukaryot Cell*. 2007;6(8):1421–1430. doi:10.1128/EC.00138-07
44. Muller J, Ley S, Felger I, Hemphill A, Muller N. Identification of differentially expressed genes in a Giardia lamblia WB C6 clone resistant to nitazoxanide and metronidazole. *J Antimicrob Chemother*. 2008;62(1):72–82. doi:10.1093/jac/dkn142
45. Hanevik K, Bakken R, Brattbakk HR, Saghaug CS, Langeland N. Whole genome sequencing of clinical isolates of Giardia lamblia. *Clin Microbiol Infect*. 2015;21(2):192.e191–e193. doi:10.1016/j.cmi.2014.08.014
46. Adam RD, Dahlstrom EW, Martens CA, et al. Genome sequencing of Giardia lamblia genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol Evol*. 2013;5(12):2498–2511. doi:10.1093/gbe/evt197
47. Ansell BR, McConville MJ, Baker L, et al. Divergent transcriptional responses to physiological and xenobiotic stress in Giardia duodenalis. *Antimicrob Agents Chemother*. 2016;60(10):6034–6045. doi:10.1128/AAC.00977-16
48. Nixon JE, Wang A, Field J, et al. Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to Giardia lamblia and Entamoeba histolytica. *Eukaryot Cell*. 2002;1(2):181–190.
49. Rice EW, Schaefer FW 3rd. Improved in vitro excystation procedure for Giardia lamblia cysts. *J Clin Microbiol*. 1981;14(6):709–710.
50. Keister DB. Axenic culture of Giardia lamblia in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg*. 1983;77(4):487–488.
51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–359. doi:10.1038/nmeth.1923
52. Aurrecoechea C, Brestelli J, Brunk BP, et al. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis. *Nucleic Acids Res*. 2009;37(Database issue):D526–D530. doi:10.1093/nar/gkn631
53. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845–858. doi:10.1038/nprot.2015.053
54. Nixon JE, Wang A, Morrison HG, et al. A spliceosomal intron in Giardia lamblia. *Proc Natl Acad Sci U S A*. 2002;99(6):3701–3705.
55. Fukuyama K, Matsubara H, Tsukihara T, Katsube Y. Structure of [4Fe-4S] ferredoxin from Bacillus thermoproteolyticus refined at 2.3 A resolution. Structural comparisons of bacterial ferredoxins. *J Mol Biol*. 1989;210(2):383–398.
56. Duee ED, Fanchon E, Vicat J, Sieker LC, Meyer J, Moulis JM. Refined crystal structure of the 2[4Fe-4S] ferredoxin from Clostridium acidurici at 1.84 A resolution. *J Mol Biol*. 1994;243(4):683–695.
57. Franzen O, Jerlstrom-Hultqvist J, Castro E, et al. Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog*. 2009;5(8):e1000560.
58. Thompson A. Human giardiasis: genotype-linked differences in clinical symptomatology. *Trends Parasitol*. 2001;17(10):465.
59. Feng Y, Xiao L. Zoonotic potential and molecular epidemiology of Giardia species and giardiasis. *Clin Microbiol Rev*. 2011;24(1):110–140.
60. Ellis JE, Wingfield JM, Cole D, Boreham PF, Lloyd D. Oxygen affinities of metronidazole-resistant and -sensitive stocks of Giardia intestinalis. *Int J Parasitol*. 1993;23(1):35–39.
61. Smith NC, Bryant C, Boreham PF. Possible roles for pyruvate: ferredoxinoxidoreductase and thiol-dependent peroxidase and reductase activities in resistance to nitroheterocyclic drugs in Giardia intestinalis. *Int J Parasitol*. 1988;18(7):991–997.
62. Goodwin A, Kersulyte D, Sisson G, Veldhuyzen van Zanten SJ, Berg DE, Hoffman PS. Metronidazole resistance in Helicobacter pylori is due to null mutations in a gene (rdxA) that encodes an oxygen-insensitive NADPH nitroreductase. *Mol Microbiol*. 1998;28(2):383–393.
63. Sisson G, Jeong JY, Goodwin A, et al. Metronidazole activation is mutagenic and causes DNA fragmentation in Helicobacter pylori and in Escherichia coli containing a cloned H. pylori RdxA(+) (Nitroreductase) gene. *J Bacteriol*. 2000;182(18):5091–5096.
64. Kobori T, Sasaki H, Lee WC, et al. Structure and site-directed mutagenesis of a flavoprotein from Escherichia coli that reduces nitrocompounds: alteration of pyridine nucleotide binding by a single amino acid substitution. *J Biol Chem*. 2001;276(4):2816–2823.
65. Paulish-Miller TE, Augostini P, Schuyler JA, et al. Trichomonas vaginalis metronidazole resistance is associated with single nucleotide polymorphisms in the nitroreductase genes ntr4Tv and ntr6Tv. *Antimicrob Agents Chemother*. 2014;58(5):2938–2943.
66. Bradic M, Warring SD, Tooley GE, et al. Genetic indicators of drug resistance in the highly repetitive genome of Trichomonas vaginalis. *Genome Biol Evol*. 2017;9(6):1658–1672.

67. Testa F, Mastronicola D, Cabelli DE, et al. The superoxide reductase from the early diverging eukaryote Giardia intestinalis. *Free Radic Biol Med*. 2011;51(8):1567–1574.
68. Kwon DH, El-Zaatari FA, Kato M, et al. Analysis of rdxA and involvement of additional genes encoding NAD(P)H flavin oxidoreductase (FrxA) and ferredoxin-like protein (FdxB) in metronidazole resistance of Helicobacter pylori. *Antimicrob Agents Chemother*. 2000;44 (8):2133–2142.
69. Caccio SM, Ryan U. Molecular epidemiology of giardiasis. *Mol Biochem Parasitol*. 2008;160(2):75–80.
70. Marais A, Bilardi C, Cantet F, Mendz GL, Megraud F. Characterization of the genes rdxA and frxA involved in metronidazole resistance in Helicobacter pylori. *Res Microbiol*. 2003;154(2):137–144.
71. Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. *Brief Bioinform*. 2016;17(6):980–995.
72. Elmendorf HG, Singer SM, Pierce J, Cowan J, Nash TE. Initiator and upstream elements in the alpha2-tubulin promoter of Giardia lamblia. *Mol Biochem Parasitol*. 2001;113(1):157–169.
73. Mastronicola D, Testa F, Forte E, et al. Flavohemoglobin and nitric oxide detoxification in the human protozoan parasite Giardia intestinalis. *Biochem Biophys Res Commun*. 2010;399 (4):654–658.
74. Gardner PR, Gardner AM, Martin LA, Salzman AL. Nitric oxide dioxygenase: an enzymic function for flavohemoglobin. *Proc Natl Acad Sci U S A*. 1998;95(18):10378–10383.
75. Hausladen A, Gow AJ, Stamler JS. Nitrosative stress: metabolic pathway involving the flavohemoglobin. *Proc Natl Acad Sci U S A*. 1998;95(24):14100–14105.
76. Mastronicola D, Falabella M, Testa F, et al. Functional characterization of peroxiredoxins from the human protozoan parasite Giardia intestinalis. *PLoS Negl Trop Dis*. 2014;8(1):e2631.
77. Martinez-Julvez M, Rojas AL, Olekhnovich I, Espinosa Angarica V, Hoffman PS, Sancho J. Structure of RdxA–an oxygen-insensitive nitroreductase essential for metronidazole activation in Helicobacter pylori. *FEBS J*. 2012;279(23):4306–4317.
78. Townson SM, Hanson GR, Upcroft JA, Upcroft P. A purified ferredoxin from Giardia duodenalis. *Eur J Biochem/FEBS*. 1994;220 (2):439–446.
79. Di Matteo A, Scandurra FM, Testa F, et al. The O2-scavenging flavodiiron protein in the human parasite Giardia intestinalis. *J Biol Chem*. 2008;283(7):4061–4068.
80. Leitsch D, Muller J, Muller N. Evaluation of Giardia lamblia thioredoxin reductase as drug activating enzyme and as drug target. *Int J Parasitol Drugs Drug Resist*. 2016;6(3):148–153.
81. Brown DM, Upcroft JA, Edwards MR, Upcroft P. Anaerobic bacterial metabolism in the ancient eukaryote Giardia duodenalis. *Int J Parasitol*. 1998;28(1):149–164.
82. Hofstetrova K, Uzlikova M, Tumova P, Troell K, Svard SG, Nohynkova E. Giardia intestinalis: aphidicolin influence on the trophozoite cell cycle. *Exp Parasitol*. 2010;124(2):159–166.
83. Sanchez LB. Aldehyde dehydrogenase (CoA-acetylating) and the mechanism of ethanol formation in the amitochondriate protist, Giardia lamblia. *Arch Biochem Biophys*. 1998;354(1):57–64.
84. Brown DM, Upcroft JA, Upcroft P. A H2O-producing NADH oxidase from the protozoan parasite Giardia duodenalis. *Eur J Biochem/FEBS*. 1996;241(1):155–161.
85. Sanchez LB, Elmendorf H, Nash TE, Muller M. NAD(P)H:menadioneoxidoreductase of the amitochondriate eukaryote Giardia lamblia: a simpler homologue of the vertebrate enzyme. *Microbiology*. 2001;147(Pt 3):561–570.
86. Muller J, Sterk M, Hemphill A, Muller N. Characterization of Giardia lamblia WB C6 clones resistant to nitazoxanide and to metronidazole. *J Antimicrob Chemother*. 2007;60(2):280–287.

Erratum: Genetic Variation in Metronidazole Metabolism and Oxidative Stress Pathways in Clinical *Giardia lamblia* Assemblage A and B isolates

Christina S. Saghaug[1,2], Christian Klotz[3], Juha P. Kallio[4], Hans-Richard Brattbakk[1,5], Tomasz Stokowy[1,5], Toni Aebischer[3], Inari Kursula[4,6], Nina Langeland[1,2,7], Kurt Hanevik[1,2]

[1]Department of Clinical Science, University of Bergen, Bergen, Hordaland, Norway.

[2]Norwegian National Advisory Unit on Tropical Infectious Diseases, Department of Medicine, Haukeland University Hospital, Bergen, Hordaland, Norway.

