UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A HYBRID APPROACH TO ASSESSING *GIARDIA INTESTINALIS* GENETIC DIVERSITY IN TWO TRADITIONAL HUMAN POPULATIONS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF ARTS

By

ABIGAIL GAMBLE
Norman, Oklahoma
2020

A HYBRID APPROACH TO ASSESSING *GIARDIA INTESTINALIS* GENETIC DIVERSITY

IN TWO TRADITIONAL HUMAN POPULATIONS

A THESIS APPROVED FOR THE
DEPARTMENT OF ANTHROPOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Cecil Lewis, Chair

Dr. Courtney Hofman

Dr. Tassie Hirschfeld

© Copyright by ABIGAIL GAMBLE 2020
All Rights Reserved.

TABLE OF CONTENTS

Abstract……………………………………………………………………………………...v

Chapter 1: Introduction………………………………………………………………………1

Chapter 2: Background…………………………………………………………………...21

Chapter 3: Materials and Methods…………………………………………………………34

Chapter 4: Results and Discussion…………………………………………………………44

Chapter 5: References Cited………………………………………………………………...62

Supplemental Data…………………………………………………………………………...69

ABSTRACT

The protozoan parasite *Giardia intestinalis* is traditionally viewed as a harmful agent of diarrheal disease in humans, and is the target of prevention and treatment efforts around the world. However, most cases of *G. intestinalis* are asymptomatic, and prevalence rates of greater than 80% in several human populations without adverse effects raise the possibility that this microeukaryote could be a commensal member of the human gut microbiome. The *Giardia intestinalis* species has been divided into eight genetic assemblages, and some research suggests particular assemblages may be more pathogenic than others. In addition, research into the presence and assemblage diversity of *G. intestinalis* in traditional and hunter-gatherer societies can provide insight on the conditions under which this microeukaryote may be commensal in humans, as the gut microbiomes of these individuals have not been affected by industrialized practices that have historically depleted microbial diversity in the human gut.

This thesis addresses this relatively underexplored area of research by attempting to determine the *G. intestinalis* assemblages present in asymptomatic hunter-gatherers from Peru and individuals from a traditional community of Burkina Faso. A hybrid approach consisting of targeted amplification of the *glutamate dehydrogenase* (*gdh*) gene and next-generation sequencing using the Illumina MiSeq was developed, and the presence of *G. intestinalis* was confirmed in the hunter-gatherer population of Peru. While poor data quality and probable database bias precluded a confident declaration of the assemblage, we may have identified a novel subassemblage of *Giardia intestinalis* in one individual from the Peruvian Amazon. This suggests that hunter-gatherer populations may harbor a greater diversity of assemblages in a commensal state.

CHAPTER 1

INTRODUCTION

This study aims to fill the gap in research on natural *Giardia intestinalis* assemblage variation in humans by exploring the prevalence and genetic diversity of *G. intestinalis* in a hunter-gatherer population of Peru and a traditional community in the West African country of Burkina Faso. To our knowledge, only five studies in the last decade have assessed *G. intestinalis* assemblage diversity among traditional peoples, and this is the first study to assess the genetic diversity of *G. intestinalis* in hunter-gatherers of South America. High prevalence of asymptomatic *G. intestinalis* has been previously found in Peru (85% in a periurban community south of Lima [1] and 29% in Amazonian hunter-gatherers [2]) and Burkina Faso (31.6% in the Centre-Ouest region [3]), suggesting that *Giardia*-host relationships are complex, and not exclusively disease bearing. From these prior studies, we anticipate a high prevalence of *Giardia* in the two populations included in this study. Furthermore, while *G. intestinalis* assemblages A and B are most commonly found in humans, zoonotic transmission of other assemblages has been observed on multiple occasions [4,5,6,7,8]. Thus, it is possible the non-industrialized lifestyles of these traditional populations, which frequently involve close interactions with a variety of animals, may result in a diverse range of *G. intestinalis*. We aim to assess such genetic diversity, as well as test the hypothesis that natural variation in *G. intestinalis* human assemblages may be obscured by the bias in data from urban settings.

What is *Giardia*?

Taxonomy

Giardia is a flagellated protozoan parasite [9,10] and one of the most common parasites infecting humans and animals around the world today [11]. There are currently six accepted species of the genus *Giardia*, each host-specific to a wide range of animals: *G. agilis* (amphibians), *G. ardeae, G. psittaci* (birds), *G. microti* (voles), *G. muris* (rodents), and *G. intestinalis* (mammals). These species were distinguished based on microscopic morphology and molecular variations identified from DNA sequencing [9,11]. A seventh species, *G. varani* (lizards), has been proposed, but has not yet been confirmed by molecular biological characterizations [11].

Giardia intestinalis, which also goes by *G. lamblia* and *G. duodenalis*, has been subdivided further into eight genetic assemblages, or strains, each of which is also primarily, though not always, host-specific. These assemblages and their typical hosts are presented in Table 1. The phylogenetic relationships among these assemblages at the *glutamate dehydrogenase* (*gdh*) locus were assessed by Feng and Xiao (2011) using the Neighbor-Joining method and distance calculated by the Kimura two-parameter model. With a collection of 33 *G. intestinalis gdh* isolates obtained from the National Center for Biotechnology Information (NCBI) GenBank database, we recreated these phylogenetic relationships using the same tree-building conditions as Feng and Xiao (2011) (Figure 1).

As is evident from the table, assemblages A and B are the most versatile, infecting humans but also a wide range of other mammals. Allozyme analysis revealed additional sub-structure within assemblage A into subassemblages AI and AII, which was later confirmed by phylogenetic analysis of sequences at the *gdh* locus. A third subgroup within assemblage A, AIII, was also identified that has significant sequence differences from subassemblages AI and AII at all loci examined so far [11]. It is also the rarest subassemblage of A to be found in

humans. Allozyme electrophoretic studies determined two subassemblages within assemblage B, subgroups BIII and BIV, however these groupings are not supported by DNA sequence analyses and thus subgrouping within assemblage B remains unclear [11].

Assemblage	Typical hosts
A	Humans, nonhuman primates, domestic ruminants, alpacas, pigs, horses, domestic and wild canines, cats, ferrets, rodents, marsupials, other mammals
B	Humans, nonhuman primates, cattle, dogs, horses, rabbits, beavers, muskrats
C	Domestic and wild canines
D	Domestic and wild canines
E	Domestic ruminants, pigs
F	Cats
G	Mice, rats
H	Seals

Table 1. *G. intestinalis* assemblages and corresponding hosts [11]

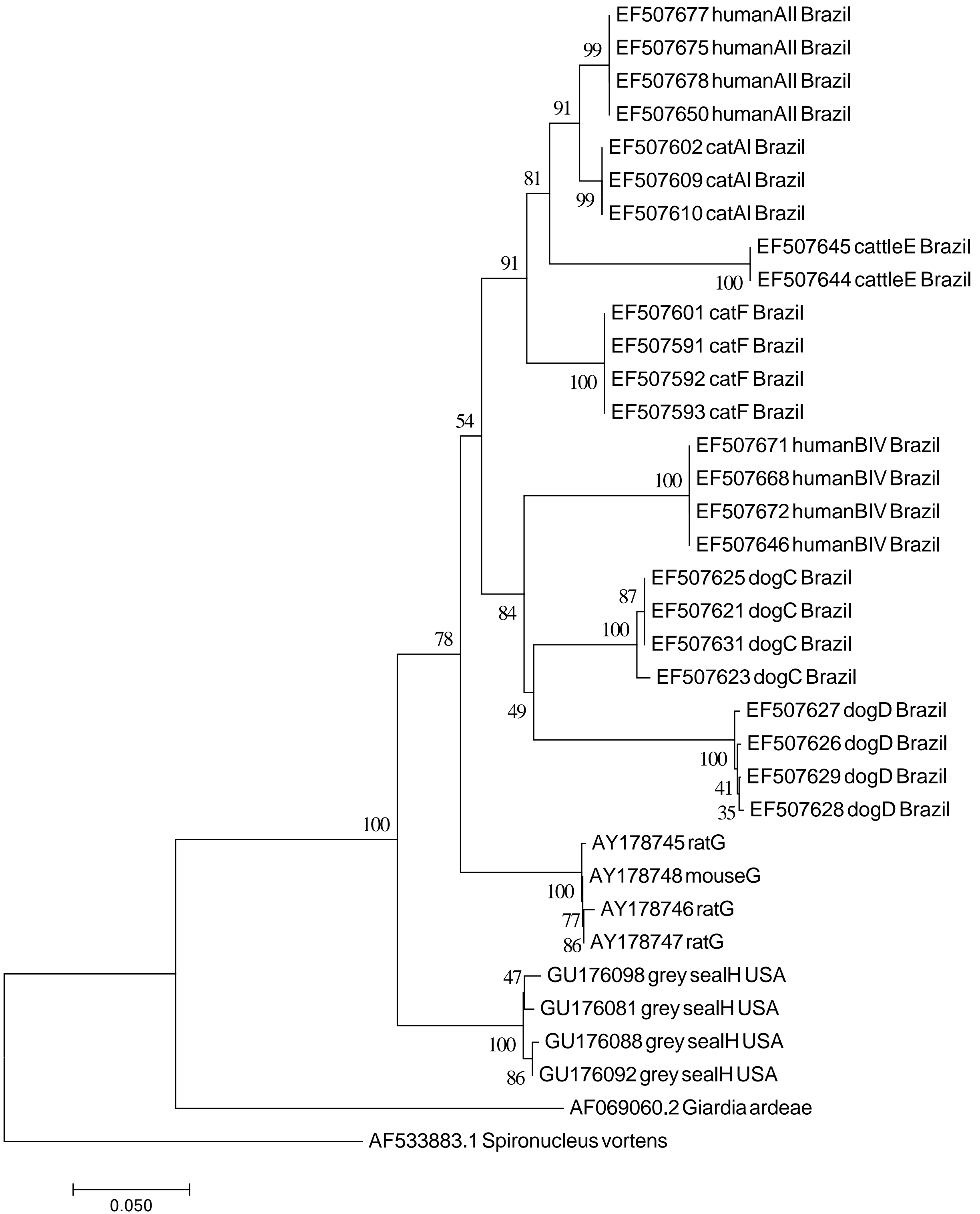

Figure 1. Phylogenetic relationships among assemblages of *G. intestinalis* at the *gdh* locus. The Neighbor-Joining method and Kimura two-parameter model were used, with 500 bootstrap replicates. The scale indicates substitutions per site.

Biology and phylogeny

While *Giardia* is a eukaryote, it is a unique one, thought to be one of the most primitive extant eukaryotic organisms and sharing many characteristics with anaerobic prokaryotes [10]. For example, *Giardia* lacks a mitochondria, peroxisomes, and a traditional Golgi apparatus [9,10]. It has been suggested that *Giardia spp.* are pre-mitochondriate organisms, but there is some debate surrounding this idea due to *Giardia intestinalis* possessing genes considered to be of mitochondrial origin [12,13]. Trophozoites (the second form *Giardia* takes in its life cycle) of *G. intestinalis* have two identical nuclei and are polyploid, with at least four and possibly eight or more copies of each of five chromosomes per organism [10]. The genome of *G. intestinalis* is about 11.7 million base pairs (bp) long with 6,470 predicted protein-coding genes [14].

The flagella and two nuclei of *Giardia*, among other biological features, place the organism into the taxonomic order Diplomonadida. Early phylogenetic studies of the Diplomonadida using conserved loci determined that *Giardia* belongs to an early branching lineage [15]. While this finding led many to think *Giardia* was one of the more primitive Diplomonadida, a phylogenetic analysis of 23 morphological characters suggests that it is actually one of the most highly adapted [16]. Furthermore, molecular studies have shown that *Giardia*'s lack of mitochondria is not a primitive feature, reflective of divergence from an ancestral eukaryote before the acquisition of the organelle, but rather is a secondary evolutionary loss [17]. This aligns with the finding of genes considered to be of mitochondrial origin in *G. intestinalis*. However, in the majority of phylogenies in which eukaryotes are rooted using prokaryotic outgroups, diplomonads and their relatives in general branch out early in the eukaryotic tree [18].

Molecular phylogenies consistently divide diplomonads into two major clades, Hexamitinae and Giardiinae, both of which fall under the taxonomic family Hexamitidae [18]. In addition, the diplomonad *Octomitus sp.* has been confirmed as a sister lineage to *Giardia* within Giardiinae [19]. Phylogenetic analysis based on SSU rRNA genes has also confirmed Giardiinae as a monophyletic group, whereas Hexamitinae form a clade with enteromonads (close relatives to diplomonads)[18].

Lifecycle and transmission

While both cysts and trophozoites can be found in the feces of infected organisms, *Giardia* cysts are responsible for transmission. The cysts are moderately chlorine tolerant and environmentally robust [20], able to survive for months without a host in surface water and soil [9]. Infection occurs when cysts are ingested through contaminated water, food, or by the fecal-oral route, and as few as ten cysts may be sufficient to cause giardiasis infection. Ingestion of more than twenty-five cysts results in a 100% infection rate [21]. Excystation occurs in the small intestine, where each cyst releases two trophozoites, which then multiply by longitudinal binary fission. In the small bowel, the trophozoites are either free roaming or attach to the mucosa by a ventral sucking disk. Encystation occurs as the trophozoites move toward the colon, and ultimately cysts are excreted by the host [22]. Individuals can shed 1×10^8 to 1×10^9 cysts in their stool each day for several months [20]. Upon excretion, *Giardia* cysts are immediately able to infect a new host [9].

The common mechanisms of *Giardia* transmission are person to person, animal to animal, zoonotic, waterborne through contaminated drinking or recreational water, and foodborne [9]. Waterborne transmission is likely the most common. *Giardia* has been found in as many as 80% of raw water supplies from lakes, streams, and ponds [21] and more than 130

waterborne giardiasis outbreaks have been reported worldwide from the early 1900s to 2004 [9,11]. Furthermore, deficiencies in the drinking water treatment process are among the most frequently reported reasons for giardiasis outbreaks [9]. Farm animals excreting high numbers of *Giardia intestinalis* cysts around water supplies can be a major source of water contamination. The *G. intestinalis* assemblages that most commonly infect humans, A and B, are often reported for livestock species, such as goats, sheep, cattle, and pigs, around the world [9]. Additionally, some *G. intestinalis* outbreaks in North America and Spain have been linked to contamination of water by cysts excreted from wild animals, such as muskrats, beaver, and wild otter. Foods frequently consumed raw like fruit, vegetables, and shellfish, pose the greatest risk of infecting humans with *G. intestinalis* cysts [9]. Several studies have shown that *G. intestinalis* cysts can be present in oysters [23,24] and mussels [25], and others have linked food handlers or produce for sale to the transmission of *G. intestinalis* [26,27].

Epidemiology

Over 200 million people around the world are estimated to have acute or chronic giardiasis [10]. According to the Centers for Disease Control and Prevention (CDC), the illness infects nearly 2% of adults and 6% to 8% of children in developed countries and nearly 33% of people in developing countries, though rates of giardiasis approaching 90% in endemic areas have been reported [28]. The global prevalence of giardiasis and of the *Giardia intestinalis* parasite itself in humans varies greatly because the disease is not always reported, diagnostic methods used differ in sensitivity and many people are asymptomatic, especially in endemic areas. In the United States, *G. intestinalis* infection is the most common intestinal parasitic disease affecting humans, according to the CDC. In 2017, there were a total of 15,214 cases in the United States, with the incidence rate per 100,000 population ranging from 1.5 in Nevada to

12.2 in Alaska. By region, the Northwest had the highest incidence rate at 9.2 cases per 100,000 population and highest percentage of cases at 25.2%. The lowest incidence rate was in the South at 4.4 cases per 100,000 population. In general, incidence rates were consistently higher in the northern states. Demographically, more males than females had giardiasis (61.6% compared to 38.1%, respectively) and incidence rates were highest among those aged 1-4 years (11.4), 25-29 years (6.9) and 55-59 years (6.7) [29].