[3]Department of Infectious Diseases, Unit 16 Mycotic and Parasitic Agents and Mycobacteria, Robert Koch-Institute, Berlin, Germany.

[4]Department of Biomedicine, University of Bergen, Bergen, Hordaland, Norway.

[5]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Hordaland, Norway.

[6]Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland.

[7]Department of Medicine, Haraldsplass Deaconess Hospital, Bergen, Hordaland, Norway

ORCID numbers: Christina Saghaug: 0000-0003-0836-7889, Kurt Hanevik: 0000-0002-1466-2326, Nina Langeland: 0000-0003-0278-1616, Anton Aebischer: 0000-0002-4875-8457, Hans-Richard Brattbakk: 0000-0001-6503-6387, Tomasz Stokowy: 0000-0003-0017-8338, Inari Kursula: 0000-0001-5236-7056. Christian Klotz: 0000-0003-2609-9654

Correspondence: Christina Skaar Saghaug, University of Bergen, Department of Clinical Science, 8th floor, Lab-building, N-5021 Bergen, Norway

Tel: +47 90 13 24 14

E-mail: christina.saghaug@uib.no

This is an erratum to Infect Drug Resist. 2019; 12: 1221-1235, indicating that the nitroreductase abbreviations should be updated according to recent practice.[1]

There is confusion around the naming of *Giardia* nitroreductases in the literature. In our paper we consistently used NR-1 to denote the gene named in GiardiaDB as "Nitroreductase family protein fused to ferredoxin domain Fd-NR1" (DHA2_153380/GSB_153178), while in recent literature this is often denoted NR2/GlNR2.[2-7] Likewise, we used NR-2 to denote the gene named in GiardiaDB as "Nitroreductase Fd-NR-2" (DHA2_22677/GSB_22677), while in recent literature this is often denoted NR1/GlNR1.[2,3,5-7]

Thus, the NR gene names were confusingly abbreviated and cited in our published article.[1]

We acknowledge that using the NR gene names as annotated in GiardiaDB may cause confusion and therefore, we bring this to the attention of readers that more correctly what is named NR-1 in the publication should be read as NR-2, and NR-1 should be read as NR-2, in line with the terms used in recent publications.

1. Saghaug CS, Klotz C, Kallio JP, et al. Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical Giardia lamblia assemblage A and B isolates. *Infection and drug resistance.* 2019;12:1221-1235.
2. Müller J, Braga S, Heller M, Müller N. Resistance formation to nitro drugs in Giardia lamblia: No common markers identified by comparative proteomics. *International journal for parasitology Drugs and drug resistance.* 2019;9:112-119.
3. Müller J, Heller M, Uldry AC, Braga S, Müller N. Nitroreductase Activites in Giardia lamblia: ORF 17150 Encodes a Quinone Reductase with Nitroreductase Activity. *Pathogens (Basel, Switzerland).* 2021;10(2).
4. Emery SJ, Baker L, Ansell BRE, et al. Differential protein expression and post-translational modifications in metronidazole-resistant Giardia duodenalis. *GigaScience.* 2018;7(4).
5. Ansell BR, Baker L, Emery SJ, et al. Transcriptomics Indicates Active and Passive Metronidazole Resistance Mechanisms in Three Seminal Giardia Lines. *Frontiers in microbiology.* 2017;8:398.
6. Ansell BR, McConville MJ, Baker L, et al. Divergent Transcriptional Responses to Physiological and Xenobiotic Stress in Giardia duodenalis. *Antimicrobial agents and chemotherapy.* 2016;60(10):6034-6045.
7. Muller J, Muller N. Nitroreductases of bacterial origin in Giardia lamblia: Potential role in detoxification of xenobiotics. *MicrobiologyOpen.* 2019;8(10):e904.

II

Infection and Drug Resistance

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genetic Diversity of the Flavohemoprotein Gene of *Giardia lamblia*: Evidence for High Allelic Heterozygosity and Copy Number Variation

This article was published in the following Dove Press journal:
Infection and Drug Resistance

Christina S Saghaug[1,2]
Christian Klotz[3]
Juha P Kallio[4]
Toni Aebischer[3]
Nina Langeland[1,2,5]
Kurt Hanevik[1]

[1]Department of Clinical Science, University of Bergen, Bergen, Norway; [2]Norwegian National Advisory Unit on Tropical Infectious Diseases, Department of Medicine, Haukeland University Hospital, Bergen, Norway; [3]Department of Infectious Diseases, Unit 16 Mycotic and Parasitic Agents and Mycobacteria, Robert Koch-Institute, Berlin, Germany; [4]Department of Biomedicine, University of Bergen, Bergen, Norway; [5]Department of Medicine, Haraldsplass Deaconess Hospital, Bergen, Norway

Purpose: The flavohemoprotein (gFlHb) in *Giardia* plays an important role in managing nitrosative and oxidative stress, and potentially also in virulence and nitroimidazole drug tolerance. The aim of this study was to analyze the genetic diversity of *gFlHb* in *Giardia* assemblages A and B clinical isolates.

Methods: *gFlHb* genes from 20 cultured clinical *Giardia* isolates were subjected to PCR amplification and cloning, followed by Sanger sequencing. Sequences of all cloned PCR fragments from each isolate were analyzed for single nucleotide variants (SNVs) and compared to genomic Illumina sequence data. Identical clone sequences were sorted into alleles, and diversity was further analyzed. The number of *gFlHb* gene copies was assessed by mining PacBio de novo assembled genomes in eight isolates. Homology models for assessment of SNV's potential impact on protein function were created using Phyre2.

Results: A variable copy number of the *gFlHb* gene, between two and six copies, depending on isolate, was found. A total of 37 distinct sequences, representing different alleles of the *gFlHb* gene, were identified in AII isolates, and 41 were identified in B isolates. In some isolates, up to 12 different alleles were found. The total allelic diversity was high for both assemblages (>0.9) and was coupled with a nucleotide diversity of <0.01. The genetic variation (SNVs per CDS length) was 4.8% in sub-assemblage AII and 5.4% in assemblage B. The number of non-synonymous (ns) SNVs was high in *gFlHb* of both assemblages, 1.6% in A and 3.0% in B, respectively. Some of the identified nsSNV are predicted to alter protein structure and possibly function.

Conclusion: In this study, we present evidence that gFlHb, a putative protective enzyme against oxidative and nitrosative stress in *Giardia*, is a variable copy number gene with high allelic diversity. The genetic variability of *gFlHb* may contribute metabolic adaptability against metronidazole toxicity.

Keywords: Giardia, genetic diversity, copy number variation, flavohemoprotein, oxidative stress, nitrosative stress, allele

Introduction

Giardia lamblia is a microaerophilic protozoan parasite that infects up to 280 million humans annually by causing giardiasis.[1] This gastrointestinal infection is more common in developing countries, and may negatively affect growth properties and cognitive functions in children.[2,3] In developed countries giardiasis is usually related to sporadic waterborne outbreaks, or seen in travelers returning from endemic areas.[4,5] To treat giardiasis, the prodrug nitroimidazole antibiotic known as metronidazole (MTZ)

Correspondence: Christina S Saghaug
Department of Clinical Science, University of Bergen, 8th Floor, Lab-Building, Bergen N-5021, Norway
Tel +47 90 13 24 14
Email christina.saghaug@uib.no

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/IDR.S274543

is often used as first-line treatment.[6] Over the past few years, treatment failures with MTZ have been reported more frequently, and 10–40% of the cases will not be eradicated after a 5–7 day course of MTZ treatment.[7–9] The currently understood mode of action of MTZ depends on partial reduction (activation) resulting in highly reactive intermediates that further initiate damage to DNA and proteins, or on the other side, regeneration of MTZ, where oxygen radicals are created ('futile cycling').[10–12] The enzymes in the detoxification system in *Giardia* must then be able to handle oxidative stress resulting from metabolizing MTZ, as well as being able to neutralize substances such as molecular oxygen (O_2) and nitric oxide (NO) encountered in the gut habitat and released by the host.[13,14] There are several known enzymes connected to free radical neutralization, including flavoprotein, desulfoferredoxin (SOR), NADH oxidase and flavohemoprotein (gFlHb).[13,15] The gFlHb functions as a nitric oxide dioxygenase, responsible for catalyzing the formation of nitric oxide (NO) to nitrate (NO_3^-) by using O_2 as a co-factor.[16,17] The gFlHb enzyme may also possess NADH/NAD(P)H oxidase activities, similar to flavodiiron protein, by catalyzing O_2 to H_2O in low NO level conditions.[18] The *gFlHb* gene was characterized a decade ago, and is likely to be an important detoxification enzyme in *Giardia*.[15,18] It has been shown to be up-regulated during exposure to oxidative stresses caused by both O_2 and H_2O_2, in addition to being upregulated during nitrosative stress.[14,15,19] It was also recently shown that gFlHb protein levels were increased during MTZ exposure in an MTZ and nitazoxanide (NTZ) resistant *Giardia* isolate.[20] In another anaerobic or microaerophilic pathogen, *Trichomonas vaginalis*, it has been observed that in vitro MTZ-resistant parasites can handle higher levels of oxygen than susceptible ones, probably linking resistance to increased tolerance or better mechanisms to handle oxidative stress.[21] It has been proposed that refractory *Giardia* has a higher tolerance towards O_2, as O_2 will compromise the activation of MTZ through futile cycling, but also potentially through an O_2 induced resistance mechanism.[22] Because gFlHb uses O_2 as a co-factor for converting NO to NO_3^-, it could also be relevant for increased MTZ tolerance in this way.

Activities of enzymes having similar properties as the gFlHb have been shown to be inhibited by bulky imidazoles (such as azoles; miconazole, econazole and ketoconazole) by binding to the heme pocket of the protein and generating reactive oxygen species (ROS), which may well mean that MTZ could potentially affect the function of the gFlHb enzyme, and link it to *Giardia*'s ability to handle and tolerate the toxic effects of MTZ.[23,24] *Giardia* is a functionally tetraploid organism, with two diploid nuclei, each of them harboring two sets of its 5 chromosomes.[25–27] However, aneuploidy with unequal distribution of chromosomes may also occur.[25,28] Thus, even single-copy genes may occur in up to four versions, ie, alleles, per strain.[25,28]

The degree of allelic sequence heterozygosity (ASH), has been shown to differ between assemblages of *G. lamblia*. Current genome data suggests a comparably low ASH in sub-assemblage AI (<0.01%), a little higher in sub-assemblage AII (based on reference strain DH with an ASH of 0.04%), and highest in assemblage B isolates (cf GS reference strain's ASH of 0.5%).[25,29,30] Genetic diversity at the allele level, especially in assemblage B, has been analyzed mainly for typical genotyping genes or housekeeping genes (β-giardin (*bg*), glutamate dehydrogenase (*gdh*), elongation factor 1-alpha (*ef-1*), mlh1 (*mlh*), the FLORF-C4 (*C4*) and triosephosphate isomerase (*tpi*).[29,31–34] The allelic forms, however, have rarely been determined by cloning.[35–41] In a recent study on the genetic diversity of genes involved in MTZ induced oxidative and nitrosative stress management, we found indications that *gFlHb* genes may not only present allelic variation, but also be present at variable copy numbers per haploid genome.[42] The current study is a follow-up study of this recent published article.[42] As gFlHb may play a role in the ability of *Giardia* to tolerate MTZ, we performed this study to further explore the copy number variability and the allelic diversity of *gFlHb* in *G. lamblia* strains representing recent sub-assemblage AII and assemblage B isolates.

Materials and Methods

Giardia Lamblia Isolates

Trophozoite cultures of a recently established *G. duodenalis* biobank at the Robert Koch-Institute in Berlin, containing twelve *Giardia* sub-assemblage AII and eight *Giardia* assemblage B isolates, were cultured according to the methods of Keister.[43] Collection, DNA extraction, concentration measurements and Illumina-based whole genome sequencing were carried out as previously stated.[42] No clinical data have been collected from the clinical samples of *Giardia.*

Cloning of Flavohemoprotein

The coding regions and approximately 150 bp up/downstream of the gene flavohemoprotein, *gFlHb*, in sub-

assemblage AII (DHA2_154000) and B (GSB_151570) were amplified from genomic DNA by PCR. The reference genomes for *Giardia* sub-assemblage AII and assemblage B (v.26) were downloaded from *Giardia* DB January 2016 versions 2013–11-25.[44] Specific primers for the two orthologs of *gFlHb* of sub-assemblage AII and assemblage B were designed in Geneious Prime® 2019.0.3 (Biomatters Ltd., Auckland, New Zealand) (see Table 1) based on conserved pre-and post-CDS regions of the *gFlHb* gene represented in twelve assemblage A and eight assemblage B whole genome sequenced isolates previously presented in a former study.[42] Each 25 µL PCR reaction contained 1X Q5 Reaction Buffer (cat. nr: B90276, New England BioLabs (NEB) Ipswich, MA, USA), 2 mM $MgCl_2$ (included in buffer), 0.2 mM dNTP's (catalog nr N0447L, NEB), 0.5 U Q5 High-Fidelity DNA Polymerase (catalog nr M0491L NEB) 0.2 µM each of forward and reverse primer, 1 µL template DNA and nuclease-free water up to 25 µL. Negative controls using master mix reactions with nuclease-free water were included in all PCR experiments. All PCR reactions included an initial denaturation at 98°C for 30s, followed by 35 cycles of denaturation at 98°C for 10 s, annealing for 30 s (55°C for sub-assemblage AII and 60°C for assemblage B), extension at 72°C for 45–60 s and final extension at 72°C for 3 minutes. Amplified PCR products were run on 1% Agarose gels stained with GelGreen® Nucleic Acid Stain (catalog nr: 41,005, Biotium, San Fransisco, CA, USA) and positive bands were cut from the gel using blue light illuminator (Serva, Heidelberg, Germany) and extracted using Wizard SV gel and PCR Clean-Up System (catalog nr: A9282, Promega, Madison, WI, USA) according to manufacturer's descriptions, the only exception was using 70°C nuclease-free water in the last elution step. Concentrations of the PCR products were measured using Quantus™ Fluorometer (catalog Nr: E6150, Promega). DNA was available from 19 of the original Illumina sequenced samples, and two of the isolates were not included in the *gFlHb* gene cloning experiments due to unsuccessful amplification of the gene target.

The pJET 1.2/blunt cloning vectors (catalog nr: K1232, CloneJET PCR Cloning kit, Thermo Fisher Scientific, Waltham, MA, USA) were used in all of the cloning experiments. The blunt end protocol for ligation of PCR-product to vector was followed according to the manufacturer's instructions. The ligation mixtures were used immediately or kept at −20°C until usage. The ligation mixtures were introduced into *Escherichia coli* DH5a competent cells by heat-shock transformation using standard protocols.[45] The *E. coli* was plated on lysogeny broth (LB) agar plates containing ampicillin (100 µg/mL) (catalog nr: 10,835,269,001 (Sigma-Aldrich), Merck KGaA, Darmstadt, Germany) and cultured overnight at 37°C in a CO_2 incubator. Approximately 20 clones from each *Giardia* isolate were picked for colony PCR. 14–20 positive clones were cultured overnight in a shaking incubator at 37°C in Falcon® 17x100 mm, 14 mL high-clarity polypropylene (PP) round bottom test tubes (Item nr: T7597-14F, Corning Life Sciences, NY, USA). The plasmids were purified from the overnight cultures using Zyppy plasmid miniprep kit (catalog nr: D4037, Zymo Research Corp., CA, USA) as advised by the manufacturer and DNA was eluted using nuclease-free water. The plasmid concentrations were measured using Quantus and sequenced using pJET1.2 F and R sequencing primers provided in the cloning kit together with the BigDye™ Terminator v3.1 Cycle Sequencing kit (catalog nr: 4,337,455, Thermo Fisher Scientific). The PCR products were Sanger sequenced at the Sequencing Laboratory of the Robert-Koch Institute, Berlin, Germany.

Data Analysis of Sequences and Single Nucleotide Variation

SNV-called Illumina sequences of *gFlHb* presented in a former study, accession numbers MK043521.1-MK0435

Table 1 Primers Designed for Gene-PCR of the *gFlHb* of *Giardia lamblia Sub*-Assemblage AII and Assemblage B All Primers Were Ordered from Eurofins Genomics (Ebersberg, Germany)

Primer Set	Type[a]	Sequence, 5'-3'	Amplicon Size	Melting Temperature [°C]
DHA2_154000	F	CGCCACCACAAGCGATCATT	1433	59.4
DHA2_154000	R	GTGTGTAGAGCGATTACAT		52.4
GSB_151570	F	CGGCCTTCAGGTACTTCCCC	1807*	63.5
GSB_151570	R	GAGACCAAAAGTCCATATGAACT		57.1

Notes: *The GSB *gFlHb* gene exists in a longer version which is not likely to be the true length, however, the whole annotated gene obtained from the reference genome was covered by the primers. Only the CDS length of 1377 bp (same as in sub-assemblage AII) was analyzed in the present study.
Abbreviations: [a]F, forward; R, reverse.

40.1 from Genbank (NCBI), were used in the present study to investigate SNVs.[42] A SNV in the Illumina sequencing data was defined by a read coverage of minimum 10 and variant frequency above 0.1 (minimum 2 reads would need to have the SNV if the coverage was 10). The Illumina sequenced files with read data were used to compare and verify SNV positions in the cloned sequences in Geneious Prime® (Biomatters Ltd.). The forward and reverse sequences of one clone were aligned against the respective reference gene orthologs DHA2_154000 and GSB_151570 and trimmed at start and end of the sequences before further analysis. The chromatograms were inspected manually, and sequences with ambiguous base calling, or not covering the full CDS were not included. To manage potential random amplification or sequencing errors in the present study, we set criteria to exclude likely false SNVs by verifying SNVs found in clones with Illumina read data. SNVs found in the clones in the present project were termed high confidence (HC) and SNVs and included in the analysis if one or several of the following criteria were met:

- the SNV position was also found to be present in Illumina data for the same isolate.
- multiple clones would have the same SNV within one isolate.
- if the SNV was not found in Illumina data as a major nucleotide, the reads in the Illumina were checked for the presence of the SNV and the nucleotide had to be presented by two reads or more.

The SNVs that were not classified as HC were discarded from the analyses and termed low confidence (LC) SNVs. De novo assemblies were generated for eight of the 20 *Giardia* isolates at RKI using the software tools HGAP 2.0 (PacificBiosciences) as described elsewhere (Klotz et al 2020, manuscript in preparation). The gFlHb copy number was retrieved by mapping whole or partial sequences of the *gFlHb* AII and B reference sequences to de novo assemblies of PacBio consensus genomes of each isolate. To also identify partial genes or genes split between contigs, mapping of 20 bp parts of the start, middle and end of the reference genes were also performed. The identified *gFlHb* copies in PacBio consensus sequences were then extracted, and aligned together with the cloned allele sequences of the *gFlHb* gene separately for sub-assemblage AII and B. Further phylogenetic trees (Tamura-Nei with method UPGMA) were made. All sequence analysis work was carried out in Geneious Prime®. All nonsynonymous (ns) substitutions were analyzed using Geneious and DNA Sequence Polymorphism (DnaSP) v6.[46,47] Nucleotide diversity (pi) was calculated as the average number of nucleotide differences in gene sequences for pairwise comparisons. Haplotype diversity, defined as allelic diversity in the present study, gives a measure of the uniqueness of a specific allele in a population, or the probability that two alleles differ from one another.