As in the United States, giardiasis is the most commonly reported food- and waterborne parasitic disease in the European Union (EU)/European Economic Area (EEA). The European Center for Disease Prevention and Control's (ECDC) most recent Annual Epidemiological Report for giardiasis stated that in 2017, 19,437 confirmed giardiasis cases were reported with an incidence rate of 5.5 cases per 100,000 population. The highest incidence rates were reported in Belgium (17.6), Estonia (12.2), and Sweden (11.4), and the highest number of confirmed cases was reported by the United Kingdom at 5,225 cases, followed by Germany at 3,329 cases. Together, the U.K. and Germany accounted for 44% of all confirmed giardiasis cases in the EU/EEA for 2017. The majority (60.1%) of cases in the EU/EEA were domestically acquired except in three Nordic countries, where 71%-83% of cases were travel-associated [30].

In general, the number of confirmed giardiasis cases in Europe steadily increased from 2013-2017. However, the ECDC notes that likely there is underreporting of cases throughout Eastern Europe, and one-fourth of EU Member States do not have surveillance systems for giardiasis. In congruence with the data from the United States, there were more cases in males (56%) than in females (44%) and the highest incidence rate per 100,000 population was detected in the age group 0-4 years, which accounted for 18% of cases with information on age. The *G. intestinalis* incidence rate decreased with age and was lowest in people aged 65 and older [30].

The prevalence of the *Giardia intestinalis* parasite is often quite high in developing countries. Western Nepal has reported one of the highest rates in Asia, at 73.4% of school-going children [31], and an even higher prevalence was found in Peruvian children (85%), notably without adverse effects [1]. Elsewhere in Asia, *G. intestinalis* prevalence has been 37.7% (Thailand), 24.9% (Malaysia), and 12.7% symptomatic and 18% asymptomatic (Bangladesh) [31]. Based on 33 studies published between 2002 and 2007, it has been shown that *G. intestinalis* prevalence in Asia is higher in urban than in rural areas, among poor communities, and slightly higher in males than in females. Multiple infections with other parasites were also frequently found [31].

There have been a limited number of population-based studies on *G. intestinalis* in Latin America, but there is nevertheless some data that reflects its pervasiveness. For instance, one study in Argentina detected the parasite in 33.3% of the 303 participating indigenous children [32] and another identified *G. intestinalis* in 23.8% of 366 people from Rio de Janeiro, Brazil [33]. Additionally, a *G. intestinalis* frequency of 64.8% was found in a population of children from Colombia [34].

High prevalence rates are consistently found across Africa. Out of thirty-one countries that reported the presence of *G. intestinalis*, nine had a prevalence of 40% or greater, including 41.7% in Algerian individuals, 56% in village children of Guinea-Bissau, 41.3% in individuals from Kenya, 53.4% in individuals of all ages and 62.2% in children under five from Tanzania, and 40.7% in individuals from villages in Uganda [35].

The immune status of the host has a major impact on the severity of giardiasis in general. Thus, the HIV/AIDS epidemic in African countries greatly contributes to an increased prevalence of giardiasis [35]. Other factors associated with a higher prevalence of *G. intestinalis*

and giardiasis in Africa include contact with animals and manure, residing in villages versus cities, drinking underground or tap water, and eating unwashed/raw fruit. Malnutrition is also an important risk factor for diarrhea and prolonged diarrhea caused by *G. intestinalis*, with this impact mainly affecting children under five [35]. While the prevalence of *G. intestinalis* has been found to be higher in men in many studies around the world, the opposite has been true in Ethiopia. An analysis of 393 stool samples from children attending eight rural schools in the Bahir Dar district of Ethiopia found that females were more frequently infected with *G. intestinalis* than males (61.1% vs. 49.5%, respectively), with the difference being statistically significant [36]. Other prevalence-based epidemiological studies in Ethiopia have resulted in similar findings: at Yirgalem Hospital in Ethiopia, results showed a statistically significant higher prevalence of *G. intestinalis* in female children (29.3%) compared to male children (8.1%) [37].

Implications for Human Health

There is no doubt that *Giardia intestinalis* is pathogenic in many cases. Symptoms of acute giardiasis commonly take the form of diarrhea, abdominal pain, malabsorption, bloating, fatigue, and weight loss [11,38]. Acute infection develops over a period of three weeks, peaking at eight days post infection. Healthy hosts typically are able to clear the infection within two to three weeks, but infection occasionally becomes chronic, potentially leading to the development of irritable bowel syndrome, chronic fatigue, and chronic diarrhea [38]. Some infections, especially in underweight malnourished children, can cause dramatic health impairments, including worsening of nutrition, growth retardation, and cognitive impairment. The worst

effects of *G. intestinalis* infection are related to damage of the absorptive small bowel mucosa, along with abnormal intestinal immunity [39].

The Centers for Disease Control and Prevention consider giardiasis to be a public health concern, as it is frequently diagnosed in the United States. They provide extensive information for the public on *Giardia intestinalis* and giardiasis, including illness and symptoms, diagnosis, treatment, risk factors, and prevention and control. The CDC also notes that giardiasis is a nationally notifiable disease, meaning health care providers and laboratories are required to report cases of the disease that they diagnose to local or state health departments. In addition, in 2004, *G. intestinalis* was added to the World Health Organization's Neglected Diseases Initiative. All Neglected Diseases have a common link with poverty and "exhibit a considerable and increasing global burden, and impair the ability of those infected to achieve their full potential" [40]. Particular concerns were *G. intestinalis* infections in children and individuals with AIDS, as the risk of disease complications is greater for these groups. In adding *G. intestinalis* to the Neglected Diseases Initiative, the hope was to gain more insight into the biology and impact of this parasite, especially through molecular methods, to generate better control strategies [40].

While concerns over symptomatic disease caused by *G. intestinalis* are valid, most cases of this protozoan in humans are asymptomatic, and there is increasing evidence that *G. intestinalis* is a frequent component of the gut microbiome. As part of the Malnutrition and Enteric Disease (MAL-ED) project from November 2009 to February 2014, *G. intestinalis* was detected in two-thirds of 1,741 children across eight countries in which diarrheal disease and malnutrition are endemic (Bangladesh, Brazil, India, Nepal, Peru, Pakistan, South Africa, and Tanzania) [41]. At each of the eight study sites, the parasite was found in 37 to 95% of children

within the first two years of life, yet *G. intestinalis* was not significantly associated with diarrhea regardless of site or age [42]. Similarly, a separate study in Peru detected *G. intestinalis* in 85% of children, with 87% becoming reinfected, and the infection did not affect growth nor was associated with an increased risk of diarrhea at any age [1]. Non-diarrheal *G. intestinalis* infections have not only been highly prevalent in endemic areas, but *G. intestinalis* detection has frequently been negatively associated with diarrhea, suggesting it has a protective effect against diarrhea [42]. Corroborating this, quantitative nucleic-acid based diagnostics have suggested that higher quantities of *G. intestinalis* are associated with even greater reductions in diarrhea risk [43]. Given the ubiquity of the protozoan and its failure to show an association between infection and symptomatic illness in many cases, some even argue against the administration of *Giardia*-specific drugs in endemic areas [44].

There is clearly ambiguity in the health implications of *G. intestinalis* for humans. Still, the majority of the literature on *G. intestinalis* focuses on it as a pathogen that should be controlled, recommending prevention measures such as better sanitation, hygiene, and water treatment. Out of 3,741 articles on *Giardia* published on NCBI's PubMed database between 1915 and 2018, the highest percentage of articles (27.5%) are on the biology of the organism itself, either seeking to understand more about an early diverging eukaryote, or researching the biological mechanisms by which *G. intestinalis* causes illness in humans. Articles that explore the presence of *Giardia* in various water sources or water treatment practices to eliminate it, along with *G. intestinalis*' general prevalence in humans, constitute 22.2% of all the published articles. Many of the articles on prevalence explore *G. intestinalis*' potential links with other health issues. Additionally, 11% of articles address the presence of the various genetic assemblages of *G. intestinalis* in humans, and 9% propose different treatments for *G. intestinalis*

infection. As is evident from this data, research into asymptomatic cases of *G. intestinalis* is lacking. Yet, since those cases are pervasive, it raises the possibility that *G. intestinalis* is sometimes a commensal organism in the human microbiome.

While more research needs to be conducted, there are some hypotheses about why the manifestations of *G. intestinalis* infection vary. One is the ecological perspective, where the diversity and interactions of commensal microbes in the mammalian gut affect *G. intestinalis* colonization and consequent symptoms. Physiological diversity of commensal microbiota is crucial for overall stability of the gut ecosystem; greater diversity allows more resilience and flexibility of microbial responses to external stress, aiding the immune system. If this ecosystem is disrupted, it may impact the disease manifestation of *G. intestinalis* [28]. There is evidence that interactions between *G. intestinalis* and commensal microbes could contribute to variations in pathogenesis. In one murine study, for example, it was found that mice from one commercial breeding farm were less susceptible to *G. intestinalis* infection than were mice from another facility, presumably due to differences in the composition of commensal gut microbiota between the two groups [45].

A second hypothesis for the variation in or lack of *G. intestinalis* symptoms is differences in the genetic strains, or assemblages. As assemblages A and B are the primary strains found in humans, their relations to symptoms of *G. intestinalis* infection in humans have been studied the most. Some research has found more severe symptoms, such as diarrhea, associated with assemblage B infection, compared to assemblage A [46]. A murine study determined that infection with *G. intestinalis* assemblage B caused decreased growth and mucosal histopathological changes in mice that matched what is seen in chronic human giardiasis [39]. In

addition, infection with particular subassemblages of A or B could contribute to symptom variation, as could infection with other assemblages of *G. intestinalis* in humans.

Current Understanding of the *G. intestinalis* Genetic Assemblages in Humans

An analysis of over 4,000 human isolates from different geographical locations by PCR amplification of fecal DNA extracts demonstrates that primarily only *Giardia intestinalis* assemblages A and B are found in humans [11]. However, the distribution of these two assemblages in humans varies greatly. For example, out of a total of 78 positive cases of *G. intestinalis* in Ethiopian children that were able to be typed, 14 isolates (17.9%) were assemblage A and 64 isolates (82.1%) were assemblage B [36]. In contrast, out of 60 positive *G. intestinalis* samples from Iran, 35 isolates (58.3%) were assemblage A, 17 isolates (28.3%) were assemblage B, and 7 (11.6%) contained a mix of both A and B [47]. Out of the 4,000 isolates analyzed, assemblage B was slightly more prevalent in both developed (1,589 isolates) and developing (708 isolates) countries compared to assemblage A (1,096 and 482 isolates, respectively). However, there is no clear geographic pattern in the distribution of the assemblages; their distribution can vary even within the same country [11]. For example, one study in Peru determined 10 positive *G. intestinalis* isolates as assemblage A compared to 6 of assemblage B [48], and another Peruvian study found 6 assemblage A isolates compared to 19 assemblage B [5]. While these sample sizes are small, there are numerous instances of assemblages A and B distribution variation both within and between countries [11,49] that corroborate the inability to make any conclusions about a pattern of distribution.

There is, however, a pattern in the host distribution of assemblages A and B, as well as potential host and geographic patterns in the distribution of the subassemblages AI, AII, AIII,

BIII, and BIV. An analysis by Sprong et al. (2009) of 3,812 *G. intestinalis* sequences extracted from a variety of sources found that humans are the major source of assemblage B; 56% of the sequences from humans were assemblage B, compared to less than 5% of the sequences from cats and livestock, and less than 10% of the sequences from dogs. Conversely, assemblage A had a fairly even distribution among companion animals, livestock, humans, and wildlife [50]. This pattern could relate to why assemblage B was found to be slightly more prevalent among humans globally, and also why assemblage B often causes more severe symptoms for humans.

Sprong et al. (2009) also examined distributions of the subassemblages of A and B, identifying some clear patterns. Out of 594 human cases of *G. intestinalis* assemblage A, the majority (75%) were of the subassemblage AII, and 25% were AI. In contrast, in all other hosts of assemblage A—companion animals, livestock, and wildlife—subassemblage AI was significantly more prevalent than the other subassemblages. None of the *G. intestinalis* cases in humans, dogs, goats, sheep, and pigs were subassemblage AIII, and its presence is overwhelmingly in wildlife, occurring in 52% of assemblage A cases in that source. There was a similar pattern between humans and other animals in the distribution of the subassemblages of assemblage B. Of 787 human cases of assemblage B, 56% were BIII, whereas the domestic dog and wildlife cases of assemblage B were largely subassemblage BIV (73% and 94% of cases, respectively) [50].

From the human cases of *G. intestinalis* assemblages A and B described above, some geographic patterns emerged in the distribution of the subassemblages [50]. Subassemblage AI was more prevalent in Asia and Australia (60% and 69% of cases, respectively) than AII, and AII was more prevalent than AI in all other parts of the world. However, the distribution of the two subassemblages was closer to 50% in the Americas than elsewhere in the world, suggesting

more genetic diversity of assemblage A in the Americas. Subassemblage BIII was noticeably more prevalent in Africa, Asia, the Middle East, and Central/South America compared to subassemblage BIV, and in North America the opposite was true, with BIV comprising 86% of the assemblage B cases in humans. In Australia and Europe, subassemblages BIII and BIV were almost evenly distributed, with a BIII to BIV distribution of 52% to 48% in Australia and 49% to 51% in Europe [50].

Mixed infections of assemblages A and B together can occur, as well. Of the 4,000 human isolates around the world mentioned above, 23 isolates in developed countries and 84 isolates in developing countries were mixed infections of A and B [11]. Additionally, other *G. intestinalis* assemblages have occasionally been detected in human samples. Of 1,658 *G. intestinalis* isolates extracted from European individuals, two were assemblage C, and there were four each of assemblages D, E, and F [50]. Humans were also found to be infected with assemblage C in Thailand [6], Brazil [51], and Egypt [7], assemblage E in Egypt [52,53], Brazil [8], and Australia [54], and assemblage F in Ethiopia [4] and Slovakia [55]. Furthermore, the seven instances of assemblage F in Ethiopia were in fact mixed *G. intestinalis* infections with assemblage A [4], and one case of assemblage E in Egypt was also a mixed infection with assemblage B [52]. While mixed infections of assemblages A and B are not surprising, these other combinations of mixed *G. intestinalis* assemblages in humans are unique and intriguing.

The Gaps in *Giardia* Research

Research on the potential of *Giardia intestinalis* to be a commensal member of the human gut microbiome is lacking, and more studies on asymptomatic cases in humans could provide clarity on that possibility. Furthermore, research into the presence of *G. intestinalis* in

more traditional and hunter-gatherer societies can be particularly enlightening, as the gut microbiomes of these individuals have not been affected by industrialized diets, sanitation practices, medicines, and other practices of industrialization. The gut microbiomes of traditional peoples provide unique representations of human guts without much modern interference. Thus, if *G. intestinalis* is detected, it provides strong evidence for this organism being a natural, commensal member.

Few studies have explored the presence or assemblage diversity of *G. intestinalis* in hunter-gatherers. Assemblage information in particular is crucial for understanding asymptomatic cases of *G. intestinalis*, as variation in assemblages could explain the differing human responses to this microeukaryote. Under the search term "*Giardia*" and filter "Humans" on PubMed, nine studies were published between 2010 and 2020 on *G. intestinalis* in traditional or hunter-gatherer populations. The studies were performed in such locations as the subtropical Atlantic Forest of Argentina [32], the Colombian Amazon Basin [34], Peninsular and East Malaysia [56,57,58], Northern Thailand [59], Central Taiwan [60], Northwestern Ecuador [61], and the Peruvian Amazon [2]. While the studies are fairly numerous, just three included hunter-gatherer communities [2,57,58] and only one examined the prevalence of *G. intestinalis* in hunter-gatherers of South America [2]. Furthermore, over half of the nine studies took place in Asia. This is in accordance with the noticeable gap in *G. intestinalis* research in South America in general.