Characterization of Amino Acid Changes

The bacterial genetic code was used for translation of the nucleotide codons in the gene sequences of *gFlHb* (DHA2_154000 and GSB_151570) in sub-assemblage AII and assemblage B. The abbreviations and characterization of the amino acid (aa) changes are based on the International Union of Pure and Applied Chemistry (IUPAC). As the crystal structure of the protein is not known, the estimation as to whether the aa change would affect the protein structure and function was based on homology models. Homology modeling was carried out using Protein Homology/analogy Recognition Engine V2.0 (Phyre2).[48] Protein sequences Uniprot ID: E2RTZ4 and A0A482ESB4 were selected to create models for gFlHb assemblage A and B, respectively. The model was created using the single highest scoring template, a crystal structure of *E. coli* flavohemoglobin (PDB-ID 1GVH) sharing a 40% sequence identity with the gFlHb.[49] Illustrations for the homology models were created using PyMol. A conservative replacement was identified as an aa substitution between two aa with rather small physicochemical distance, whilst a radical aa change would be considered as aa having large physicochemical characteristics, or potential rearrangement of the secondary structure.[50]

Results

gFlHb Copy Number Variation and Allele Diversity

Re-analysis of the mapping of Illumina sequencing data from the AII isolates onto the AII reference strain DH, showed that higher coverage for the *gFlHb* gene was present along the whole coding regions (CDS) of the gene and adjacent 0–230 bp upstream and 461–1271 bp downstream of the reference CDS.[42] For the B isolates the coverage was found to be higher starting from 600 to 816 bp upstream of the GS reference CDS and extending to 90–913 bp downstream of it. PacBio consensus sequences derived from de novo assembly were available for eight isolates, five assemblage B isolates and three sub-assemblage AII isolates. More than one copy of the *gFlHb*

gene was found in the PacBio sequences-derived assemblies of all three AII and five B isolates. In general, higher Illumina coverage fold difference corresponded to higher copy numbers in PacBio sequence assemblies. Isolate P344 had the same estimated copy number of *gFlHb* based on both analyses. Isolate P424 showed the highest relative coverage in Illumina sequence data with 5.1 fold the average and six copies were detected in the respective PacBio sequence derived assembly. The respective values for P392 were 3.6 and four (see Table 2 for more information). The *gFlHb* gene was successfully amplified by PCR from 17 of the 20 isolates (11 AII isolates and 6 B isolates). Between five and seventeen clones per isolate were obtained (Table 2). The cloned *gFlHb* genes were categorized into different alleles depending on their sequences whereby 37 alleles were identified in eleven sub-assemblage AII isolates, and 41 alleles were identified in six assemblage B isolates. In the AII isolates, the number of *gFlHb* alleles identified by cloning varied from two to eight, while in the B isolates two to twelve alleles per isolate were found (Table 2). The allelic diversity, the probability that two alleles are different from one another, for sub-assemblage AII was calculated to be 0.94 ± 0.01 and 0.95 ± 0.01 for assemblage B. To evaluate how alleles determined by cloning were represented in the *gFlHb* genes identified by de-novo assembled consensus genomes from PacBio sequenced isolates, we aligned all the CDS sequences and analyzed correspondence by creating phylogenetic trees (Figures 1 and 2). The phylogenetic tree of sub-assemblage AII in Figure 1 represents the relations between the alleles and PacBio copies found in the different isolates. Several allelic forms of *gFlHb* were found in different isolates, for example one allele was found to be present in four isolates represented by allele AA16 found in P033, P506, P034 and P064. A total of six alleles were found to be present in more than one isolate (table S4). A total of nine PacBio consensus sequences of *gFlHb* were available from three AII isolates, representing the five, three and two copies found in isolates P392, P407 and P064, respectively. These sequences were aligned to the cloned sub-assemblage AII alleles in order to check whether cloned alleles would match these consensus sequence copies. Three PacBio copies were found to be identical to alleles in their respective isolates (table S4), whereas the P407 AA37 allele and P407 PacBio copy 3 were identical with the exception of three, likely artefactual, indels in the PacBio consensus sequence.

Table 2 Estimates of gFlHb Copy Numbers in PacBio and Illumina Sequencing Data, Number of Clones, and Identified Alleles and SNVs in Cloned Sequences

Isolate	Average Coverage 5 Largest Contigs	Coverage *gFlHb* CDS	Estimated Number of Copies by Illumina	Copies Found in PacBio	Number of Clones Sanger Sequenced	Unique *gFlHb* Alleles per Isolate	Number of HC SNVs in all Clones	HC nsSNVs in all Clones
P033	30.6	64.9	2.1	Nd	11	3	18	6
P034	29.7	64.9	2.2	Nd	12	2	6	2
P064	94.2	97.9	1	2	11	2	6	2
P316	26.0	106.3	4.1	Nd	11	5	13	4
P324	30.1	97.9	3	Nd	NA	NA	NA	NA
P361	31.0	84.5	2	Nd	14	4	13	4
P368	42.4	139.9	3	Nd	5	2	23	9
P392	51.0	183.3	3.6	4	12	8	32	12
P403	27.8	86.9	3	Nd	8	2	26	5
P407	43.5	235.9	5.4	3	12	8	48	16
P478	11.5	106.3	3.4	Nd	9	7	34	10
P506	29.8	64.9	2	Nd	12	2	7	3
P344	23.7	45.4	2	2	15	9	22	7
P387	36	126.3	3.5	2	11	9	24	13
P413	23	110.6	4.8	Nd	13	12	28	21
P424	22.4	114.6	5.1	6	17	9	20	12
P427	17.2	24.9	1.5	2	14	3	21	10
P428	14	20.5	1	Nd	NA	NA	NA	NA
P433	13.4	13.1	1.5	Nd	NA	NA	NA	NA
P458	26.1	44.2	1.7	3	7	2	14	7

Abbreviations: Nd, not done; HC, high confidence; CDS, coding regions.

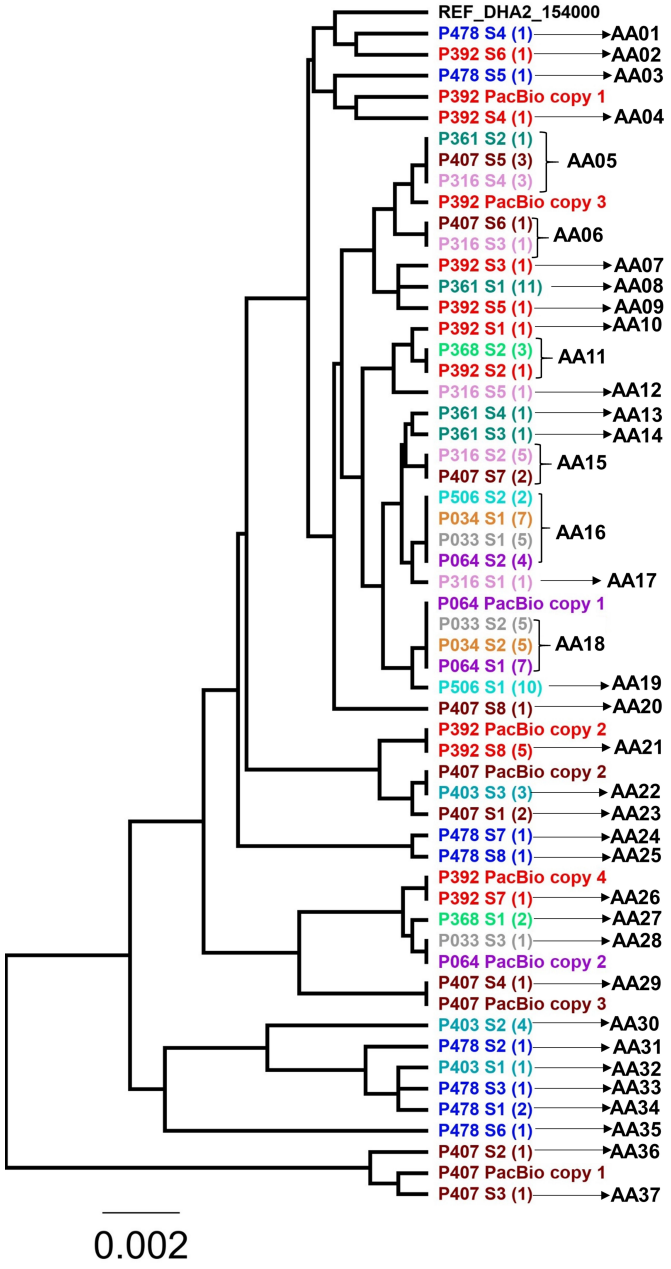


Figure 1 Phylogenetic tree of all the alleles of *gFlHb* found in the sub-assemblage AII isolates and *gFlHb* copies found in the PacBio sequencing data. A is abbreviated for allele, and the number of clones representing each allele is listed in the parenthesis.

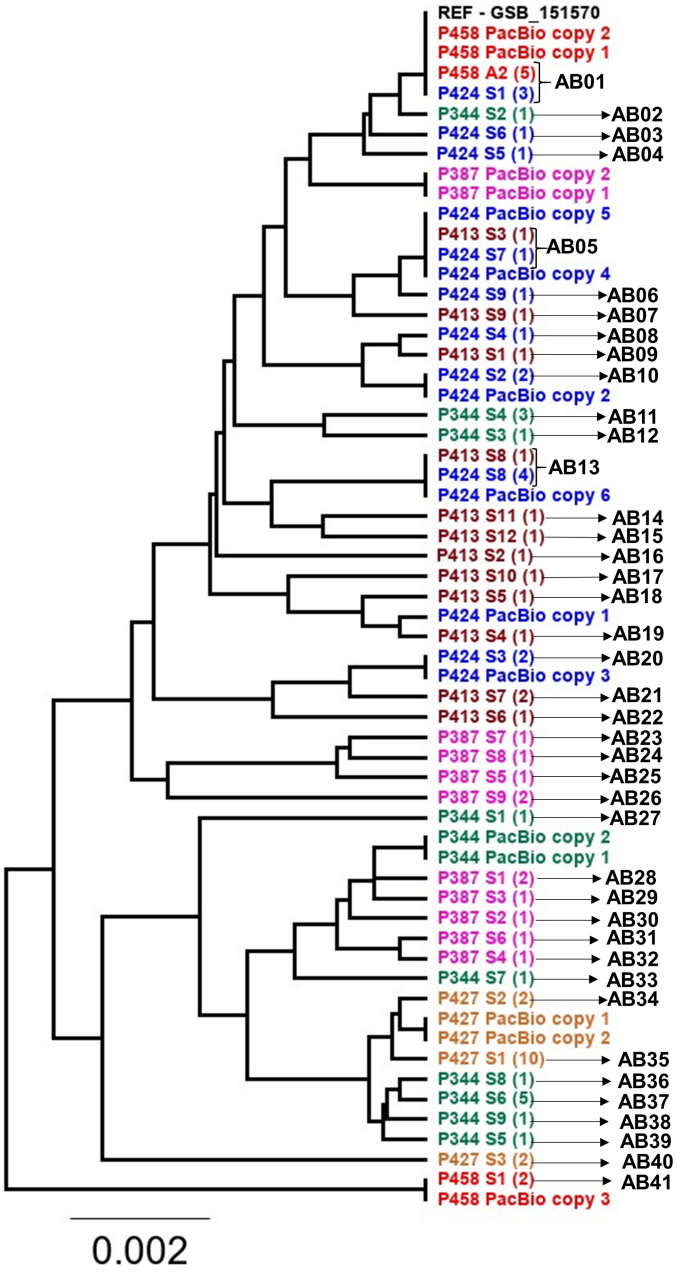


Figure 2 Phylogenetic tree of all the alleles found in the assemblage B isolates and *gFlHb* copies found by PacBio sequencing. A is an abbreviation of allele, and the number of clones representing each allele is listed in the parenthesis.