Of the nine studies, five assessed assemblage diversity of *G. intestinalis* [34,56,57,59,60], with four of those studies located in Asia and nearly all study participants being children. Only assemblages A and B were detected in all studies. The aboriginal community in Taiwan had the lowest overall prevalence of *G. intestinalis* at 3.83% and was also the only community to not

harbor assemblage B. All positive cases of *G. intestinalis* were assemblage A, specifically subassemblage AII [60]. All other communities from Asia had an even or nearly even distribution of assemblages A and B, however, distributions of the subassemblages differed by location, even within the same study. In Thailand, tribes in the Hod District had slightly more genetic diversity of *G. intestinalis* than those from the Mae-Chaem District, with subassemblages AII, BIII, and BIV detected compared to just AII and BIV, respectively. Subassemblage BIV was also more common in tribes from the Hod District, present in 60% of positive *G. intestinalis* samples, whereas AII was most prevalent in individuals from Mae-Chaem, identified in 77.8% of positive samples [59]. All cases of *G. intestinalis* assemblage A in Orang Asli tribal children in Malaysia were identified as subassemblage AII, whereas assemblage B consisted of nucleotide variation too high to enable subtyping. There were also 43 cases of mixed assemblage *G. intestinalis* present in the Orang Asli children [57].

The highest prevalence of *G. intestinalis* in these traditional societies was found in children from the Colombian Amazon, with 64.8% of screened samples being positive [34]. Additionally, there was significant genetic diversity within these cases, with subassemblages AI, AII, BIII, and BIV all detected. The frequency of each of these subassemblages depended on which gene was targeted in the analysis. According to analysis of the *gdh* gene, subassemblage AI was most common (61%), followed by subassemblage BIII (32%). In contrast, the *tpi* gene determined subassemblage BIII to be more common (59.6%) compared to AI (19.1%) [34]. While the study on *G. intestinalis* in hunter-gatherers of Peru did not examine assemblage diversity, overall prevalence of the organism was 29%. *G. intestinalis* was the second most common intestinal parasite in these Peruvians next to *Blastocystis hominis* (40%) [2].

Not only is there a lack in data on *G. intestinalis* assemblage diversity in traditional populations, but there is disproportionate reporting of assemblages around the world. This is reflected in Figure 2, which is our representation of the global assemblage data collected and made publicly available by Garcia-R. et al. (2017). The researchers extracted 4,348 records on *G. intestinalis* assemblages detected in humans and animals from 64 countries, and discovered that assemblages A and B represent more than 60% of the reported data worldwide [62], which is evident from the pie charts in Figure 2. Furthermore, the majority of available data were retrieved from China, Australia, and Brazil [62]. In fact, data from China comprises 72% of the total data from Asia and data from Brazil is 57% of all the data from South America. Excluding Brazil, data from South America comprises just 5.6% of the global data (13% with Brazil). Additionally, only 6.1% of the world data is from Africa, and a mere 0.9% is from Central America. The obvious paucity of data from Central America, most of South America, and nearly all of Africa reflects the bias in published *G. intestinalis* research, and is likely contributing to primarily assemblages A and B being reported.

Considering assemblages other than A and B have been reported in humans, but only in European or urban communities, and given the significant amount of missing data from developing countries in general and traditional or hunter-gatherer societies in particular, questions arise as to what kind of *G. intestinalis* assemblage variation may be present in these populations. If found, this information may shed light on the conditions under which *G. intestinalis* can be commensal in humans.

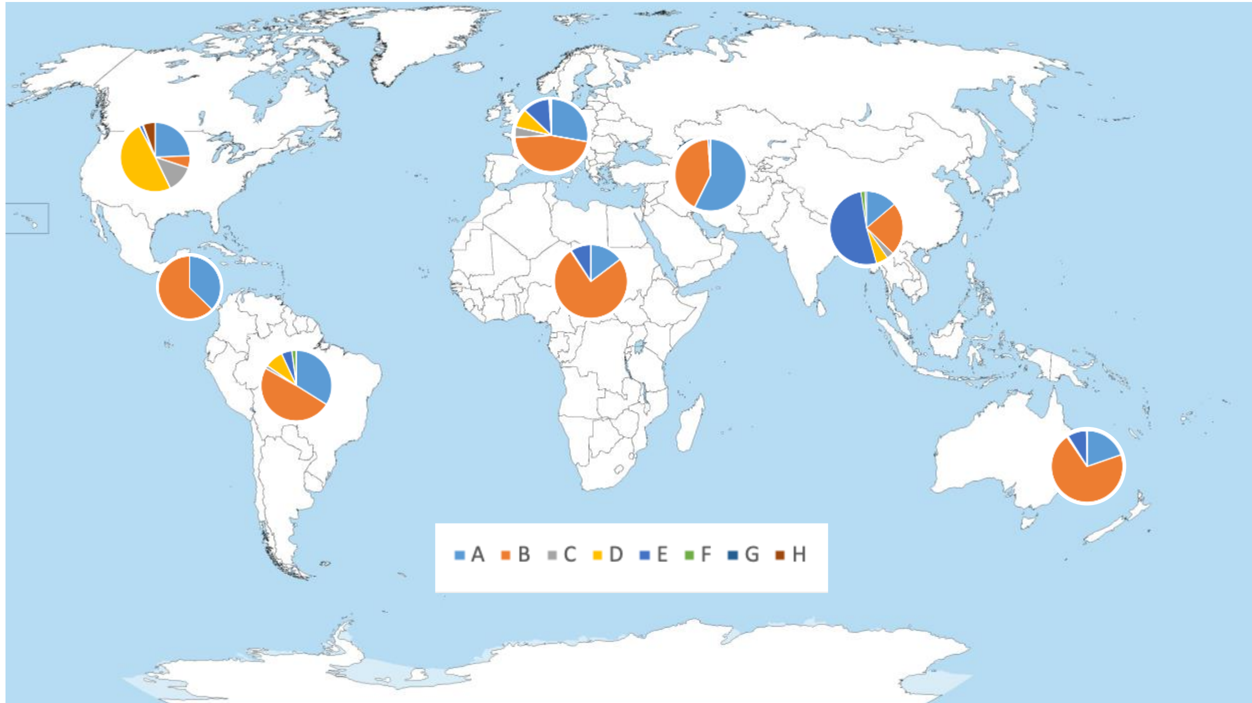

Figure 2. Global distribution of *G. intestinalis* assemblages from data presented in Garcia-R. et al. (2017). This data represents *G. intestinalis* isolates extracted from both humans and animals.

CHAPTER 2

BACKGROUND – FROM MICROBIOME TO EUKARYOME

The Human Gut Microbiome

Our bodies are not entirely human. While humans consist of 0.3 x 10^{13} nucleated cells, we are also home to approximately 3.8 x 10^{13} microbial cells that primarily inhabit four regions of the body: skin, oral, vaginal, and gut [63]. These numbers, in effect, estimate our cellular composition as only 10% human. Joshua Lederberg is credited with first suggesting the concept of the human microbiome, defining the phrase to "signify the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space" [64]. Now, "microbiome" is commonly used interchangeably to refer to all of the microbiota in a given habitat, as well as the collective genome of all those microbiota.

The most densely colonized body site is the gut, with approximately 99% of the 3.8 x 10^{13} microbial cells residing in the colon [65]. Due to its abundance of microorganisms and their associations with a variety of disorders, including immune and metabolically driven diseases, the gut microbiome has been the most studied of the microbial regions. As of 2018, a total of 15,335 publications out of almost 40,000 indexed under the search term "microbiome" in PubMed pertained to research on the gut microbiome, specifically. This was the largest category of microbiome publications, with the other major aspects of microbiome research being host processes (13,805 publications) and diet (5,709 publications) [66].

Composition and function

Microorganisms are found throughout the human gastrointestinal tract, with variations in type, density, and metabolic activity of microbiota residing in the stomach, small intestine, and

colon. Contents of the small intestine reside there for a relatively short period of time, with a median time of 255 minutes, which prevents microorganisms from reaching a high density [65]. In contrast, the long transit time in the colon of about 35 hours [67] allows more proliferation of microorganisms. In general, the microbial composition in the small intestine is also more variable among individuals and over time compared to the fecal microbiota [65]. Furthermore, the diversity of bacterial genera is greatest in the colon, with 34 major genera identified in the colon compared to 14 genera in the small intestine and eight genera in the stomach [65]. Only three genera that are found in the stomach (*Gemella, Rothia, Helicobacter*) are not shared in the colon, and seven genera that reside in the small intestine (*Megasphaera*, *Brevibacillus*, *Veillonella, Gemella*, *Leptotrichia, Neisseria, Haemophilus*) are not in the colon. Among all three areas of the gastrointestinal tract, just five major genera are shared (*Streptococcus, Porphyromonas, Prevotella, Atopobium, Fusobacterium*) [65]. It should be stressed that these are only the major bacterial genera, and estimations of the true number of microbial species present in the colon or feces vary significantly, although the count is undoubtedly much higher compared to other microbial habitats.

While the majority of microbial species in the gut are bacteria, the gut microbiome also consists of archaea, eukaryotes, and viruses. *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* are two primary species of archaea found in the gut, both of which aid in the formation of methane [65]. Intestinal fungi, referred to as gut mycobiome, are the most prominent eukaryotic members [65]. Although not well studied compared to the prokaryotes, some attempts at characterizing intestinal fungi have been made. For example, an analysis of fecal samples from 98 healthy individuals revealed 66 fungal genera and an estimated 184 species. *Saccharomyces, Candida,* and *Cladosporium* were the most prevalent genera, found in

89%, 57%, and 42% of samples, respectively [68]. Even less studied than intestinal fungi are protozoan microeukaryotes, which are also important members of the microbial community. As more research on microeukaryotes emerges, it is becoming more evident that both fungi and protozoa play important ecological and health-related roles in the gut microbiome [69]. Viruses are not considered to be living organisms, but nevertheless several thousand bacteriophage genomes have been detected in human fecal samples [70].

One important function of gut microbiota is to protect the host from pathogens and unwanted colonization, in part by supporting the immune system. However in addition to this, microbiota also provide enzymes that expand the metabolic capacity of the host, assisting in the breakdown of dietary components that the host cannot use, such as carbohydrates, proteins, and secondary plant metabolites [66]. A mutualistic relationship exists between the microbiome and the host, with immune and metabolic receptors sensing and processing microbial signals. This relationship has led to the idea of the microbiome as a "forgotten organ" that coevolved with the mammalian host [66].

The gut microbiome has been divided into two key sections: the core microbiome and the variable microbiome. The microbiomes of different individuals share a high proportion of gene functions, namely for metabolic activities; these ubiquitous functions represent the core microbiome. In contrast, there are functions that are present in some individuals but not in others, creating the variable microbiome [65]. Several studies of unrelated, healthy adults have revealed substantial diversity in their gut communities [71,72,73]. Among family members, the human gut microbiome is generally shared, but each person's gut microbial community still varies in the specific bacterial lineages present [74].

There are still many gaps in our understanding of the human gut microbiome, including the composition, exact mechanisms underlying microbe-microbe or microbe-host interactions, and the role of various microorganisms in health and disease. One central question is whether microbiome alterations are the cause or consequence of pathologies [66].

The Anthropology of the Microbiome

Microorganisms are part of us as human beings, and the microbial ecosystem likely coevolved with us [71,72]. There is substantial interindividual variation in the microbiomes of human populations [66]; therefore, the microbiome is a source of human diversity and should be considered in anthropological studies surrounding human variation both within and between populations.

The diversity of the human gut microbiome is exemplified by the examination of fecal samples from individuals of different ages and from different geographic regions. An analysis of fecal samples from 531 individuals comprised of healthy children and adults from the Amazonas of Venezuela, rural Malawi, and U.S. metropolitan areas produced some significant findings regarding variation in gut composition [73]. There were some similarities in the gut microbiomes of individuals from all three regions. For example, the functional maturation of the gut microbial community to an adult-like configuration within the first three years of life was the same in all three populations. In addition, interpersonal variation was greater among children than among adults regardless of geographic location, intrapersonal variation decreased with age, and bacterial diversity increased with age in all populations. However, differences in the composition of fecal microbiota were detected between individuals from the different countries as well, especially between the U.S. and the Malawian and Amerindian gut communities. Of the three populations,

the fecal microbiota of U.S. adults was the least diverse, and there were distinct microbial community signatures for Western (U.S.) and non-Western individuals. Out of 92 species-level Operational Taxonomic Units (OTUs), 73 were over-represented in non-Western adults. Interestingly, the distinctions were less in infants: only 28 OTUs discriminated U.S. and non-U.S. infants [73]. These findings suggest that a Western lifestyle significantly alters the microbe composition of the gut and it does so systematically.

An important area of anthropological study is understanding how factors such as age, environment, and cultural traditions shape us as human beings. These factors directly affect the composition of the gut microbiome, which in turn influences human physical, mental, and pathophysiological states. Thus, examination into the variations in gut microbiomes between individuals and populations, especially between industrialized and non-industrialized societies, can answer a variety of questions surrounding the source of diverse human conditions.

The Gut Microbiome and the "Hygiene Hypothesis"

In 1989, David P. Strachan introduced a concept that became known as the "hygiene hypothesis". In a longitudinal epidemiological study of hay fever and atopic diseases among British children, Strachan noted associations between hay fever and family size and position in the household in childhood. Hay fever was inversely related to family size in that larger families had a lower prevalence of hay fever. Additionally, eczema in the first year of life was independently related to the number of older children in the household. These findings suggested that allergic diseases were prevented by infection in early childhood, acquired from unhygienic contact with older siblings [75]. Strachan deduced that increased standards of hygiene, as well as

declining family sizes, have reduced early childhood infections and consequently increased the prevalence of allergic diseases [75].

Since Strachan's initial proposal, there have been some critiques to the hygiene hypothesis. It is argued that the increase in allergic disorders does not correlate with the decrease in pathogenic infections and cannot entirely be explained by changes in domestic hygiene. What has been suggested instead is the "microbial exposure" or "microbial deprivation" hypothesis, where decreased exposure to certain microbial species has inhibited the development of immunoregulatory mechanisms [76]. Humans co-evolved with and learned to tolerate numerous harmless organisms, such as bacteria, archaea, fungi, and protozoa, that occupied the natural environment and entered the body on a regular basis. Isolated hunter-gatherers and other traditional populations have been able to retain this tolerance and benefit from the immunoregulation microbes provide [77], while industrialization has depleted microbial colonization in many other human populations. Several changes in recent decades in addition to hygiene practices—in food and water quality and general public health—likely altered human exposure to commensal microorganisms and components of microbes such as bacterial endotoxins [76]. In the attempt to reduce exposure to pathogens, humans have unintentionally enabled greater expression of atopy, or the genetic predisposition to develop allergic diseases, such as asthma, due to a heightened inflammatory response to common allergens [78]. This is especially true in Westernized countries, where a more sterile lifestyle prevents exposure to a variety of microbes that could serve protective functions. Both the incidence and prevalence of atopy is usually much lower in most developing countries [76]. The International Study of Asthma and Allergies in Childhood (ISAAC) showed an asthma prevalence of 2-3% in developing countries and 20-40% in developed countries in self-reporting 13-14 year-olds [79].

Evidence in support of the “microbial exposure” concept can be found in many areas, including place of residence, food-borne and gastrointestinal disease, gut flora, and intestinal parasites. Some studies have noted that living on a farm can have a protective effect against allergic diseases, even in comparison to other rural environments. For example, over a 16-year period, a statistically significant increase in the incidence and severity of hay fever and asthma was seen in rural children with no contact with agriculture, while children on farms had less atopic disease and lower levels of seroprevalence to a wide range of allergens [76]. The specific exposure responsible for this difference is uncertain, but some scholars are exploring bacterial endotoxins as the possible protective factor [76].

Bacterial flora in the gut can play an important protective role against atopy. In a study of allergic versus nonallergic Estonian and Swedish children, Bjorksten et al. (1999) discovered that allergic children in both countries were less often colonized with lactobacilli and bifidobacteria compared to the nonallergic children. In addition, the allergic children had both higher counts and proportions of aerobic microorganisms, particularly coliforms and *Staphylococcus aureus* [80].