P392 PacBio copy 1 and 3 and P407 PacBio copy 1 and 2, were not identical to alleles from the same isolate, but closely related (2–6 SNVs difference), and were likely to be part of the alleles resulting in the consensus PacBio copy. Intriguingly, two PacBio copies, P064 copy 2 and P407 copy 2 were identical to alleles found in other isolates, suggesting that sequencing of more clones would be necessary to pick up all possible alleles of each isolate. The relations between the 41 alleles and PacBio copies that were discovered in assemblage B are presented in phylogenetic tree in Figure 2. Fewer alleles were shared between the assemblage B isolates than in the sub-assemblage AII isolates, where only three alleles were found in more than one isolate, such as AB01 in isolates P424, P458 and reference sequence GSB_151570, and alleles AB05 and AB13 in isolates P424 and P413 (see table S5).

Fifteen PacBio consensus sequences of *gFlHb* were obtained from five assemblage B isolates, and the copies were checked for similarity to the cloned alleles. Eight of the copies, all in P458 and P424, were found to match cloned alleles in the same isolates. For the isolates P344, P387 the PacBio copies were not identified in the clones, however the alleles from isolate P427 were highly similar to the two identical PacBio copies found in the isolate (1 SNV difference). The PacBio copies found in isolate P387 had 14 bp shorter sequences, probably due to assembly artefacts. Isolate P344 had two identical PacBio copies and the closest allele, P344 AB33, had a total of four SNVs different from the PacBio sequences. Some of the *gFlHb* PacBio copies in assemblage B were also found to be identical, such as P458 copy 1 and 2, P387 copy 1 and 2, P424 copy 4 and 5 (table S5). See supplementary box 1 for more information about the PacBio copies.

The copy number found in the PacBio data was compared to the number of cloned alleles. For the three sub-assemblage AII isolates with available PacBio data, more alleles were identified in isolates with a higher copy number (P064 two copies and two alleles, P392 four copies and eight alleles and P407 three copies and eight alleles). Among the assemblage B isolates, two isolates with two identified PacBio copies (P344 and P387) both had a total of nine identified alleles, thus higher than the theoretical maximum for a tetraploid organism with two gene copies. Isolates P427 and P458, both with two identified PacBio copies, had three and two alleles, respectively. In the isolate with the highest number of identified copies, P424, with six copies, nine alleles were identified.

Sequence Variation of the gFlHb Gene in Giardia Assemblages A and B

The identified SNVs in the cloned sequences were compared against Illumina sequencing data and categorized into high confidence (HC) and low confidence (LC) SNVs using the algorithm described in methods. LC SNVs were discarded and not used in the analysis, see Supplementary Table S1 for full overview of HC and LC SNVs. Sequence analysis showed a higher proportion of nsSNVs/SNVs in assemblage B isolates (53%) than sub-assemblage AII isolates (38%) (Table S1). The number of SNV positions found in the respective cloned *gFlHb* CDS for each isolate varied from 6 to 48 for sub-assemblage AII and 14–28 for assemblage B sequences (Table 2). One AII isolate, P407, had the most SNV positions of all isolates with 48 SNV positions found, while the numbers of SNVs were more congruent among the B isolates with approx. 20 SNVs per isolate. A total of 4.8% and 5.4% of all positions of the *gFlHb* CDS showed some variation in AII and B, respectively (see Tables S2–S5 and S6-S7). The total number of nsSNV positions per CDS was lower for AII isolates (22 positions, 1.6%), than for B isolates (42 positions, 3.1%). The average nucleotide diversity, pi, between cloned alleles was calculated to be 0.007 in sub-assemblage AII isolates' sequences and 0.009 in assemblage B isolates. Generally, there were more nsSNVs found in just one or a few isolates in assemblage B isolates, compared to sub-assemblage AII, where numerous nsSNV positions were common for several isolates (see Tables S3-S4).

Amino Acid Changes and Relation to Predicted gFlHb Protein Structure

Based on homology modelling gFlHb shares the same domain structures as the homologous FlHb from bacteria[49,51] and yeast.[52] It is a protein formed of 3 structural domains: (1) heme binding globin domain; (2) a FAD-binding domain; (3) a C-terminal NAD binding domain. Homology models for gFlHb for assemblages A and B were generated and are shown in Figure 3, where the detected nsSNVs that induced aa mutations are visualized. As noted before, assemblage B has a considerably higher number of aa substitutions than sub-assemblage AII. Most of the mutations are predicted to concern residues at the protein surface and, more accurately, those located in loop regions. Also, the majority is considered to only moderately affect the physicochemical

properties, eg, simply altering the size of a hydrophobic side chain. Some mutations, however, will cause a reversal of charge, hence affect local electrostatic properties of the protein surface. Some of the changes are predicted to be in near vicinity of the heme or FAD-binding sites and could have an effect for accommodating these co-factors. These changes are marked in red on Figure 3. The most drastic change amongst sub-assemblage AII alleles is a premature stop codon at position 49 in alleles AA24 and AA25 in isolate P478 that likely results in early truncation within the first domain. In addition, there was a change of aromatic Y at position 211 to basic and positively charged H in several alleles (AA07, AA10, AA11, AA12, AA13 and AA15) observed in isolates P316, P361, P368, P392 and P407. This mutation is located in a putative FAD-binding region of domain 2. For assemblage B, the most different allelic forms are obviously those with the deletion Del-$_{75}$HEL$_{77}$ that is located to the helix above the heme binding site, and the E80G variant is located on the same helix. Both the deletion and the E80G mutation are potential helix breakers that could disturb the secondary structure and affect heme binding.

Discussion

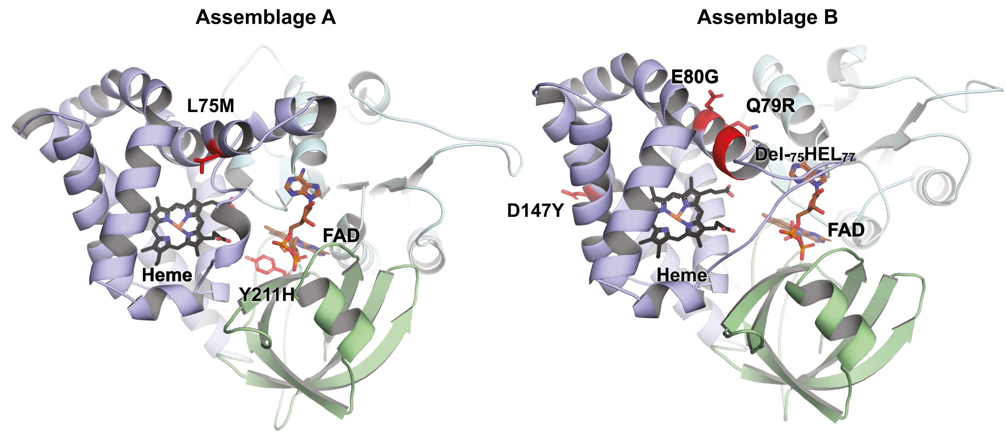
In this study, the allelic diversity of *gFlHb* was investigated through combined analyses of cloned sequences, genomic Illumina sequencing data, and de-novo assembled genomes derived from PacBio sequencing.

Sequencing Methods, PCR, Cloning and SNVs

Studies using a clone-based approach to identify SNPs in *Giardia* genes have been conducted earlier.[37,38,41] In some of these studies a Taq-based polymerase was used for the PCR-amplification, and may have caused an inflated high diversity as the error rate of standard Taq is around 1 in every 3500 bp.[53] However, some of the studies have used Q5 or other high-fidelity polymerases that have a lower error rate.[36,39,40,54] When identifying allele sequences it is important to utilize high-fidelity polymerases to minimize the chance of introducing errors during amplification.[40] The error rate of the Q5 polymerase used in this study has earlier been reported to be 1 per million bp.[53] Still, some SNVs identified in the clones could potentially be caused by amplification or sequencing errors, specifically within reads in the beginning or end of the Sanger sequences. The 35 cycles used in the initial gene-PCR reactions could introduce amplification errors leading to false-positive SNVs in the clones. Validation of SNVs therefore becomes important. Indeed, identification of HC SNVs, and excluding LC SNVs from the analysis reduced the number of SNVs from an average of 23 to 21 in eleven sub-assemblage AII isolates and from 37 to 22 for six assemblage B isolates (table S1). We, therefore, consider the reported number of alleles and diversity as a conservative interpretation of the data. We could see that the majority of the SNVs identified in the clones were also identified in Illumina data for the same isolate. The clones had generally more unique SNVs than the number identified in Illumina data. However, for one isolate (P344), two different batches of DNA (trophozoites cultured an additional time to obtain enough DNA) were used for the Illumina sequencing data and the cloning experiments. Although derived by limiting dilution, P344 may represent a mixed isolate with variable contribution to, and dominance of, co-existing lineages between batches, and possibly introduction of new mutations. These could all be reasons why this isolate had the highest number of LC SNVs, and *gFlHb* alleles matching other alleles in the B assemblage isolates.

General Features of gFlHb Genetic Diversity

Our analysis of SNVs in the CDS and copy number variation of the two ortholog *gFlHb* genes, DHA2_154000 and GSB_151570 in *Giardia* sub-assemblage AII and assemblage B reference strains DH and GS, respectively, show a high degree of genetic variation (number of variable SNV positions per CDS length) for both assemblages compared to the set of 29 MTZ-metabolizing and other metabolism genes analyzed in a previous study.[42] One interesting finding in the present study is that sub-assemblage AII isolates, shown to generally have less genetic variation than assemblage B, possessed a higher number of SNVs than assemblage B isolates in *gFlHb*.[30,38,55,56] In addition, a relatively high allelic heterozygosity of the *gFlHb* gene coupled with nucleotide diversity values <0.01 was found in both assemblages. Other studies looking at nucleotide diversity in single copy genes glutamate dehydrogenase gene (*gdh*), beta-giardin (*bg*) and triosephosphate isomerase (*tpi*), that are often used for genotyping, have found similar nucleotide diversity ranging from 0.003 to 0.02.[36,39] It is challenging to culture clinical isolates of *Giardia*, especially for assemblage B parasites.[57,58] *Giardia* strains that are able to grow in



	Isolate number	Amino acid change																			
Domain 1	Assemblage A																				
	P033	N42D				D137N	I138V		E158D												
	P034	N42D				D137N	I138V														
	P064	N42D				D137N	I138V														
	P316	N42D			L75M	D137N	I138V														
	P361	N42D				D137N	I138V														
	P368	N42D			L75M	D137N	I138V														
	P392	N42D			L75M	D137N	I138V														
	P403	N42D					I138V														
	P407	N42D			L75M	D137N	I138V														
	P478	N42D	STOP-49Q*	A66V		D137N	I138V														
	P506	N42D				D137N	I138V														
	Assemblage B																				
	P344			V43I			Q79R														
	P387			V43I			Q79R	E80G	E88G	L93P	I119V										
	P413	E5G	D6A		I72V	Del-75HEL77						D135N									
	P424					Del-75HEL77	Q79R					D135N									
	P427			V43I			Q79R						D147Y								
	P458						Q79R														
Domain 2	Assemblage A																				
	P033																				
	P034				G264D																
	P064				G264D																
	P316				G264D																
	P361			Y211H	G264D																
	P368			Y211H	G264D																
	P392			Y211H	G264D																
	P403			Y211H	G264D																
	P407		G202D		G264D																
	P478	V195I		Y211H	G264D																
	P506		G202D		G264D																
					G264D																
	Assemblage B																				
	P344			I185V			T233I			K268E											
	P387		A183V				T233I		K249R	K268E											
	P413	E180K		I185V				T239M		K268E											
	P424	E180K		I185V				T239M		K268E											
	P427			I185V	F190S	A197T	T233I			K268E											
	P458							T239M		K268E	A270V										
Domain 3	Assemblage A																				
	P033		P288A	A289V	Q295P			N354D	A366T			A413V		N443D	S444N						
	P034		P288A	A289V	Q295P			N354D	A366T												
	P064		P288A	A289V	Q295P			N354D	A366T												
	P316		P288A	A289V	Q295P			N354D	A366T												
	P361		P288A	A289V	Q295P				A366T												
	P368		P288A	A289V	Q295P			N354D	A366T			A413V		N443D	S444N						
	P392		P288A	A289V	Q295P			N354D	A366T		I405V	A413V	A416V	N443D	S444N						
	P403	A283V	P288A	A289V	Q295P			N354D		R376H	I405V										
	P407		P288A	A289V	Q295P	M313T		N354D	A366T	R376H	I405V	A413V			S444N	R445S					
	P478	A283V	P288A	A289V	Q295P	M313T		N354D		R376H				N443D	S444N						
	P506		P288A	A289V	Q295P				A366T												
	Assemblage B																				
	P344												D384N						K437Q		
	P387		G285E						K336R								E412G				
	P413			G290R	L316P	S318P	M329L	E335G			V363I			R399H	S401G	D403N		Q434K	K437Q	P442S	R445H
	P424			G290R										R399H				Q434K	K437Q	P442S	R445H
	P427	R281S								W347R			D384N								
	P458						M329L					A365I							K437Q		

Figure 3 Cartoon representation of homology models created for gFlHb from assemblage A and B. Heme binding globing domain is presented in purple, FAD-binding domain in green and C-terminal FAD- binding domain in cyan. Heme and FAD were fitted to the homology model and are presented by stick representation. Detected nsSNV-induced mutations to the amino acid sequences for individual isolates are presented in the table. Mutations possibly affecting directly to protein function by disturbing the heme or FAD binding are indicated with red in both cartoon models and in the table. The mutation marked as STOP-49Q introduces a stop codon into the sequence interrupting the protein synthesis.

culture could have an adapted metabolism of MTZ and detoxification processes of free radicals and O_2 and may represent a selected version of the isolates causing disease in humans. The isolates in the present study could therefore be biased with regard to gFlHb copy numbers or be the reason for the low nucleotide diversity and many different, but similar, alleles of *gFlHb*. With this caveat in mind, the alleles and copy number variation identified in the present study will nevertheless enable future studies and analysis of *gFlHb* in non-culturable isolates, as it provides a start of a respective database of gene variants.

Alleles of the gFlHb Gene

In the present study, several alleles of the *gFlHb* gene were identified and PacBio sequencing confirmed that the number of copies of *gFlHb* in each isolate is variable. As several copies of *gFlHb* are sometimes found on the same contig, we use the term allele rather than the term haplotype which has been used in previous studies of *Giardia* single-copy gene variants.[39,40,54] The phylogenetic trees in Figures 1 and 2 represent all the identified alleles in each isolate from sub-assemblages AII and assemblage B. Assemblage B isolates had less available samples and clones than AII, but had more unique alleles (41 vs 37). No sub-assemblage AII derived allele was found in assemblage B isolates, or vice versa, thus indicating no inter-assemblage recombination. However, it is difficult to define full tetraploid *gFlHb* genotypes with the number of clones available for each isolate, further complicated by the presence of copy number variation (CNV) which allows for more possible combinations of *gFlHb* alleles. Phylogenetic trees for each assemblage, on the other hand, are consistent with the occurrence of intra-assemblage recombination and genetic exchange although mixed isolate infections cannot completely be ruled out. Some alleles were common and found to be present in more than one isolate, while the same isolates also harbor several other, and different, alleles only occurring in one isolate (see tables S4 and S5). The high allelic diversity indicates that the overall diversity in the population may be much larger. As *gFlHb* is a variable-copy gene, the maximum number of alleles one tetraploid *Giardia* strain may harbor is four times the number of copies. In sub-assemblage AII isolates, higher copy numbers correlated with higher numbers of identified alleles but no breach to said rule was observed. In previous studies investigating haplotypes of single copy genes such as (*bg*, *gdh* and *mlh*, *tpi*), more than the expected maximum four haplotypes per paralog were encountered.[36–39] At present, we cannot rule out that this could be due to unidentified additional copies of the gene hidden in the current versions of de novo assembled genomes. Also, mixed infections consisting of multiple *Giardia* strains as demonstrated by previous studies, cannot be discounted completely (see above).[29,33,37–39,59] However, we favor the idea that supernumerary alleles reflect genome variations that have occurred during culture causing co-existence of daughter lineages as described by Choy et al.[31] Remarkable genetic variation due to selective pressure during axenization has been recently described also for other parasites such as *Leishmania donovani*.[60–62] In both assemblages, some of the identified alleles were not matching the identified PacBio copies. Minor alleles, representing 1:4 of the possible alleles in a tetraploid organism, are likely to be missed in consensus sequences as nucleotides of the major allele would be decisive. This, in addition to potential sequencing errors in combination with correction routines or assembly errors in the PacBio genomes might be the reason why not all copies are represented by the identified alleles.

Copy Number Variation

gFlHb is an important enzyme in the detoxification processes in *Giardia*, to eliminate the harmful effects of NO. One study showed that this was the sole enzyme that was upregulated in several different stress-exposures such as O_2 and H_2O_2, although there was no correlation between transcription responses to H_2O_2 and MTZ.[14] Our finding of a variable number of *gFlHb* gene copies may be important in relation to tackling the oxidative stress caused by MTZ treatment. A recent study by Müller et al, reported upregulated gFlHb protein levels in an MTZ and nitazoxanide resistant isolate.[20] One can hypothesize that it is beneficial for the parasite to have more *gFlHb* gene copies, as it may promote survival in an environment with higher O_2 levels, where activation of MTZ is lower, or in the oxidative stress condition induced by MTZ. Interestingly, a major antibiotic function of MTZ is to cause oxidative stress, and it is tempting to speculate whether gFlHb could be involved in neutralizing the harmful effects of the drug as *Giardia* lacks conventional antioxidant enzymes.[63] In a previous study, the nitro-drug resistant line C4 of *Giardia* was compared to its corresponding wild-type WBC6 isolate with respect to mRNA expression, O_2 consumption and functional assays.[66] It was found that the C4

line exhibited lower nitroreductase activity, while mRNA levels of gFlHb were not significantly different before and after drug exposure while reduction of nitro-drug activation was found to be lower.[66] In another study by Ansell et al 2017, strand-specific RNA analysis was carried out in three resistant laboratory lines.[67] gFlHb was found to be strongly induced in one of the lines (713-r), and the authors suggested that gFlHb could play the role of an alternative MTZ detoxification enzyme.[67] This may be assay dependent, but is more likely to be strain specific, especially as different *Giardia* isolates harbor a variable number of *gFlHb* copies. Taken together these findings indicate that different *Giardia* isolates could have different MTZ resistance formation strategies, ie, isolates with a higher *gFlHb* copy number may be more prone to use an active detoxifying MTZ resistance strategy, while other isolates may adopt a strategy of reduced MTZ activation.[66] For higher eukaryotes, such as the insects *Anopheles gambiae and Culex pipiens*, increased CNV has been linked to insecticide resistance.[64] Aneuploidy causing increased copy numbers of several genes in fungi has also been linked with drug resistance.[65] Although several copies of the *gFlHb* gene in some isolates were found, we must acknowledge that we do not know how the copies, and their various alleles would be expressed or regulated in an MTZ exposed *Giardia* isolate in need of more potent free radical neutralization. Further studies combining single-cell DNA and RNA sequencing are needed to elucidate the effect of having multiple gFlHb copies on the ability to overcome MTZ toxicity or improve tolerance of oxidative stress. In the current reference genome of sub-assemblage AII there are two shorter paralogs of *gFlHb* (DHA2_152971 and DHA2_153995) that could potentially result in falsely higher coverage for the *gFlHb* gene. There is no reason to believe that these shorter fragments contribute to the higher coverage found in the Illumina sequence data, and this was shown in a previous study.[42] Furthermore, the PacBio consensus sequences were used to determine the CNV in the *gFlHb* gene, and the shorter paralogs could not affect the results, as full-length genes were investigated.