While the term “parasite” carries a negative connotation, there is research to suggest some intestinal parasites can be protective against other afflictions, such as atopy. A case-control study of Italian males found that men with atopy had significantly lower serum levels of antibodies to the parasite *Toxoplasma gondii*, as well as the bacteria *Helicobacter pylori*, than the non-atopic controls. Similarly, among the 245 participants exposed to at least two of the three infections studied (*T. gondii, H. pylori*, and hepatitis A virus), allergic asthma and allergic rhinitis were rare, present in only one individual (0.4%) and sixteen men (7%), respectively [81]. *T. gondii, H. pylori*, and hepatitis A virus are all transmitted through the fecal-oral or food-borne

route, thus the investigators of the Italian study concluded that the increased prevalence of allergic asthma and rhinitis in developed countries has been facilitated by greater hygiene and a semi-sterile diet [81]. These Western practices limit exposure to commensal or pathogenic microorganisms and influence their overall pattern in the gut.

Other parasites industrialized populations are less exposed to are helminths, which could also contribute to the atopy epidemic in the Western world. Several studies on helminths have noted their strong immunomodulatory effects [82], as well as their associations with lower prevalence of allergic diseases, particularly when infection is heavy or chronic [76]. An analysis of individuals from Ethiopia who self-reported wheezing found that the risk of wheeze was independently reduced by hookworm infection, and this seemed to be related to the intensity of infection. This suggests that parasitic infection by certain helminths might prevent asthma symptoms in atopic individuals [83].

T. gondii and hookworm are eukaryotes, and while it is unclear whether these two parasites in particular are regular residents of the healthy human gut microbiome, many eukaryotes are. Additionally, these other eukaryotes are often not pathogenic, but commensal or beneficial for the health of the host.

The "Eukaryome"

Many uni- and multicellular eukaryotes (protists and helminths) are thought of as "parasites", which implies that they are alien to the human gut. However, microbial eukaryotes (microeukaryotes) are common inhabitants of the human gut microbiome worldwide [84]. While the eukaryotic component of the gut microbiome—which has been called the "eukaryome" [85]—has been understudied in comparison to bacterial microbes, it is becoming increasingly

clear that microeukaryotes play important roles in the ecology of the human gut microbiome and in human health [69,84,85].

Eukaryotes, which are defined by the presence of nuclei, are most visibly represented by animals, plants, and fungi, but there are in fact more than 70 lineages of eukaryotes, most of which are microbial [84]. In the healthy human gut, fungi (such as filamentous fungi and yeasts) are the most prominent members [66,84], but the eukaryome also includes protozoans and helminths [69].

Historically, microeukaryotes have been studied for their parasitic or pathogenic relationships with the host. While some intestinal protists and helminths are in fact pathogenic, such as *Cryptosporidium* sp*., Entamoeba histolytica,* and *Ascaris lumbricoides*, most of the eukaryome is commensal or beneficial [69,85,84]. There are some microeukaryotes that are probiotics and thus benefit the host. A well-known example is the yeast *Saccharomyces boulardii*, which was an effective treatment for cholera and is now a general cure for diarrhea [84]. Others are occasionally beneficial, depending on the context, such as the example previously mentioned of the role of hookworm infection in reducing asthma and other allergic diseases [83]. Still other microeukaryotes are neither beneficial nor harmful (commensals), including *Pentatrichomonas* and *Entamoeba dispar* [84], or are pathogenic in some individuals and commensal in others.

One intriguing example of a potentially commensal protist that has frequently been associated with gastrointestinal disease in the literature is *Blastocystis* sp. [69,85]. This microeukaryote is quite common in individuals around the world, colonizing the intestine of 0.5 to 30% of people in industrialized countries and 30 to 100% of those in non-industrialized societies [69]. While the presence of *Blastocystis* in both healthy and diseased individuals creates

controversy over its role in human health, it clearly has associations with the bacterial ecology of the gut. Its presence is linked to a significant increase in bacterial alpha diversity [86] and to compositional shifts in abundant bacterial taxa [87]. One possible explanation for these effects of *Blastocystis* applies the idea of food web theory. It has been suggested that without the presence of *Blastocystis* in the gut microbiome, one strong bacteria can dominate the community, monopolizing nutrients and limiting species richness. *Blastocystis* may serve a predatory role on abundant bacterial taxa, thus lowering competition for resources and allowing more bacterial evenness to flourish [69]. More studies are needed to test this theory. *Blastocystis* may positively influence the host immune system as well by stimulating mucus production, which relieves symptoms of colitis and generally improves gut health. This ability to maintain a healthy mucus layer in the intestine is one potential reason why *Blastocystis* is more common in healthy individuals [85].

Just as there is substantial interindividual variation in the gut microbiomes of humans in general, the composition of the eukaryome varies among individuals as well, as do host responses to colonization by individual microeukaryotes, especially parasites. This results in some people experiencing disease while others are asymptomatic in the presence of common eukaryotic organisms. There are many factors behind the differences in both eukaryome composition and manifestations of pathology. One relates to environment and lifestyle: isolated communities, such as hunter-gatherers, have retained tolerance to certain infections by helminths and gut parasites from our evolutionary past [77]. Other major factors include host immune response, prior exposure to the organism, host genetics, host nutritional status, and co-infection with multiple parasites. The variations in bacterial and archaeal residents with which parasites and other microeukaryotes interact also likely contribute to disparate responses to them by hosts

[84]. Furthermore, disruptions to an individual's normal gut microbial community often causes disease symptoms, similar to how symptoms from *Clostridioides difficile* (formerly *Clostridium difficile* [88]) result from changes in the gut microbial community. Some microeukaryotes that are present in the microbial communities of healthy individuals cause disease symptoms when their numbers increase, such as the flagellate *Chilomastix* and the fungus *Candida albicans*. Additionally, diseases such as Irritable Bowel Syndrome (IBS) and Inflammatory Bowel Disease (IBD) are known to be caused by *Blastocystis* and fungi, respectively, when complex changes in the microbiome occur. However, it is unclear whether these changes are restricted to the site of disease or occur as systemic changes in the overall gut [84].

The most notable difference in composition of the human gut microbiome, including the eukaryome, exists between individuals from industrialized versus non-industrialized societies. The diversity of microeukaryotes in the gut is significantly lower in industrialized populations [69,85], due to several differences in lifestyle and environment that have altered the routes of dispersal of many protozoans [69]. Industrialized societies have experienced improved personal hygiene and increased food and water sanitation, in addition to targeted removal of particular protozoans. Because many eukaryotes in the human gut have been labeled as parasites and thus assumed to have negative effects on the host, Western doctors almost always remove them, even in asymptomatic individuals [85]. Furthermore, copious antibiotic use in industrialized populations can have negative consequences for not only the beneficial bacteria in the gut, but for the microeukaryotes as well. For example, the modifications bacteria experience as a result of antibiotics can in turn influence the colonization and survival of gut fungi [69].

High diversity of the gut microbiome in general is associated with healthy individuals and fewer cases of autoimmune and inflammatory disease [77]. Thus, it is likely that greater

diversity of the eukaryome is also beneficial for host health, as gut microeukaryotes can encourage the diversity of the microbiome overall [85], as in the *Blastocystis* example above. The industrialized gut's paucity in microeukaryotes from the practices previously mentioned can have a "domino effect" on the rest of the microbial community and consequently on host immunity [69], explaining the abundance of autoimmune diseases and more frequent symptomatic responses to protozoan colonization in industrialized populations compared to more traditional societies.

The eukaryome has been studied less than the bacterial microbes in the human gut and as such it is far less understood. In particular, more research is needed to better understand how microeukaryotes interact with bacteria and other members of the gut microbiome, given the variation in pathology manifestation that can occur, and the likely important ecological roles of eukaryotes in the gut. The eukaryome has historically been viewed as inherently harmful, yet most intestinal microeukaryotes are commensal or beneficial, and ignoring those positive roles may actually be hurting human health [85]. Examining the gut microbiomes—and their eukaryotic components in particular—of healthy individuals from non-industrialized, traditional societies, will be important for a broader understanding of what constitutes a normal eukaryotic resident in the microbiome.

Our perceptions of certain intestinal parasites may especially change. Previously mentioned were some intestinal parasites (*Toxoplasma gondii* and helminths) that are occasionally beneficial for their host, protecting against atopy. In addition, *Blastocystis* is a eukaryote that, while generally very common worldwide, is less prevalent in industrialized populations and also more likely to cause disease symptoms in that demographic compared to traditional societies. Yet, there is strong evidence to suggest it is an important ecological player

in the gut microbiome [69]. Another intestinal parasite that is common around the world—in fact, it is the most common intestinal parasite—is *Giardia intestinalis* (also known as *G. duodenalis* and *G. lamblia*). Similar to *Blastocystis*, it is more prevalent in non-industrialized societies and is also frequently asymptomatic [39,42], though it can cause diarrheal disease [38]. Given that there are other intestinal parasites that can be commensal or beneficial, and considering the similarities *G. intestinalis* shares with *Blastocystis*, it is reasonable to suspect that *Giardia* is more than a pathogen.

CHAPTER 3

MATERIALS AND METHODS

Samples

Human fecal samples previously collected from two geographically distinct populations were used for the purposes of this study. These consist of the Matses (N=19), a hunter-gatherer population from the Amazon jungle of Peru, and traditional individuals from the West African country of Burkina Faso (N=46). The Matses population (denoted by SM) consisted of 7 males and 12 females, with an age range of 2 to 52 years; the Burkina Faso population (denoted by TM) comprised 26 males and 20 females, aged between 5 and 87 years.

DNA was previously extracted from the fecal samples using the Qiagen AllPrep PowerViral DNA/RNA kit, following manufacturer's protocol. For the purposes of this study, genomic DNA from two strains of *Giardia intestinalis*, WB clone C6 and Be-1, was obtained from BEI Resources. This DNA was used for PCR optimization and as a positive control.

Targeted amplification

There is a great diversity of bacteria and microeukaryotes in the human gut, many of which are present in high abundance, which complicates molecular methods used for characterizing *Giardia*. Consider that a *Giardia intestinalis* genome is roughly 11.7 Mb [39,42] in size while a bacterial genome averages 3.6 Mb [89]. Even being conservative, it takes less than five bacterial genomes, on average, to equal the *G. intestinalis* genome size; however, in all likelihood, given the billions of bacterial cells in the human gut [65], such cells are likely to outnumber *G. intestinalis* cells by the millions, although there has yet to be accurate assessments for the latter. Shotgun metagenomics methods will sample genomes at random, based on the

relative frequencies/genome size in the DNA extract, and thus capturing *G. intestinalis* variation would require a depth of sequencing that is unreasonable by current methods. Therefore, a targeted approach to detecting *G. intestinalis* in extracted human fecal DNA samples was deemed necessary. A commonly targeted region for detecting microeukaryotes in the gut microbiome is the ITS (Internal Transcribed Spacer) region. However, this region is extremely truncated in *Giardia*, with an ITS1 region of 35-40 bp likely among the shortest in all eukaryotes [90]. Amplification of a specific gene was therefore the best option for identifying *G. intestinalis* in these samples.

Selection of genes

An extensive literature review of publications under the search term "*Giardia*" in the database of the U.S. National Library of Medicine (PubMed) was conducted to determine the best genes to target for its detection in fecal samples, as well as potential primer sets to use. From this review, five genes of *G. intestinalis* were found to have been previously successfully used for detection: *elongation factor 1 alpha (ef1α), glutamate dehydrogenase* (*gdh*), *triose phosphate isomerase* (*tpi*), *beta-giardin* (*bg*), and SSU rRNA, with the *gdh* and *tpi* genes being especially common for discriminating between *G. intestinalis* assemblages. Feng and Xiao (2011) provide a total of ten primer sets targeting four out of these five genes (excluding *elongation factor 1 alpha*) that are known to be successful in identifying *G. intestinalis*. Due to the abundance of reference sequences available on the National Center for Biotechnology Information (NCBI) GenBank database for the *G. intestinalis gdh* gene and its previous success in discriminating between assemblages, *gdh* was chosen for this study.

PCR Optimization

Four primer sets targeting the *gdh* gene of *G. intestinalis*, according to Feng and Xiao (2011), namely, gdh1/gdh2, gdh3/gdh4, GDH1/GDH4, and gdhFi/gdhRi were tested on the *G. intestinalis* positive controls, strains C6 and Be-1, using a polymerase chain reaction (PCR). The annealing temperature for each primer set was determined using ThermoFisher's oligo Tm calculator and testing a range of temperatures on the positive controls using a gradient PCR. The primer sequences and annealing temperatures are given in Table 1 of this chapter.

Out of these, three primer sets best showed amplicons of the expected size: gdh1/gdh2 (about 750 bp), gdh3/gdh4 (about 530 bp) and gdhFi/gdhRi (about 430 bp), with the region amplified by the gdh3/gdh4 primer set located within the same region of the *gdh* gene amplified by the gdh1/gdh2 set. The PCR conditions for each primer set were the same, except for the annealing temperatures: the forward and reverse primers were at a final concentration of 0.5 µM, dNTPs at a final concentration of 0.2 µM, BSA at a final concentration of 0.1 mg/mL, and Phusion HS II enzyme was used in a volume of 0.1 µL per reaction. Three µL of DNA at an initial concentration of approximately 5 ng/µL was used. The final volume of the reaction was 20 µL. The PCR conditions were: an initial denaturation of 98°C for 30 seconds; 35 cycles of 98°C for 15 seconds, 60°C (gdh1/gdh2), 63°C (gdh3/gdh4) or 61.5°C (gdhFi/gdhRi) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 7 minutes for a final extension. A gel extraction and purification of the bands of the expected size was performed following the Qiagen QIAquick Gel Extraction Kit protocol. The purified products were provided to the Biology Core Molecular Laboratory at the University of Oklahoma for Sanger sequencing in both primer directions to verify that the expected regions of the *G. intestinalis* genome were amplified by each primer set.

Primer Set	Sequences (5'-3')	Optimal Annealing Temperature
gdh1/gdh2	gdh1: TTCCGTRTYCAGTACAACTC gdh2: ACCTCGTTCTGRGTGGCGCA	60°C
gdh3/gdh4	gdh3: ATGACYGAGCTYCAGAGGCACGT gdh4: GTGGCGCARGGCATGATGCA	63°C
GDH1/GDH4	GDH1: ATCTTCGAGAGGATGCTTGAG GDH4: AGTACGCGACGCTGGGATACT	62°C
gdhFi/gdhRi	gdhFi: CAGTACAACTCYGCTCTCGG gdhRi: GTTRTCCTTGCACATCTCC	61.5°C
gdh1I/gdh2I	gdh1I: GGATGGCTGCGAGAAGGCTAGATTCCGTRTYCAGTACAACTC gdh2I: AAGTCTGCACACGAGAAGGCACCTCGTTCTGAGTGGCGCA	62°C
gdh3I/gdh4I	gdh3I: GGATGGCTGCGAGAAGGCTAGAATGACYGAGCTYCAGAGGCACGT gdh4I: AAGTCTGCACACGAGAAGGCGTGGCGCARGGCATGATGCA	66°C

Table 1. Primer sets used in this study targeting the *gdh* gene of *G. intestinalis*

The Sanger sequencing data were analyzed using Geneious Prime 11.1.5 (http://www.geneious.com). The consensus sequences for the three amplicons were queried against the NCBI nt database in a Basic Local Alignment Search Tool (BLAST) search [91]. The sequences showed a match for the *G. intestinalis gdh* gene, confirming that the primer sets worked. Analysis using Geneious Prime showed that the consensus between the forward and reverse primer was slightly better for the gdh1/gdh2 primer set. Since the gdh1/gdh2 primer set amplified a region of 750bp (much longer than the regions amplified by gdhFi/gdhRi and gdh3/gdh4), the gdh1/gdh2 primer set was chosen for screening the human samples.

Screening of SM samples

Preliminary testing of eight SM DNA extracts was performed using the gdh1/gdh2 primer set and the PCR conditions described above. Amplicons were visualized using gel electrophoresis on a 1.5% agarose gel. Visualization of a band of approximately 750 bp

suggested that all eight samples contained *G. intestinalis*. Following this, the remaining SM DNA extracts (N = 11) were tested using the same PCR conditions.

Due to the presence of non-specific bands in most samples, the three SM samples (SM2, SM23, and SM29) showing bright and most distinct 750 bp bands were extracted and purified using the Qiagen QIAquick Gel Extraction Kit. Subsequently, a PCR was performed using the gdh1/gdh2 primer set and the purified SM2, SM23, and SM29 amplicons as the template DNA. Visualization of the products of this second round of amplification showed a reduction in the amounts of non-specific bands. Gel extraction was performed for the bands at a size of 750 bp, followed by purification and Sanger sequencing. This process of double-amplification followed by Sanger sequencing was repeated for other SM samples (total N = 16); however, ultimately, only four samples (SM2, SM23, SM29, and SM43) yielded usable sequences. The sequences from the remaining samples were deemed unusable, due to the many ambiguities and interference from non-specific DNA. A BLAST search was used to determine the best hit for the consensus sequences. Only one sequence, from sample SM23, showed *G. intestinalis* as the best and unambiguous hit.

Overall, there were difficulties in obtaining high-quality, specific amplicons from the SM samples, resulting in ambiguous sequence data using Sanger sequencing. Thus, we decided to adopt a hybrid-sequencing approach using a targeted amplification of the *G. intestinalis gdh* gene and next-generation sequencing (NGS) using Illumina technology. We designed this approach such that we could potentially sequence not just the targeted *G. intestinalis gdh* amplicon, but also all non-specific products to determine which other species were being amplified by the primer sets. This approach would be useful for a holistic understanding of the gut microbiomes of these individuals, as well as for understanding cohabitation of *G. intestinalis* and other

microorganisms. Furthermore, determining what other species are amplified by these primer sets would also inform us about further improvements for primer design.

Hybrid approach

We added Illumina-compatible P5/P7 adapters to the gdh1/gdh2 and gdh3/gdh4 primers, such that the primers would amplify the *gdh* gene as well as provide an attachment site for Illumina indexing adapters. These new primers were designated as gdh1I/gdh2I and gdh3I/gdh4I (Table 1) and a PCR optimization was performed using *G. intestinalis* C6 DNA as a positive control. The expected lengths of the amplified regions of the *gdh* gene were about 770 bp for gdh1I/gdh2I and about 550 bp for gdh3I/gdh4I. The PCR conditions remained the same as previously described with a few alterations: the forward and reverse primers were at a final concentration of 0.5 µM, dNTPs at a final concentration of 0.2 µM, and BSA at a final concentration of 0.1 mg/mL. Phusion HS II enzyme was used in a volume of 0.1 µL per reaction. One µL of DNA at an initial concentration of approximately 5 ng/µL was used. The final volume of the reaction was 20 µL. The PCR conditions were: an initial denaturation of 98°C for 30 seconds; 35 cycles of 98°C for 15 seconds, 62°C (gdh1I/gdh2I) or 66°C (gdh3I/gdh4I) for 30 seconds, and 72°C for 1 minute, followed by 72°C for 7 minutes for a final extension.

Screening of SM samples

A quantitative PCR (qPCR) was performed to identify the appropriate cycle number for amplification. The FastStart Essential DNA Green MasterMix and the gdh3I/gdh4I primers were used to test the SM samples, the *G. intestinalis* C6 DNA (positive control), and water (non-template control, NTC). The qPCR conditions were as follows: 95°C for 10 minutes for the

initial denaturation and 50 cycles of 95°C for 10 seconds, 66°C for 20 seconds, and 72°C for 30 seconds. Most samples amplified between 35 to 50 cycles, suggesting that if the targeted *G. intestinalis gdh* amplicon was present, it was likely at very low amounts.

Following these results, the SM samples and the positive control were amplified using the gdh3I/gdh4I primers and 50 cycles (annealing temperature 66°C and all other conditions the same as previously). The amplicons were visualized using gel electrophoresis, and purified using a 1.8X bead cleanup using the SeraMag Magnetic SpeedBeads [92] to eliminate all bands shorter than 150 bp. After the SpeedBeads were added to each PCR product, the bead cleanup protocol, briefly, was as follows: incubation at room temperature for 5 minutes, placing on a magnet for 5 minutes, removing the supernatant, two washes on the magnet with 150 µL of 80% ethanol, drying for 4 minutes, resuspending in 20 µL of water, followed by another 5-minute incubation and removal of the supernatant. These purified products were then used as the template DNA for an indexing PCR.

In the indexing PCR, each sample was tagged with Illumina-compatible sequencing adapters containing unique indices. The conditions were as follows: the i5 and i7 indices were at a final concentration of 0.5 µM, dNTPs at a final concentration of 0.2 µM, and BSA at a final concentration of 0.125 mg/mL. Phusion HS II enzyme was used in a volume of 0.2 µL per reaction. Four µL of the template DNA and one µL of each unique index were used. The final volume of each reaction was 20 µL. The indexing PCR conditions were: an initial denaturation of 98°C for 30 seconds; 30 cycles of 98°C for 15 seconds, 60°C for 30 seconds, and 72°C for 45 seconds, followed by 72°C for 5 minutes for a final extension. This amplification was performed over up to five replicates per sample. The replicates of each sample were combined and purified using the 1.8X SpeedBead cleanup.

Screening of TM samples

In addition to the 19 SM samples, Burkina Faso (TM) samples were also amplified using the hybrid approach. All 46 TM samples were screened using the gdh1I/gdh2I and gdh3I/gdh4I primer sets; the amplification with the gdh1I/gdh2I primer set resulted in less non-specific amplification. This was in contrast to the SM samples, for which the gdh3I/gdh4I primer set showed less non-specific amplification. A subset of 25 TM samples that showed bright bands at the desired length of about 770 bp were chosen for indexing (this study was limited by only 48 unique indices being available for a single Illumina sequencing run). These 25 TM samples which had been amplified using gdh1I/gdh2I primers, were purified, indexed, and purified as described for the SM samples. This resulted in a final number of 45 samples (SM=19, TM=25, and *G. intestinalis* C6 positive control) which were included on the Illumina MiSeq sequencing run.

The remaining 19 TM samples, which were not indexed for sequencing using the Illumina MiSeq, were instead sequenced using the gdh1I/gdh2I primers and Sanger sequencing. Following the initial PCR using gdh1I/gdh2I primer set and gel electrophoresis, all samples showed a band of about 770 bp. All samples underwent a gel extraction and purification following the Qiagen QIAquick Gel Extraction Kit protocol. To increase the final DNA concentrations of the samples after the purification process, two replicates of each sample were included. Ultimately, 12 out of these 19 samples were Sanger-sequenced in both primer directions at the Biology Core Molecular Laboratory at the University of Oklahoma.

Illumina MiSeq sequencing

All indexed samples (N=45) were quantified using the Kapa Library Quantification Kit and following the kit protocol. Kapa Library Standards and NTC were included. Samples were also analyzed using either a Fragment Analyzer or TapeStation system to acquire their DNA concentration (ng/µL), average bp sizes, and molarity (nmol/L). For the Fragment Analzyer, the High Sensitivity Small Fragment Kit was used, whereas the D1000 assay was used for the TapeStation.

All samples were pooled in equimolar ratios into two pools: one for fragments ranging between 150 bp and 500 bp (which should comprise all non-specific amplicons), and one for fragments ranging from 501 bp to 1000 bp (which should include the desired fragments of the *G. intestinalis gdh* gene). The average bp sizes and molarity (nmol/L) for both the 151-500 bp range and 501-1000 bp range, along with the concentration (ng/µL) and target nanomole (nM) for each sample were used to calculate the amount of each sample to add to each pool.

To purify the pools and eliminate fragments outside of the desired range, each pool was run on a PippinPrep. A 2% agarose cassette was used for the 151-500 bp pool (pool 1) and a 1.5% agarose cassette for the 501-1000 bp pool (pool 2). After the runs, both eluted pools were quantified using the D1000 High Sensitivity assay on the TapeStation as well as a Kapa Library Quantification qPCR assay. For the latter, dilutions of each pool were assessed in triplicate to accurately quantify the molarity of each pool prior to sequencing. The average molarity across all dilutions for each pool was calculated to be 6.896 nM for pool 1 and 6.259 nM for pool 2. Each pool was diluted to 4 nM. The final sequencing reaction comprised 15% pool 1 and 85% pool 2 and was loaded at a final concentration of 12 pM. The run also included a 15% PhiX spike-in to increase diversity. The sequencing was conducted using 2 x 250 bp V2 chemistry on the Illumina MiSeq machine at the Consolidated Core Laboratory at the University of Oklahoma.

Data analysis

The Illumina sequence reads were demultiplexed using the bcl2fastq software. The numbers of raw reads across all samples ranged from 317 to 2,249,066. The reads were trimmed using AdapterRemoval v2 [93] using the following parameters: --trimns, --trimqualities, --minquality 30, --maxns 0. Paired reads were merged. Analysis-ready reads (merged as well as unmerged trimmed reads) were then queried in a BLASTX search against the NCBI non-redunant (nr) database using DIAMOND, with default parameters [94]. The analysis-ready reads were also mapped to *G. intestinalis* reference genome using bowtie2 with default parameters and the –no-unal option to discard unmapped reads [95].

Phylogenetic analyses were performed using the sequence obtained from Sanger sequencing of the SM23 sample and 33 published *G. intestinalis* isolates belonging to different assemblages acquired from NCBI GenBank. Two outgroups were included for reference and to more accurately resolve the relationships among the *G. intestinalis* assemblages: *Giardia ardeae* (another species of *Giardia*) and *Spironucleus vortens* (another organism of the taxonomic family Hexamitidae). Using MEGA version 7 [96], the SM23 sequence and all 33 published sequences were aligned with ClustalW and phylogenetic trees were built using the Neighbor-Joining (NJ) method and the Kimura two-parameter model. The NJ method and Kimura two-parameter model are consistent with previous phylogenetic analyses of *Giardia intestinalis* assemblage data that produced well supported topologies [11].

CHAPTER 4

RESULTS AND DISCUSSION

Results

Molecular detection of *Giardia* has frequently used a taxonomically diagnostic gene approach, which we replicated by targeting the *glutamate dehydrogenase* (*gdh*) gene. However, we noted that a targeted PCR-based approach produced non-specific amplicons, despite the use of previously published primers that were supposedly specific for the *G. intestinalis gdh* gene. We then adopted a hybrid-sequencing approach using targeted amplification and next-generation sequencing (NGS) with Illumina technology. The goal of our hybrid approach was to obtain high quality sequences of the *G. intestinalis gdh* gene from Matses (SM) and Burkina Faso (TM) samples as well as to sequence non-specific products to determine which other species were being amplified by our primer sets. Ultimately, we aimed to assess the *G. intestinalis* assemblage diversity in these traditional populations.

Using the hybrid approach, 45 samples were sequenced on the Illumina MiSeq. Since the *G. intestinalis gdh* amplicons were expected to be greater than 500 bp in length (gdh1I/gdh2I = 770 bp, gdh3I/gdh4I = 550 bp), we expected that paired reads covering these amplicons would not overlap and would not be merged. As such, any merged reads in our data could be attributed to either our non-specific amplicons of smaller fragment lengths or adapter heteroduplexes formed during the amplification process. However, across all samples, the proportion of quality-filtered, merged reads ranged from 60-95% (Supplemental Table 1), suggesting that our target *G. intestinalis gdh* amplicons were either not sequenced or were covered by a very low proportion of reads. Mapping of the trimmed reads (merged and unmerged) to both the *G. intestinalis* genome as well as the *G. intestinalis gdh* gene showed that a negligible number of unique reads

mapped to these targets across all samples, including the *G. intestinalis* C6 positive control (Supplemental Table 1). The lack of reads mapping to the *G. intestinalis* genome for the positive control suggests that our hybrid sequencing approach did not work. A further analysis of the trimmed (merged and unmerged) reads employing a BLASTX search conducted using DIAMOND showed that the reads were attributed to various bacterial species, including commonly found gut microbes such as *Prevotella* and *Firmicutes*. None of the samples, including the positive control, comprised any reads attributed to *G. intestinalis*. Therefore, no further analyses were conducted on the Illumina data.

Out of the 19 SM samples, we were able to obtain four usable sequences using the targeted amplification and Sanger sequencing (for samples SM2, SM23, SM29, and SM43). The success rate was higher for the TM samples, with 12 out of 19 samples producing usable sequences. However, when the consensus sequences for these amplicons were queried against the NCBI nt database using a BLAST search, only one (sample SM23) showed *G. intestinalis* as the best and unambiguous hit. Other samples showed different bacterial species as the best hit. For example, *Ruminococcus champanellensis* NADP-dependent *gdh* gene was the best hit for SM2, whereas *G. intestinalis* was not in the results. The best hit for SM29 was *Flavonifractor plautii* NADP-specific *gdhA* gene; however *G. intestinalis gdh* gene of assemblage F was the 14th hit and nearly all of the rest of the 100 hits were *G. intestinalis*. The best hit for SM43 was the *Bifidobacterium adolescentis gdh* gene. The top 50 hits for SM43 were almost exclusively *Bifidobacterium* species, but several hits lower on the scale were the *G. intestinalis gdh* gene.

We further analyzed the SM and TM Sanger sequence data in Geneious Prime. The SM sequences have some noticeable differences between them - all except SM43 contain many ambiguities and, when aligned, have only 52.2% identical sites. In contrast, the TM sequences

are nearly identical, except two sequences (TM5-2 and TM12-4) that are not as clean as the others. With TM5-2 and TM12-4 excluded, an alignment of the remaining TM sequences shows 90.8% identical sites. The best BLAST hit for all TM sequences was the *Flavonifractor plautii gdhA* gene, with no *G. intestinalis* hits. However, the TM sequences show 65.6% identical sites and 92.5% pairwise identity with the *G. intestinalis* C6 control. Individually, the identical sites and pairwise identity of each TM sequence with C6 range from 67% to 71%.

Since SM23 was the only sequence to unambiguously register as *G. intestinalis* in the BLAST results, further analyses primarily only involved sample SM23. This sequence had numerous ambiguous sites; thus, we attempted to determine if two assemblages of *G. intestinalis* were coexisting in the same individual or if *G. intestinalis* was co-occurring with another identifiable organism. Additionally, we aimed to place this sample in a phylogeny of *G. intestinalis* assemblages by building a tree that included the SM23 sequence and 33 published *G. intestinalis* isolates belonging to different assemblages acquired from NCBI GenBank.

To address these questions, five versions of the SM23 sequence were created by editing the ambiguity codes in different ways. The first sequence was the original with zero ambiguity code alterations. To confirm the BLAST designation of assemblage C, phylogenetic analyses were conducted in MEGA involving the SM23 sequence and 33 published *G. intestinalis* isolates (Figure 1). Multiple reference sequences of each assemblage were included, as well as the two outgroups previously described: *Giardia ardeae* and *Spironucleus vortens*. SM23 did not form a

Figure 1. Neighbor-Joining tree with the original SM23 sequence. The Kimura 2-parameter model and 500 bootstrap replicates were used. The scale indicates substitutions per site.

clade with *G. intestinalis* assemblage C reference sequences. Rather, SM23 branched off independently from all published *G. intestinalis* strains, but with a relatively weak bootstrap support of 48. This finding calls into question the true assemblage designation of our sequence.

In the second and third versions, the ambiguity codes of the SM23 sequence were subsequently edited to firstly, examine the effect that assuming assemblage C had on the phylogenetic placement of this sequence and secondly, determine if a second assemblage of *G. intestinalis* was co-occurring in this individual. In the second version, all ambiguous sites of SM23 were individually examined and changed to a nucleotide that matched the assemblage C reference sequences, if possible. If the base call(s) suggested by the ambiguity code did not match the reference base in assemblage C, but matched that in another *G. intestinalis* assemblage, the position was edited to match the other assemblage. If neither option matched the base present in any assemblage, the ambiguity code was replaced with an N. The resulting SM23 sequence was used to build the NJ tree. In this tree, SM23 formed a clade with the *G. intestinalis* assemblage C reference sequences, but with low bootstrap support of 38 (Figure 2). When this edited SM23 sequence was queried against the NCBI nt database using a BLAST search, *G. intestinalis* was no longer the best hit; the best hit was *Methanomassiliicoccaceae archaeon*, a common methanogen. However, the *G. intestinalis gdh* gene of assemblage C was the 7_{th} hit, and every subsequent hit after was also *G. intestinalis*. Clearly, the assumption of assemblage C does affect the phylogenetic placement of our amplicon and adds credence to the original BLAST designation. However, we cannot make a valid conclusion from an assumption, and the low bootstrap support and incongruous BLAST results of the version 2 sequence reflect assemblage uncertainty of the SM23 isolate.