Putative Effect of SNV-Induced Mutations on gFlHb Proteins

The *gFlHb* gene in both sub-assemblage AII and assemblage B was found to have a high number of nsSNVs that would alter the protein's amino acid sequence. Some previous studies have also reported high numbers of nsSNPs in housekeeping genes, especially for assemblage B isolates, but these SNVs were commonly only causing conservative aa changes, unlikely to affect protein function or structure.[37,40] The crystal structure of gFlHb protein is not yet known, and structural mapping of the location of the mutated amino acids was consequently based on homology models. The premature stop codon found in two alleles in sub-assemblage AII clones will cause a truncated protein without its normal function. Interpretation of the mutations that seem to be located on the surface of the protein is more challenging in terms of predicting the effect. Altering the surface charge could potentially have an effect on inter-molecular interactions. It is clear that further research is needed to investigate whether the aa changes can potentially affect the function of the respective protein variants. Specifically locating the exact site for NAD binding would help us investigate the effect of the mutations in the domain 3, the NAD binding domain. However, some of the mutations described might indeed affect heme and FAD binding, therefore altering protein function.

Limitations

For some of the isolates, few clones were obtained, and rare alleles may have been missed in some isolates. For isolate P064, one of the two *gFlHb* copies found in the PacBio consensus sequence, was not found among the clones obtained in this study, showing that several more alleles may exist for this isolate. One other explanation for missing alleles, is that the PCR primers targeting *gFlHb* in the initial amplification may not have matched all alleles and did not bind to the template DNA and amplify these. However, PacBio consensus copies may not be identical to the clones due to sequencing errors, or due to PacBio sequence correction routines. The study population in this study is rather small and future studies of non-cultured isolates are likely to extend information about the number of copies and genetic variation of the gene. The SNVs analyzed in the present study were all limited to the CDS of the *gFlHb* of assemblage A and B isolates. There have been previous studies linking SNPs in intergenic areas to MTZ resistance in *T. vaginalis*, specifically 12 SNPs, that potentially could be markers of resistance.[68] SNVs in the up- or downstream regions of the *gFlHb* CDS could affect the transcription and therefore be of potential interest for investigating MTZ susceptibility and oxidative stress responses markers in future studies.

Conclusion

In this study, we show evidence that the *gFlHb* gene in *Giardia* sub-assemblage AII and assemblage B, is a variable copy number gene with a high allelic diversity. The high genetic diversity seems to be due to both high copy number variation, many similar but unique, alleles per isolate, and a high number of SNV positions among both *Giardia* sub-assemblage AII and assemblage B isolates. Relatively high percentages of nonsynonymous SNVs were identified in both assemblages and some changes could potentially affect the protein function. The variable copy number nature of the *gFlHb* gene may allow some *Giardia* strains to better adapt to nitrosative and oxidative stress and could thereby potentially play a role in MTZ susceptibility.

Abbreviations

CDS, coding region; gFlHb, *Giardia lamblia* Flavohemoprotein; HC, High Confidence; LC, Low confidence; MTZ, Metronidazole; ns, nonsynonymous; SNV, single nucleotide polymorphism; s, synonymous.

Data Sharing Statement

The sequences for the identified gFlHb alleles are available in NCBI genbank with accession numbers: MT713149-MT713229, MT713231-MT713243, MT713245-MT713263, MT713267, MT713269.

Ethics Statement

No patient data was collected and used in the present study. Any link between individual parasite data and patient information had been removed before WGS of trophozoites and cloning experiments were carried out.

Acknowledgments

The personnel at the Genomics Core Facility at Oslo University Hospital (oslo.genomics.no) should be acknowledged due to their work with the Illumina whole-genome sequencing of *Giardia* DNA. We would also like to thank the personnel at the Department of Clinical Science, University of Bergen, for providing technical assistance with the sequencing and assembly of genome data. We acknowledge Prof Dr. Ralf Ignatius from Institute of Tropical Medicine and International Health, Charité, Berlin, Germany for kindly providing biological samples. We would also like to thank Simone Caccio for support in the de-novo PacBio sequencing of the *Giardia* isolates. We would also like to thank Petra Gosten-Heinrich, Robert Koch-Institute Berlin for providing excellent technical assistance during culturing experiments and sequencing clones of gFlHb.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study has been funded by Norwegian Surveillance System for Antimicrobial Drug Resistance (NORM), Centre for Pharmacy, University of Bergen, Helse-Vest (grant number 912245), and Norwegian PhD School of Pharmacy (NFIF). All of the experiments, data analysis and evaluation of the results in the present study were performed independently by the authors, without any interference from any of the funding institutions.

Disclosure

Some of the data in this paper has been published as an abstract and was presented at the VIIth International Giardia and Cryptosporidium Conference (IGCC), 2019, as a conference talk with interim findings. The abstract was published in the "Conference abstracts on USB Key" and is available from [https://en.rouentourisme.com/wp-content/uploads/2019/06/Programme-d%C3%A9finitif-modifi%C3%A9-200619.pdf].

The authors report no conflicts of interest for this work.

References

1. Lane S, Lloyd D. Current trends in research into the waterborne parasite Giardia. *Crit Rev Microbiol*. 2002;28(2):123–147. doi:10.1080/1040-840291046713
2. Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. *Lancet*. 2002;359 (9306):564–571. doi:10.1016/S0140-6736(02)07744-9
3. Rogawski ET, Platts-Mills JA, Seidman JC, et al. Early antibiotic exposure in low-resource settings is associated with increased weight in the first two years of life. *J Pediatr Gastroenterol Nutr*. 2017;65 (3):350–356. doi:10.1097/MPG.0000000000001640
4. Craun GF, Calderon RL, Craun MF. Outbreaks associated with recreational water in the United States. *Int J Environ Health Res*. 2005;15 (4):243–262. doi:10.1080/09603120500155716
5. Farthing MJ. Giardiasis. *Gastroenterol Clin North Am*. 1996;25 (3):493–515. doi:10.1016/S0889-8553(05)70260-0