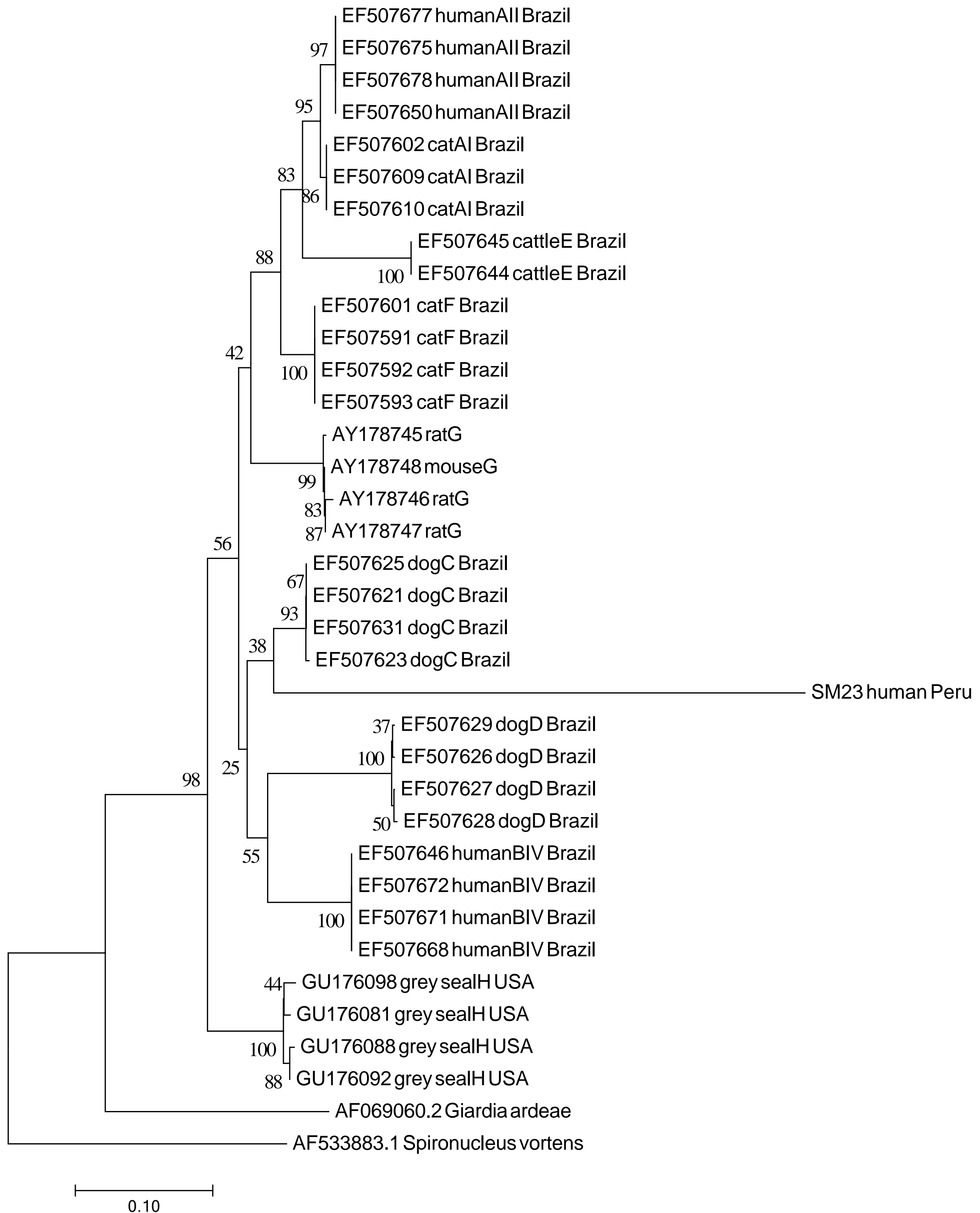

Figure 2. Neighbor-Joining tree with the SM23 sequence tailored to assemblage C. The Kimura 2-parameter model and 500 bootstrap replicates were used. The scale indicates substitutions per site.

Out of the total 59 ambiguous sites that were edited in version two, 30 were changed to a nucleotide that matched all *G. intestinalis* reference sequences. In the third sequence version, any ambiguous sites that did not match all assemblages were changed to the alternate nucleotide from the previous edit, according to the ambiguity code and provided it still matched one or more *G. intestinalis* assemblages. All "N"s, which were representative of ambiguities that did not match any assemblage, were kept as an "N". In the phylogenetic analysis that followed these alterations to the sequence, SM23 formed a clade with assemblage H sequences, with bootstrap support of 46 (Figure 3). BLAST results were essentially identical to those of version two of SM23, with *Methanomassiliicoccaceae archaeon* as the top hit and *G. intestinalis* assemblage C the 7th. Considering the low bootstrap value of the version three phylogeny, the fact that only 14 of the 59 ambiguities in the SM23 sequence were a difference between two *G. intestinalis* assemblages, and the identical BLAST results between SM23 sequence versions two and three, it is unlikely there is a co-occurrence of assemblages in our amplicon.

In the fourth version, to assess whether the ambiguities reflected the amplification of another organism's DNA in addition to *G. intestinalis*, all ambiguous sites in SM23 of which one nucleotide previously matched all *G. intestinalis* assemblages were changed to the alternative nucleotide. Additionally, any positions that had been deemed "N" were changed to one of the possible nucleotides according to the original sequence's ambiguity code. BLAST results were fairly consistent with the results from the two previous edits that had tailored SM23 to either *G. intestinalis* assemblage C or an alternative assemblage: *Methanomassiliicoccaceae archaeon* was again the best hit, with *G. intestinalis* the 6th hit and every hit after that. Since *M. archaeon* was the best hit in BLAST after all three edits to the SM23 sequence, it is possible we amplified the DNA from this archaeal species in addition to the *gdh* gene of *G. intestinalis*.

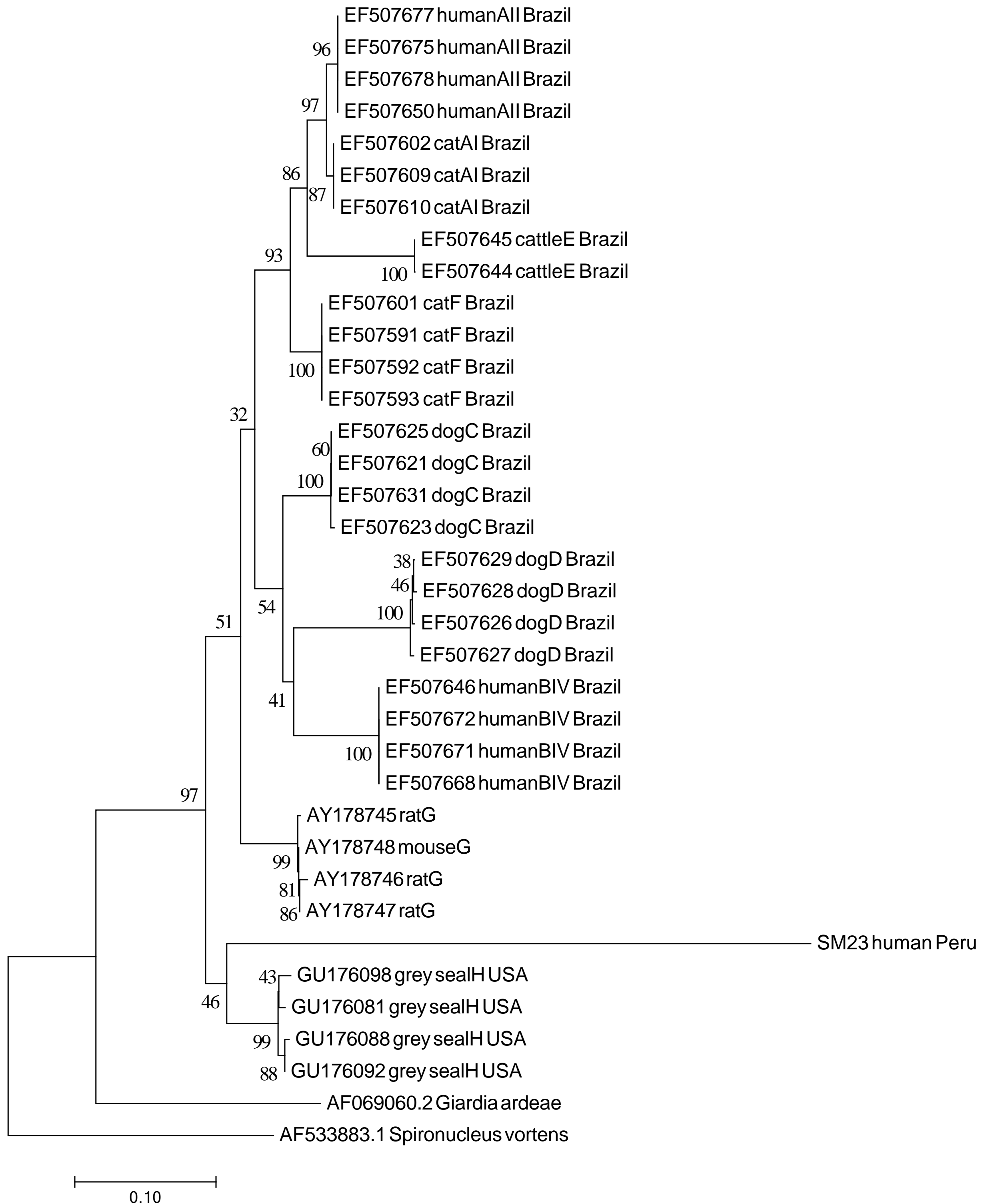

Figure 3. Neighbor-Joining tree with version three of the SM23 sequence. The Kimura 2-parameter model and 500 bootstrap replicates were used. The scale indicates substitutions per site.

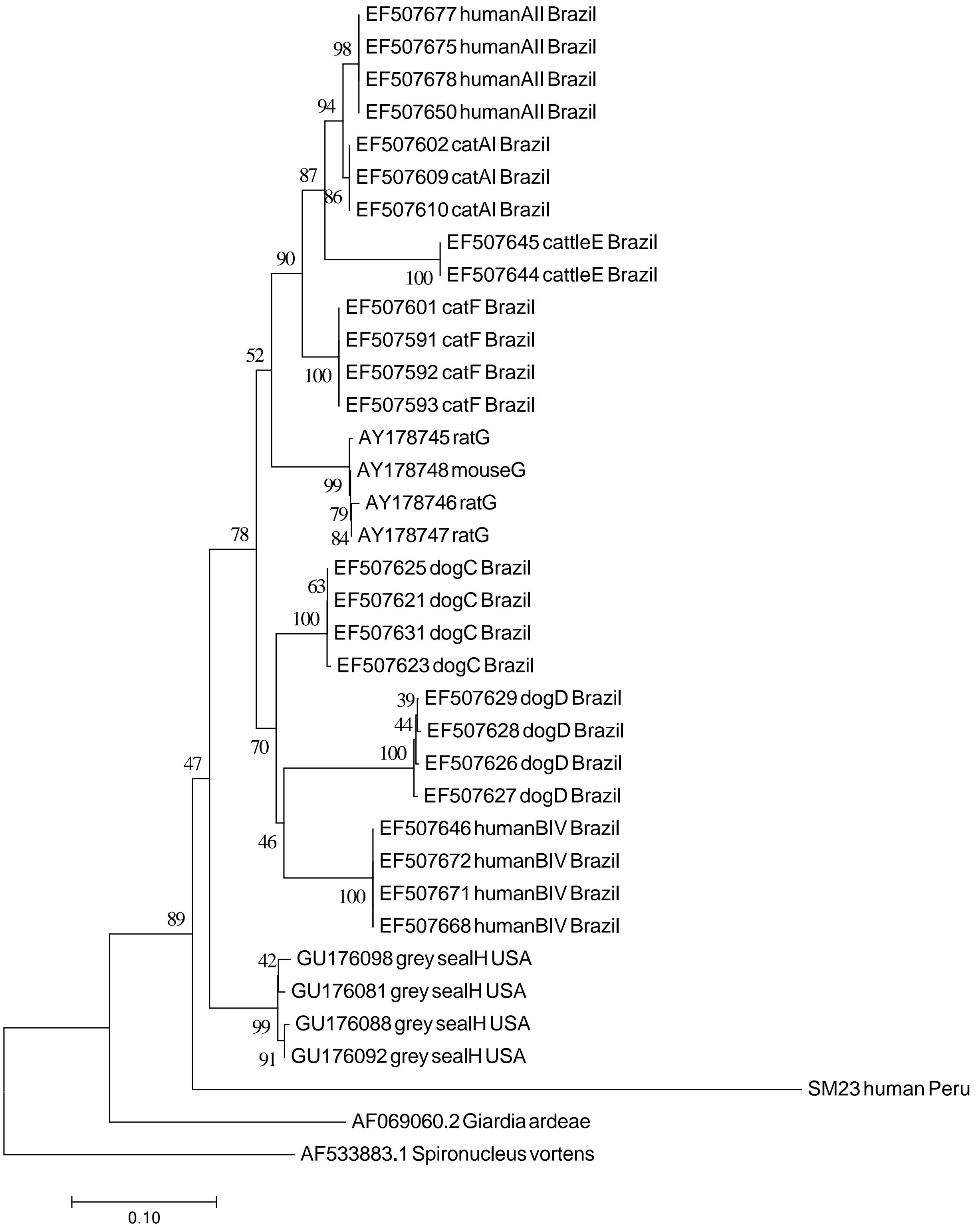

Figure 4. Neighbor-Joining tree with version five of the SM23 sequence. The Kimura 2-parameter model and 500 bootstrap replicates were used. The scale indicates substitutions per site.

The final, fifth version of the SM23 sequence aimed to eliminate any bias by changing all ambiguous sites to "N". The phylogenetic relationship of SM23 to all *G. intestinalis* reference strains was then re-assessed following a similar tree building method as used previously. As with the analysis of the original SM23 sequence, this NJ phylogenetic analysis placed SM23 outside of all assemblages, with bootstrap support of 89 (Figure 4). In discordance with the phylogeny, BLAST identified this version of SM23 once again as assemblage C, with every BLAST hit as *G. intestinalis gdh* gene.

It is apparent that when ambiguities in the SM23 sequence are kept—either by the original ambiguity codes or in the form of "N"—BLAST recognizes the sequence better as *G. intestinalis* generally, and assemblage C particularly, compared to when it is tailored to match one or more of the *G. intestinalis* assemblages. In contrast, phylogenetically the ambiguities result in SM23 being placed outside of known *G. intestinalis* assemblages.

A further analysis of the SM23 sequence in excel determined that after position 278, the number of ambiguity sites significantly increases, suggesting less reliable data. When truncated, the 278 bp SM23 sequence is still identified as *G. intestinalis* assemblage C in BLAST. However, a NJ phylogenetic analysis places our sequence in a clade with assemblage H (bootstrap=51) (Figure 5). While removing the ambiguous sites after position 278 does not impact the BLAST designation of our amplicon, it does affect the phylogeny, placing the isolate in a clade with published *G. intestinalis* assemblages. This further confirms that the SM23 amplicon is *G. intestinalis*; however, a relatively low bootstrap support value precludes any confident conclusion about its true assemblage.