6. Gardner TB, Hill DR. Treatment of giardiasis. *Clin Microbiol Rev*. 2001;14(1):114–128.
7. Munoz Gutierrez J, Aldasoro E, Requena A, et al. Refractory giardiasis in Spanish travellers. *Travel Med Infect Dis*. 2013;11(2):126–129. doi:10.1016/j.tmaid.2012.10.004
8. Requena-Mendez A, Goni P, Rubio E, et al. The use of quinacrine in nitroimidazole-resistant giardia duodenalis: an old drug for an emerging problem. *J Infect Dis*. 2017;215(6):946–953. doi:10.1093/infdis/jix066
9. Nabarro LE, Lever RA, Armstrong M, Chiodini PL. Increased incidence of nitroimidazole-refractory giardiasis at the hospital for tropical diseases, London: 2008–2013. *Clin Microbiol Infect*. 2015;21(8):791–796. doi:10.1016/j.cmi.2015.04.019
10. Dingsdag SA, Hunter N. Metronidazole: an update on metabolism, structure-cytotoxicity and resistance mechanisms. *J Antimicrob Chemother*. 2018;73(2):265–279. doi:10.1093/jac/dkx351
11. Leitsch D. A review on metronidazole: an old warhorse in antimicrobial chemotherapy. *Parasitology*. 2017;1–12.
12. Church DL, Laishley EJ. Reduction of metronidazole by hydrogenase from clostridia. *Anaerobe*. 1995;1(2):81–92. doi:10.1006/anae.1995.1002
13. Mastronicola D, Falabella M, Forte E, Testa F, Sarti P, Giuffre A. Antioxidant defence systems in the protozoan pathogen Giardia intestinalis. *Mol Biochem Parasitol*. 2016;206(1–2):56–66. doi:10.1016/j.molbiopara.2015.12.002
14. Ma'ayeh SY, Knorr L, Svard SG. Transcriptional profiling of Giardia intestinalis in response to oxidative stress. *Int J Parasitol*. 2015;45(14):925–938. doi:10.1016/j.ijpara.2015.07.005
15. Mastronicola D, Testa F, Forte E, et al. Flavohemoglobin and nitric oxide detoxification in the human protozoan parasite Giardia intestinalis. *Biochem Biophys Res Commun*. 2010;399(4):654–658. doi:10.1016/j.bbrc.2010.07.137
16. Gardner PR, Gardner AM, Martin LA, Salzman AL. Nitric oxide dioxygenase: an enzymic function for flavohemoglobin. *Proc Natl Acad Sci U S A*. 1998;95(18):10378–10383.
17. Gardner PR. Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases. *J Inorg Biochem*. 2005;99(1):247–266. doi:10.1016/j.jinorgbio.2004.10.003
18. Rafferty S, Luu B, March RE, Yee J. Giardia lamblia encodes a functional flavohemoglobin. *Biochem Biophys Res Commun*. 2010;399(3):347–351. doi:10.1016/j.bbrc.2010.07.073
19. Ansell BR, McConville MJ, Baker L, et al. Divergent transcriptional responses to physiological and xenobiotic stress in giardia duodenalis. *Antimicrob Agents Chemother*. 2016;60(10):6034–6045. doi:10.1128/AAC.00977-16
20. Müller J, Braga S, Heller M, Müller N. Resistance formation to nitro drugs in Giardia lamblia: no common markers identified by comparative proteomics. *Int J Parasitol Drugs Drug Resist*. 2019;9:112–119. doi:10.1016/j.ijpddr.2019.03.002
21. Rasoloson D, Tomkova E, Cammack R, Kulda J, Tachezy J. Metronidazole-resistant strains of Trichomonas vaginalis display increased susceptibility to oxygen. *Parasitology*. 2001;123(\(Pt1)):45–56. doi:10.1017/S0031182001008022
22. Ansell BR, McConville MJ, Ma'ayeh SY, et al. Drug resistance in Giardia duodenalis. *Biotechnol Adv*. 2015;33(6 Pt 1):888–901. doi:10.1016/j.biotechadv.2015.04.009
23. Forrester MT, Foster MW. Protection from nitrosative stress: a central role for microbial flavohemoglobin. *Free Radic Biol Med*. 2012;52(9):1620–1633. doi:10.1016/j.freeradbiomed.2012.01.028
24. Helmick RA, Fletcher AE, Gardner AM, et al. Imidazole antibiotics inhibit the nitric oxide dioxygenase function of microbial flavohemoglobin. *Antimicrob Agents Chemother*. 2005;49(5):1837–1843. doi:10.1128/AAC.49.5.1837-1843.2005
25. Yu LZ, Birky CW Jr, Adam RD. The two nuclei of Giardia each have complete copies of the genome and are partitioned equationally at cytokinesis. *Eukaryotic Cell*. 2002;1(2):191–199. doi:10.1128/EC.1.2.191-199.2002
26. Bernander R, Palm JE, Svard SG. Genome ploidy in different stages of the Giardia lamblia life cycle. *Cell Microbiol*. 2001;3(1):55–62. doi:10.1046/j.1462-5822.2001.00094.x
27. Kabnick KS, Peattie DA. In situ analyses reveal that the two nuclei of Giardia lamblia are equivalent. *J Cell Sci*. 1990;95(Pt 3):353–360.
28. Tumova P, Uzlikova M, Jurczyk T, Nohynkova E. Constitutive aneuploidy and genomic instability in the single-celled eukaryote Giardia intestinalis. *MicrobiologyOpen*. 2016;5(4):560–574. doi:10.1002/mbo3.351
29. Ankarklev J, Svard SG, Lebbad M. Allelic sequence heterozygosity in single Giardia parasites. *BMC Microbiol*. 2012;12:65. doi:10.1186/1471-2180-12-65
30. Franzen O, Jerlstrom-Hultqvist J, Castro E, et al. Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog*. 2009;5(8):e1000560.
31. Choy SH, Mahdy MA, Al-Mekhlafi HM, Low VL, Surin J. Population expansion and gene flow in Giardia duodenalis as revealed by triosephosphate isomerase gene. *Parasit Vectors*. 2015;8:454. doi:10.1186/s13071-015-1084-y
32. Caccio SM, Ryan U. Molecular epidemiology of giardiasis. *Mol Biochem Parasitol*. 2008;160(2):75–80. doi:10.1016/j.molbiopara.2008.04.006
33. Hussein AI, Yamaguchi T, Nakamoto K, Iseki M, Tokoro M. Multiple-subgenotype infections of Giardia intestinalis detected in Palestinian clinical cases using a subcloning approach. *Parasitol Int*. 2009;58(3):258–262.
34. Lalle M, Pozio E, Capelli G, Bruschi F, Crotti D, Cacciò SM. Genetic heterogeneity at the beta-giardin locus among human and animal isolates of Giardiaduodenalis and identification of potentially zoonotic subgenotypes. *Int J Parasitol*. 2005;35(2):207–213. doi:10.1016/j.ijpara.2004.10.022
35. Gabin-Garcia LB, Bartolome C, Abal-Fabeiro JL, Mendez S, Llovo J, Maside X. Strong genetic structure revealed by multilocus patterns of variation in Giardia duodenalis isolates of patients from Galicia (NW-Iberian Peninsula). *Infect Genet Evol*. 2017;48:131–141. doi:10.1016/j.meegid.2016.12.014
36. Kosuwin R, Putaporntip C, Pattanawong U, Jongwutiwes S. Clonal diversity in Giardia duodenalis isolates from Thailand: evidences for intragenic recombination and purifying selection at the beta giardin locus. *Gene*. 2010;449(1–2):1–8. doi:10.1016/j.gene.2009.09.010
37. Siripattanapipong S, Leelayoova S, Mungthin M, et al. Clonal diversity of the glutamate dehydrogenase gene in Giardia duodenalis from Thai isolates: evidence of genetic exchange or mixed infections? *BMC Microbiol*. 2011;11:206. doi:10.1186/1471-2180-11-206
38. Lasek-Nesselquist E, Welch DM, Thompson RC, Steuart RF, Sogin ML. Genetic exchange within and between assemblages of Giardia duodenalis. *J Eukaryot Microbiol*. 2009;56(6):504–518. doi:10.1111/j.1550-7408.2009.00443.x
39. Mizuno T, Matey EJ, Bi X, Songok EM, Ichimura H, Tokoro M. Extremely diversified haplotypes observed among assemblage B population of Giardia intestinalis in Kenya. *Parasitol Int*. 2020;75:102038. doi:10.1016/j.parint.2019.102038
40. Lecova L, Tumova P, Nohynkova E. Clone-based haplotyping of Giardia intestinalis assemblage B human isolates. *Parasitol Res*. 2019;118(1):355–361. doi:10.1007/s00436-018-6161-7
41. Aguiar JM, Silva SO, Santos VA, et al. Evidence of heterozygosity and recombinant alleles in single cysts of Giardia duodenalis. *Revista Brasileira De Parasitologia Veterinaria*. 2016;25(2):187–195. doi:10.1590/S1984-29612016031
42. Saghaug CS, Klotz C, Kallio JP, et al. Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical Giardia lamblia assemblage A and B isolates. *Infect Drug Resist*. 2019;12:1221–1235. doi:10.2147/IDR.S177997
43. Keister DB. Axenic culture of Giardia lamblia in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg*. 1983;77(4):487–488. doi:10.1016/0035-9203(83)90120-7

44. Aurrecoechea C, Brestelli J, Brunk BP, et al. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis. *Nucleic Acids Res*. 2009;37(Databaseissue):D526–530. doi:10.1093/nar/gkn631
45. Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. *Gene*. 1990;96(1):23–28. doi:10.1016/0378-1119(90)90336-P
46. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299–3302. doi:10.1093/molbev/msx248
47. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003;19(18):2496–2497. doi:10.1093/bioinformatics/btg359
48. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845–858. doi:10.1038/nprot.2015.053
49. Ilari A, Bonamore A, Farina A, Johnson KA, Boffi A. The X-ray structure of ferric Escherichia coli flavohemoglobin reveals an unexpected geometry of the distal heme pocket. *J Biol Chem*. 2002;277(26):23725–23732. doi:10.1074/jbc.M202228200
50. Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974;185(4154):862–864. doi:10.1126/science.185.4154.862
51. Ermler U, Siddiqui RA, Cramm R, Friedrich B. Crystal structure of the flavohemoglobin from Alcaligenes eutrophus at 1.75 A resolution. *EMBO J*. 1995;14(24):6067–6077. doi:10.1002/j.1460-2075.1995.tb00297.x
52. El Hammi E, Warkentin E, Demmer U, Marzouki NM, Ermler U, Baciou L. Active site analysis of yeast flavohemoglobin based on its structure with a small ligand or econazole. *FEBS J*. 2012;279(24):4565–4575. doi:10.1111/febs.12043
53. Potapov V, Ong JL. Examining sources of error in PCR by single-molecule sequencing. *PLoS One*. 2017;12(1):e0169774. doi:10.1371/journal.pone.0169774
54. Teodorovic S, Braverman JM, Elmendorf HG. Unusually low levels of genetic variation among Giardia lamblia isolates. *Eukaryotic Cell*. 2007;6(8):1421–1430. doi:10.1128/EC.00138-07
55. Adam RD, Dahlstrom EW, Martens CA, et al. Genome sequencing of Giardia lamblia genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol Evol*. 2013;5(12):2498–2511. doi:10.1093/gbe/evt197
56. Morrison HG, McArthur AG, Gillin FD, et al. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science*. 2007;317(5846):1921–1926. doi:10.1126/science.1143837
57. Benere E, Geurden T, Robertson L, Van Assche T, Cos P, Maes L. Infectivity of Giardia duodenalis Assemblages A and E for the gerbil and axenisation of duodenal trophozoites. *Parasitol Int*. 2010;59(4):634–637. doi:10.1016/j.parint.2010.08.001
58. Binz N, Thompson RC, Lymbery AJ, Hobbs RP. Comparative studies on the growth dynamics of two genetically distinct isolates of Giardia duodenalis in vitro. *Int J Parasitol*. 1992;22(2):195–202. doi:10.1016/0020-7519(92)90101-P
59. Caccio SM, Sprong H. Giardia duodenalis: genetic recombination and its implications for taxonomy and molecular epidemiology. *Exp Parasitol*. 2010;124(1):107–112. doi:10.1016/j.exppara.2009.02.007
60. Bussotti G, Gouzelou E, Côrtes Boité M, et al. Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *mBio*. 2018;9:6. doi:10.1128/mBio.01399-18
61. Imamura H, Downing T, Van den Broeck F, et al. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *Elife*. 2016;5.
62. Domagalska MA, Imamura H, Sanders M, et al. Genomes of Leishmania parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. *PLoS Negl Trop Dis*. 2019;13(12):e0007900. doi:10.1371/journal.pntd.0007900
63. Argüello-García R, Leitsch D, Skinner-Adams T, Ortega-Pierres MG. Drug resistance in giardia: mechanisms and alternative treatments for giardiasis. *Adv Parasitol*. 2020;107:201–282.
64. Weetman D, Djogbenou LS, Lucas E. Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? *Current Opinion Insect Sci*. 2018;27:82–88. doi:10.1016/j.cois.2018.04.005
65. Wertheimer NB, Stone N, Berman J. Ploidy dynamics and evolvability in fungi. *Philos Trans R Soc Lond B Biol Sci*. 2016;371:1709.
66. Müller J, Hemphill A, Müller N. Physiological aspects of nitro drug resistance in Giardia lamblia. *Int J Parasitol Drugs Drug Resist*. 2018;8(2):271–277. doi:10.1016/j.ijpddr.2018.04.008
67. Ansell BR, Baker L, Emery SJ, et al. Transcriptomics indicates active and passive metronidazole resistance mechanisms in three seminal giardia lines. *Front Microbiol*. 2017;8:398. doi:10.3389/fmicb.2017.00398
68. Bradic M, Warring SD, Tooley GE, et al. Genetic indicators of drug resistance in the highly repetitive genome of trichomonas vaginalis. *Genome Biol Evol*. 2017;9(6):1658–1672. doi:10.1093/gbe/evx110

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230869185 (print)
9788230869932 (PDF)