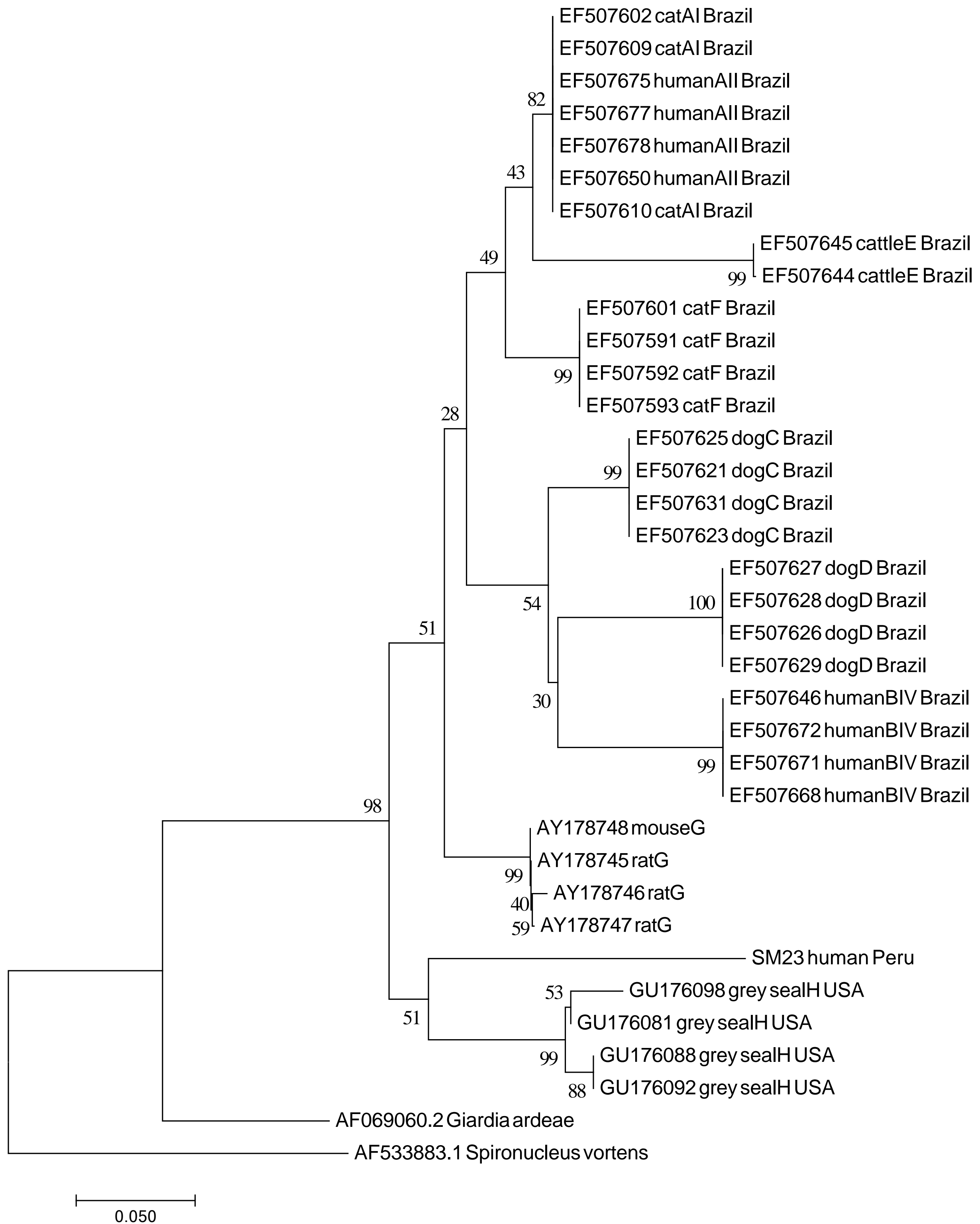

Figure 5. Neighbor-Joining tree with the truncated, 278 bp SM23 sequence. The Kimura 2-parameter model and 500 bootstrap replicates were used. The scale indicates substitutions per site.

Discussion

Methodology

Our study was greatly challenged by the quality of our data. The hybrid-sequencing approach was designed to better explore the non-specific amplifications produced by targeted amplification of the *glutamate dehydrogenase* gene, and to improve the sequence quality of our desired amplified *gdh* regions suspected to be *Giardia intestinalis*. While the theoretical basis of this approach is sound, it was unsuccessful in practice.

Sanger sequencing is only able to produce one sequence for each sample, making it difficult to separate non-specific amplifications from our intended amplicon in the sequence. In contrast, next-generation sequencing with the Illumina MiSeq uses a massively parallel approach, sequencing millions of fragments simultaneously. Not only will our intended amplicon be sequenced, but additionally any other DNA fragments that were amplified in the sample will have a separate sequence. Naturally, this improves the sequence quality of our desired *gdh* region by removing the non-specific interference bioinformatically. During sample preparation, each sample received unique indices that should have attached to each amplified DNA fragment in the sample. These indices enable the pooled sequences to be separated and organized by their unique barcodes during the data analysis portion of the MiSeq run, which in turn informs us as to which sample each sequence belongs. Ultimately, this process should create a depiction of precisely which organisms are present in each of our samples.

When the hybrid approach was put into practice, it did not produce the intended results. The expected *G. intestinalis gdh* amplicon lengths of greater than 500 bp and our selected read length of 2 x 250 should have precluded any overlap of forward and reverse reads during the

Illumina MiSeq run. However, the merged reads proportion ranged from 60-95%, indicating that the amplicons read were shorter than our targeted amplicons, thus our amplicons were not read. Further confirmation was that the C6 *G. intestinalis* positive control read did not map to the *G. intestinalis* genome, suggesting it was not properly sequenced. Ideally, there should be some overlap—about 30-50 bp—of the forward and reverse reads, to ensure the full amplicon is sequenced. Yet, when fragments are longer, such as our expected *G. intestinalis gdh* amplicons, the chances of overlap, and therefore complete sequencing, are less. We suggest that in future attempts of this approach, a smaller amplicon should be targeted to enable some overlap, but not so small such that there is redundancy in the reads. This would require redesigning the primers to target a region of the *gdh* gene of *G. intestinalis* that is shorter, yet still sufficiently variable within the species to enable distinguishing among the assemblages. In addition to this, the selected read length for the MiSeq could be adjusted to allow some overlap of the reads.

The primers we selected for the hybrid approach may not be optimal for reasons other than length of the amplified region as well. While the gdh1/gdh2 and gdh3/gdh4 primers are listed as "commonly used primers for the species/genotype and subtype differentiation of *Giardia* isolates in animal and human specimens" [11], it is unclear how well they can discriminate *G. intestinalis* from bacteria, archaea, and other microeukaryotes present in the human gut. The two primer sets were newly designed by Cacció et al. (2008) to amplify a fragment of the *G. intestinalis gdh* gene through nested PCR [97]. However, the fecal samples of that study had first been microscopically screened for *Giardia* cysts by immunofluorescence, and subsequently underwent DNA extraction specifically designed to increase the sensitivity of protozoan parasite DNA detection in fecal samples [98]. Thus, presumably the primer sets were used to screen DNA extracts that already primarily consisted of *G. intestinalis* DNA. In contrast,

we screened extracts that had captured all DNA present in human fecal samples, including DNA from bacteria and archaea, and we had no prior knowledge as to whether the samples contained *G. intestinalis*.

It may be that the gdh1/gdh2 and gdh3/gdh4 primer sets are optimized for genotyping *G. intestinalis* isolates in order to discriminate assemblages and subassemblages, but are not specific enough to distinguish *G. intestinalis* from other gut microorganisms with homologous genes. As described above, these primer sets were designed to amplify *G. intestinalis* in the absence of other microorganisms typically present in the human gut. Thus, they likely do not account for recurrent mutations that may have occurred over evolutionary time in that region of the *gdh* gene of other microorganisms, making that region homologous and preventing discrimination of *Giardia*. While we did seek primers that could discriminate among the various assemblages, primer sets targeting smaller, more *G. intestinalis*-specific regions may be more ideal for the initial detection of *G. intestinalis* from fecal extracts using PCR, as well as for NGS as described above. The BLAST results from SM29 and SM43 sequence queries included *Giardia intestinalis*, thus it is possible that the microeukaryote is present in these individuals and further optimization to our methodology by designing new primers is needed to unambiguously detect it.

Another issue of our method may lie in the choice of gene. The *gdh* gene (along with SSU rRNA, *tpi, ef1α,* and *bg*) is among the five most common genes targeted for the differentiation of *Giardia* at the species/assemblage and genotype levels and the identification of *G. intestinalis* genotypes in clinical samples [11]. These loci vary in terms of substitution rates, and *gdh* is in the middle on the conserved to variable range, with a 0.06 substitutions per nucleotide rate [99]. While this allows the *gdh* gene to have broad applications, with a traditional PCR-based approach alone it may be too conserved to distinguish *G. intestinalis* from other

common microorganisms in the human gut. According to BLAST, *Ruminococcus champanellensis, Flavonifractor plautii,* and *Bifidobacterium adolescentis* were the other top organisms we amplified in our PCR-based approach, and upon further inquiry into the NCBI database, it was the *gdh* gene specifically being amplified in each organism. Other studies on *G. intestinalis* in traditional populations used a multilocus approach for amplification and genotyping of *G. intestinalis* [59,57,56,34]. Thus, targeting more than one of the five common genes listed above may be a way to optimize our method in the future.

Giardia intestinalis in the Matses

The resolution of our data confirms the presence of *Giardia intestinalis* in our hunter-gatherer population from the Peruvian Amazon. Unambiguously, BLAST results identified the amplicon from an asymptomatic, healthy seven-year-old male (SM23) as the *glutamate dehydrogenase* gene of *G. intestinalis*. However, there is uncertainty surrounding to which assemblage the amplicon belongs. Phylogenetic analysis of the original, unedited sequence places SM23 outside of known, published *G. intestinalis* assemblages, as does changing all ambiguity codes to “N". Yet, the removal of many ambiguities after position 278 causes SM23 to form a clade with assemblage H, and the sequence forms a clade with assemblage H or C when the ambiguities are tailored to match *G. intestinalis* assemblages. In contrast to all but one of these phylogenies, BLAST consistently identifies all versions of the SM23 sequence as assemblage C.

There are two competing explanations for this pattern of uncertainty: our *G. intestinalis* isolate is outside of known assemblages, or homologous genes confounded the data. The first possibility relates to research and database biases that have favored urban settings and left gaps

in data from much of South America. It is likely from BLAST results and phylogenetically informative markers that the SM23 isolate is not assemblage A or B. Yet, the majority (60%) of reported *G. intestinalis* assemblage data worldwide is assemblage A or B [62], which significantly limits the number of varied assemblage reference sequences to which our sequence can be compared. We currently only know of two studies that found an assemblage of *G. intestinalis* other than A or B in humans in South America, both of which were located in metropolitan regions of Brazil. There were 15 cases of assemblage E in a Rio de Janeiro nursery [8] and a few cases of assemblage C in the city of Campinas [51]. As the worldwide assemblage data from Garcia-R. et al. (2017) showed, the majority of the *G. intestinalis* data in South America comes from Brazil. Thus, there are likely various *G. intestinalis* assemblages present in humans of other parts of South America that have gone unreported. This could be especially true for hunter-gatherer and other traditional populations, whose lifestyles and gut microbiomes differ significantly from urban and rural industrialized societies that are more frequently studied.

A study of the genetic diversity of *G. intestinalis* obtained from humans in three different Brazilian biomes found higher genetic diversity of assemblage B in the Amazon rainforest [100]. Considering the *bg* (*beta-giardin*) genetic target, 23 previously undescribed assemblage B sequences were identified, which were considered new subassemblages of B. Furthermore, subassemblage AIII was detected only in the Amazon [100], whereas an extensive study of primarily European isolates and about 45% GenBank sequences did not find subassemblage AIII in any humans [50]. The Brazilian biome study attributed the higher genetic diversity of assemblage B in the Amazon to demographic and socioenvironmental differences between it and the other two Brazilian regions, though notably the Amazonian population was of an urban

nature [100]. Furthermore, there was considerable genetic diversity of *G. intestinalis* subassemblages in children of a traditional society in the Colombian Amazon [34].

Considering multiple new subassemblages of B and the rare AIII subassemblage were recently identified in the Brazilian Amazon, and that there is a great diversity of subassemblages in the Colombian Amazon, further *G. intestinalis* genetic diversity is likely in the Peruvian Amazon, especially in traditional populations. Our *G. intestinalis* isolate from a hunter-gatherer society in the Peruvian Amazon was identified as assemblage C when queried against the NCBI nucleotide database, but placed outside of known assemblages in the unbiased phylogenetic analyses, suggesting it could be a novel subassemblage. This subassemblage could belong to assemblage C or H according to the nucleotide options suggested by ambiguity codes, but phylogenetic uncertainty precludes a concrete answer.

Homologous genes could alternatively be confounding the data, preventing SM23 from forming a definitive clade with its true assemblage in the phylogenetic analysis. In the other Matses and Burkina Faso samples, we amplified regions of the *gdh* gene from other organisms (namely *Ruminococcus champanellensis, Flavonifractor plautii,* and *Bifidobacterium adolescentis*). Additionally, when ambiguities in the SM23 sequence were changed to a specific nucleotide, the top BLAST hit was consistently *Methanomassiliicoccaceae archaeon*. Thus, it is plausible that DNA from the *gdh* gene of *M. archaeon* is present in our amplicon and precluding its full taxonomic resolution.

Conclusions and Future Directions

The hybrid approach of traditional PCR-based amplification and next-generation sequencing has potential to successfully capture *G. intestinalis* genetic diversity from human

fecal samples. We confirmed the presence of *Giardia intestinalis* in the Matses population, a hunter-gatherer community from the Peruvian Amazon. However, the quality of our data, as well as probable database bias, limited our ability to confidently define the assemblage or confirm the presence of *G. intestinalis* in other samples. Given the precedence in the literature for both zoonotic transmission of assemblages C, D, E, and F, and higher genetic diversity of *G. intestinalis* in the Amazon, it is reasonable to believe our SM23 isolate from the Matses community could be a novel subassemblage of *G. intestinalis*, possibly of assemblage C.

Further optimization of the hybrid approach beyond the scope of this study is necessary to better detect and genotype *G. intestinalis* in the Matses and Burkina Faso communities. In particular, primers should be redesigned to target a shorter, more *Giardia*-specific region of the *gdh* gene, or target a different gene entirely. Additionally, a multilocus genotyping approach may produce more robust results. To our knowledge, this was the first study to assess the genetic diversity of *G. intestinalis* in a hunter-gatherer population of Peru. More research should be conducted in traditional communities of Peru specifically and South America in general to fill the gap in reported *G. intestinalis* assemblages in humans. It is likely that hunter-gatherer populations may harbor a greater diversity of assemblages in a commensal state, and more insight into this possibility will reveal much about the true nature of *Giardia intestinalis* in the human gut microbiome.

CHAPTER 5

REFERENCES CITED

1 Hollm-Delgado MG et al. 2008. Lack of an adverse effect of Giardia intestinalis infection on the health of Peruvian children. Am J Epidemiol 168:647–655.

2 Machicado JD et al. 2012. Diagnosis of soil-transmitted helminthiasis in an Amazonic community of Peru using multiple diagnostic techniques. Trans R Soc Trop Med Hyg 106:333–339.

3 Erismann S et al. 2016. Prevalence of intestinal parasitic infections and associated risk factors among schoolchildren in the Plateau Central and Centre-Ouest regions of Burkina Faso. Parasit Vectors 9:554.

4 Gelanew T et al. 2007. Molecular characterization of human isolates of Giardia duodenalis from Ethiopia. Acta Trop 102:92–99.

5 Sulaiman IM et al. 2003. Triosephosphate Isomerase Gene Characterization and Potential Zoonotic Transmission of Giardia duodenalis. Emerg Infect Dis 9:1444–1452.

6 Traub RJ et al. 2009. Transmission cycles of Giardia duodenalis in dogs and humans in Temple communities in Bangkok—A critical evaluation of its prevalence using three diagnostic tests in the field in the absence of a gold standard. Acta Trop 111:125–132.

7 Soliman RH et al. 2011. Identification of a novel Assemblage B subgenotype and a zoonotic Assemblage C in human isolates of Giardia intestinalis in Egypt. Parasitol Int 60:507–511.

8 Fantinatti M et al. 2016. Identification of Giardia lamblia Assemblage E in Humans Points to a New Anthropozoonotic Cycle. J Infect Dis 214:1256–1259.

9 Plutzer J et al. 2010. Giardia taxonomy, phylogeny and epidemiology: Facts and open questions. Int J Hyg Environ Health 213:321–333.

10 Adam RD. 2000. The Giardia lamblia genome. Int J Parasitol 30:475–484.

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

12 Hashimoto T et al. 1998. Secondary absence of mitochondria in Giardia lamblia and Trichomonas vaginalis revealed by valyl-tRNA synthetase phylogeny. Proc Natl Acad Sci 95:6860–6865.

13 Roger AJ et al. 1998. A mitochondrial-like chaperonin 60 gene in Giardia lamblia: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci National Academy of Sciences. 95:229–234.

14 Franzén O et al. 2009. Draft Genome Sequencing of Giardia intestinalis Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species? PLoS Pathog 5.

15 Hashimoto T et al. 1995. Phylogenetic Place of Mitochondrion-Lacking Protozoan, Giardia Zamblia, Inferred from Amino Acid Sequences of Elongation Factor 2. Mol Biol Evol 12(5):782-793.

16 Siddall ME et al. 1992. Phylogenetic Analysis of the Diplomonadida (Wenyon, 1926) Brugerolle, 1975: Evidence for Heterochrony in Protozoa and Against Giardia lamblia as a "Missing Link." J Protozool 39:361–367.

17 Thompson RCA and Monis PT. 2011. Taxonomy of *Giardia* Species. In: Luján HD and Svärd S, editors. Giardia: a model organism. Wien ; New York: Springer. p 3-16.

18 Kolisko M et al. 2008. Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: Implications for the evolutionary history of the double karyomastigont of diplomonads. BMC Evol Biol 8:205.

19 Keeling PJ, Brugerolle G. 2006. Evidence from SSU rRNA Phylogeny that Octomitus is a Sister Lineage to Giardia. Protist 157:205–212.

20 Reses HE et al. 2018. Risk factors for sporadic Giardia infection in the USA: a case-control study in Colorado and Minnesota. Epidemiol Infect 146:1071–1078.

21 Nazer H. 2018. Giardiasis. Medscape website, Oct 01. https://emedicine.medscape.com/article/176718-overview

22 Centers for Disease Control and Prevention. 2018. Giardia: Pathogen & Environment. https://www.cdc.gov/parasites/giardia/pathogen.html

23 Graczyk TK et al. 1998. Detection of Cryptosporidium oocysts and Giardia cysts in the tissues of eastern oysters (Crassostrea virginica) carrying principal oyster infectious diseases. J Parasitol 84:1039–1042.

24 Schets FM et al. 2007. Cryptosporidium and Giardia in commercial and non-commercial oysters (Crassostrea gigas) and water from the Oosterschelde, The Netherlands. Int J Food Microbiol 113:189–194.

25 Gomez-Couso H et al. 2005. Occurrence of giardia cysts in mussels (Mytilus galloprovincialis) destined for human consumption. J Food Prot 68:1702–1705.

26 Colli CM et al. 2015. Food handlers as a link in the chain of transmission of Giardia duodenalis and other protozoa in public schools in southern Brazil. Trans R Soc Trop Med Hyg 109:601–603.

27 Utaaker KS et al. 2017. Checking the detail in retail: Occurrence of Cryptosporidium and Giardia on vegetables sold across different counters in Chandigarh, India. Int J Food Microbiol 263:1–8.

28 Barash NR et al. 2017. Giardia Alters Commensal Microbial Diversity throughout the Murine Gut. Infect Immun 85(6) e00948-16.

29 Centers for Disease Control and Prevention (CDC). 2019. Giardiasis Summary Report--National Notifiable Diseases Surveillance System, United States, 2017. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC.

30 European Centre for Disease Prevention and Control. 2019. Giardiasis (lambliasis). In: ECDC. Annual epidemiological report for 2017. Stockholm: ECDC.

31 Dib HH et al. 2008. Prevalence of Giardia lamblia with or without diarrhea in South East, South East Asia and the Far East. Parasitol Res 103:239–251.

32 Rivero M et al. 2018. Intestinal parasitism and nutritional status among indigenous children from the Argentinian Atlantic Forest: Determinants of enteroparasites infections in minority populations. Acta Trop 187:248–256.

33 Volotão AC et al. 2007. Genotyping of Giardia duodenalis from human and animal samples from Brazil using β-giardin gene: A phylogenetic analysis. Acta Trop 102(1):10-19.

34 Sanchez A et al. 2017. Molecular Epidemiology of Giardia, Blastocystis and Cryptosporidium among Indigenous Children from the Colombian Amazon Basin. Front Microbiol 8:248.

35 Squire SA, Ryan U. 2017. Cryptosporidium and Giardia in Africa: current and future challenges. Parasit Vectors 10.

36 de Lucio A et al. 2016. Prevalence and Genetic Diversity of Giardia duodenalis and Cryptosporidium spp. among School Children in a Rural Area of the Amhara Region, North-West Ethiopia. PLoS ONE 11.

37 Firdu T et al. 2014. Intestinal Protozoal Parasites in Diarrheal Children and Associated Risk Factors at Yirgalem Hospital, Ethiopia: A Case-Control Study. Int Sch Res Not 2014:1–8.

38 Cernikova L et al. 2018. Five facts about Giardia lamblia. PLoS Pathog 14:e1007250.

39 DuPont HL. 2013. Giardia: both a harmless commensal and a devastating pathogen. J Clin Invest 123:2352–2354.

40 Savioli L et al. 2006. Giardia and Cryptosporidium join the “Neglected Diseases Initiative.” Trends Parasitol 22:203–208.

41 Rogawski ET et al. 2017. Determinants and Impact of Giardia Infection in the First 2 Years of Life in the MAL-ED Birth Cohort. J Pediatr Infect Soc 6:153–160.

42 Bartelt LA, Platts-Mills JA. 2016. Giardia: a pathogen or commensal for children in high-prevalence settings? Curr Opin Infect Dis 29:502–507.

43 Liu J et al. 2014. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. Lancet Infect Dis 14:716–724.

44 Sullivan PS et al. 1988. Illness and reservoirs associated with Giardia lamblia infection in rural Egypt: the case against treatment in developing world environments of high endemicity. Am J Epidemiol 127:1272–1281.

45 Singer SM, Nash TE. 2000. The Role of Normal Flora in *Giardia lamblia* Infections in Mice. J Infect Dis 181:1510–1512.

46 Jerez Puebla LE et al. 2017. Molecular analysis of Giardia duodenalis isolates from symptomatic and asymptomatic children from La Habana, Cuba. Parasite Epidemiol Control 2:105–113.

47 Rahimian F et al. 2018. A modified PCR-RFLP method to determine genetic diversity of Giardia lamblia human isolates based on triosephosphate isomerase (TPI) gene. Acta Trop 186:58–62.

48 Peréz Cordón G et al. 2008. Prevalence of enteroparasites and genotyping of Giardia lamblia in Peruvian children. Parasitol Res 103:459–465.

49 Cacciò SM et al. 2005. Unravelling Cryptosporidium and Giardia epidemiology. Trends Parasitol 21:430–437.

50 Sprong H et al. 2009. Identification of Zoonotic Genotypes of Giardia duodenalis. PLoS Negl Trop Dis 3.

51 Durigan M et al. 2017. Population genetic analysis of Giardia duodenalis: genetic diversity and haplotype sharing between clinical and environmental sources. Microbiologyopen 6.

52 Foronda P et al. 2008. Identification of genotypes of Giardia intestinalis of human isolates in Egypt. Parasitol Res 103:1177–1181.

53 Abdel-Moein KA, Saeed H. 2016. The zoonotic potential of Giardia intestinalis assemblage E in rural settings. Parasitol Res 115:3197–3202.

54 Zahedi A et al. 2017. Molecular typing of Giardia duodenalis in humans in Queensland - first report of Assemblage E. Parasitology 144:1154–1161.

55 Pipikova J et al. 2018. First report on Giardia duodenalis assemblage F in Slovakian children living in poor environmental conditions. J Microbiol Immunol Infect 53:148-156.

56 Choy SH et al. 2014. Prevalence and Associated Risk Factors of Giardia Infection among Indigenous Communities in Rural Malaysia. Sci Rep 4:6909.

57 Huey CS et al. 2013. Multilocus genotyping of Giardia duodenalis in Malaysia. Infect Genet Evol 17:269–276.

58 Anuar TS et al. 2012. Giardiasis among different tribes of Orang Asli in Malaysia: highlighting the presence of other family members infected with Giardia intestinalis as a main risk factor. Int J Parasitol 42:871–880.

59 Saksirisampant W et al. 2012. Prevalence of giardiasis and genotypic characterization of Giardia duodenalis in hilltribe children, Northern Thailand. Trop Biomed 29(3):331-338.

60 Liang C et al. 2012. Occurrence and genotype of Giardia cysts isolated from faecal samples of children and dogs and from drinking water samples in an aboriginal area of central Taiwan. Exp Parasitol 131:204–209.

61 Anselmi M et al. 2019. Community Epidemiology Approach to Parasitic Infection Screening in a Remote Community in Ecuador. Am J Trop Med Hyg The American Society of Tropical Medicine and Hygiene 101:650–653.

62 Garcia–R JC et al. 2017. Local and global genetic diversity of protozoan parasites: Spatial distribution of Cryptosporidium and Giardia genotypes. PLoS Negl Trop Dis 11(7):e0005736.

63 The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

64 Lederberg J, Mccray AT. 2001. `Ome Sweet `Omics--A Genealogical Treasury of Words. The Scientist 15:8–8.

65 Blaut M. 2018. Composition and Function of the Gut Microbiome. In: Haller D, editor. Gut Microbiome Health Dis. Cham: Springer International Publishing. p 5–30.

66 Haller D. 2018. Intestinal Microbiome in Health and Disease: Introduction. In: Haller D, editor. Gut Microbiome Health Dis. Cham: Springer International Publishing. p 1–3.

67 Metcalf AM et al. 1987. Simplified assessment of segmental colonic transit. Gastroenterology 92:40–47.

68 Hoffmann C et al. 2013. Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. PLoS ONE 8(6):e66019.

69 Laforest-Lapointe I, Arrieta M-C. 2018. Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies. mSystems 3:e00201-17.

70 Minot S et al. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res 21:1616–1625.

71 Eckburg PB et al. 2005. Diversity of the Human Intestinal Microbial Flora. Science 308:1635–1638.

72 Frank DN et al. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci USA 104:13780–13785.

73 Yatsunenko T et al. 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227.

74 Turnbaugh PJ et al. 2009. A core gut microbiome in obese and lean twins. Nature 457:480–484.

75 Strachan DP. 1989. Hay fever, hygiene, and household size. BMJ 299:1259–1260.

76 Bloomfield S et al. 2006. Too clean, or not too clean: the Hygiene Hypothesis and home hygiene. Clin Exp Allergy 36:402–425.

77 Rook GAW et al. 2014. Microbial 'old friends', immunoregulation and socioeconomic status. Clin Exp Immunol 177:1–12.

78 Bellanti JA, Settipane RA. 2017. The atopic disorders and atopy … "strange diseases" now better defined! Allergy Asthma Proc 38:241–242.

79 Beasley R. 1998. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The Lancet 351:1225–1232.

80 Bjorksten et al. 1999. The intestinal microflora in allergic Estonian and Swedish 2-year-old children. Clin Exp Allergy 29:342–346.

81 Matricardi PM et al. 2000. Exposure to foodborne and orofecal microbes versus airborne viruses in relation to atopy and allergic asthma: epidemiological study. BMJ 320:412–417.

82 Reynolds LA et al. 2015. Cohabitation in the Intestine: Interactions among Helminth Parasites, Bacterial Microbiota, and Host Immunity. J Immunol 195:4059–4066.

83 Scrivener S et al. 2001. Independent effects of intestinal parasite infection and domestic allergen exposure on risk of wheeze in Ethiopia: a nested case-control study. The Lancet 358:1493–1499.

84 Parfrey LW et al. 2011. Microbial Eukaryotes in the Human Microbiome: Ecology, Evolution, and Future Directions. Front Microbiol 2.

85 Lukeš J et al. 2015. Are Human Intestinal Eukaryotes Beneficial or Commensals? Knoll LJ, editor. PLOS Pathog 11:e1005039.

86 Audebert C et al. 2016. Colonization with the enteric protozoa Blastocystis is associated with increased diversity of human gut bacterial microbiota. Sci Rep Nature Publishing Group. 6:1–11.

87 Beghini F et al. 2017. Large-scale comparative metagenomics of Blastocystis , a common member of the human gut microbiome. ISME J Nature Publishing Group. 11:2848–2863.

88 Lawson PA et al. 2016. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938. Anaerobe 40:95–99.

89 diCenzo GC, Finan TM. 2017. The Divided Bacterial Genome: Structure, Function, and Evolution. Microbiol Mol Biol Rev 81:e00019-17.

90 Cacciò SM et al. 2010. Identification of *Giardia* species and *Giardia duodenalis* assemblages by sequence analysis of the 5.8S rDNA gene and internal transcribed spacers. Parasitology 137:919–925.

91 Altschul SF et al. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

92 Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 22:939–946.

93 Schubert M et al. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88.

94 Buchfink B et al. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods Nature Publishing Group. 12:59–60.

95 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods Nature Publishing Group. 9:357–359.

96 Kumar S et al. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.

97 Cacciò SM et al. 2008. Multilocus genotyping of Giardia duodenalis reveals striking differences between assemblages A and B. Int J Parasitol 38:1523–1531.

98 Da Silva AJ et al. 1999. Fast and reliable extraction of protozoan parasite DNA from fecal specimens. Mol Diagn 4:57–64.

99 Wielinga CM, Thompson RCA. 2007. Comparative evaluation of *Giardia duodenalis* sequence data. Parasitology 134:1795–1821.

100 Nunes BC et al. 2018. Genetic diversity of Giardia duodenalis circulating in three Brazilian biomes. Infect Genet Evol 59:107–112.

SUPPLEMENTAL DATA

Supplemental Table 1. MiSeq Results

Sample	Raw Reads	Analysis-ready reads	% of reads merged	Number of reads mapped to *G. intestinalis* genome	Number of unique reads mapped to *G. intestinalis* genome	Number of reads mapping to the *G. intestinalis gdh* gene
C6	7379	6588	89.29	0	0	0
SM10	28645	19434	67.85	5	3	0
SM11	376053	282703	75.18	31	12	0
SM18	86970	75167	86.43	5	3	0
SM1	250225	226412	90.49	5	5	0
SM20	115443	104711	90.71	15	7	0
SM23	220281	182823	83	57	4	0
SM28	103404	66586	64.4	14	4	0
SM29	94655	58710	62.03	6	4	0
SM2	29538	26959	91.27	1	1	0
SM31	246140	220980	89.78	2	1	0
SM32	1056431	822859	77.9	99	20	0
SM33	71819	64051	89.19	8	3	0
SM37	25497	23145	90.78	0	0	0
SM39	78685	70299	89.35	2	2	0
SM3	112756	81863	72.61	3	2	0
SM40	222122	199184	89.68	16	8	0
SM41	91813	80588	87.78	4	2	0
SM43	14232	12985	91.24	1	1	0
SM44	41111	37199	90.49	3	2	0
TM10-2	722695	650411	90	3	2	0
TM10-3	2054572	1899661	92.47	0	0	0
TM10-4	1752938	1633473	93.19	2	1	0
TM11-1	2249066	2127916	94.62	7	5	0
TM1-2	216730	204294	94.27	0	0	0
TM1-3	23565	18172	77.12	0	0	0
TM1-4	84776	73461	86.66	0	0	0
TM2-2	1098615	1028203	93.6	3	3	0
TM2-3	413915	345877	83.57	90	19	0
TM2-4	180877	172556	95.4	0	0	0
TM4-3	72663	59902	82.44	0	0	0
TM4-4	30956	28355	91.6	0	0	0

TM5-1	8878	8095	91.19	3	2	0
TM5-3	36690	26939	73.43	8	1	0
TM6-1	160813	104339	64.89	0	0	0
TM6-2	631639	483100	76.49	113	3	0
TM6-3	793690	722301	91.01	3	3	0
TM7-2	345827	300960	87.03	1	1	0
TM7-4	317	290	91.49	0	0	0
TM8-2	1568	1490	95.03	0	0	0
TM8-4	57479	51260	89.19	0	0	0
TM9-1	55682	47496	85.3	0	0	0
TM9-2	72375	59609	82.37	0	0	0
TM9-3	20578	18469	89.76	8	6	0
TM9-4	1024372	921514	89.96	3	1	0