Title: Molecular epidemiology of *Giardia* infections in the genomic era

Authors: Paul Capewell[1], Sarah Krumrie[2], Frank Katzer[3], Claire Alexander[4], William Weir*[2]

Affiliations:

[1] Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, 464 Bearsden Road, Glasgow, G61 1QH, UK

[2] School of Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, 464 Bearsden Road, Glasgow, G61 1QH, UK

[3] Moredun Research Institute, Pentlands Science Park, Edinburgh, EH26 0PZ, UK

[4] Scottish Parasitology Diagnostic and Reference Laboratories, Glasgow, G31 2ER, UK

*Correspondence: willie.weir@glasgow.ac.uk (W. Weir)

Keywords:

Giardia, genotyping, genomics, molecular epidemiology, population genetics, public health

Abstract

Giardia duodenalis is a major gastrointestinal parasite of humans and animals across the globe. It is also of interest from an evolutionary perspective as it possesses many features that are unique among the eukaryotes, including its distinctive binucleate cell structure. While genomic analysis of a small number of isolates has provided valuable insights, efforts to understand the epidemiology of the disease and the population biology of the parasite have been limited by the molecular tools currently available. We review these tools and assess the impact of affordable and rapid genome sequencing systems increasingly being deployed in diagnostic settings. While these technologies have direct implications for public and veterinary health, they will also improve our understanding of the unique biology of this fascinating parasite.

A major worldwide pathogen

Giardia duodenalis (also known as *Giardia intestinalis* or *Giardia lamblia*) is one of the most common gastrointestinal parasites in the world, causing an estimated 180 million infections annually [1]. Although giardiasis is treatable, the correct administration of therapeutics depends on accurately identifying the parasite, either in an individual or within a community during an outbreak. Asymptomatic infection can occur and may represent the majority of cases [2–5], although it should be noted that apparently asymptomatic individuals can still display signs of mild malnutrition [6]. Patients with overt clinical disease experience severe gastrointestinal disturbances for several weeks due to **trophozoites** (see Glossary) attaching to the intestinal lining of the host, disrupting nutrient and water uptake and eliciting an immune response [7]. In rare cases, some patients can develop post-infection complications that lead to long-term gastrointestinal disorders similar to irritable bowel syndrome (IBS) [8]. These symptoms are linked to a loss of barrier function and **dysbiosis** of the gut flora [9–11]. Infective **cysts** are shed into the environment by infected hosts where they can be ingested by new individuals, maintaining transmission. Outbreaks are frequently associated with contaminated water [12] or food sources [13].

Giardia genetics

Giardia species are described as early divergent eukaryotes and lack common subcellular structures such as mitochondria, a true Golgi complex and **peroxisomes** [14]. However, the identification of mitochondrial genes in the genome suggests that *G. duodenalis* and other **diplomonads** once possessed these organelles but subsequently lost them [15]. In addition to being amitochondriate, *G. duodenalis* is unusual in that it possesses two nuclei despite being unicellular. This gives rise to an unusual **ploidy** throughout the life-cycle where the

trophozoites cycle between 4N (2×2N) and 8N (4×2N) in vegetative growth, doubling to 16N (4×4N) during the transition to cysts (Figure 1A) [16]. After excystation, the cell divides without DNA replication to create four trophozoites, each with a ploidy of 4N. The relatively small genome (11.7 megabases (Mb)) is distributed across five chromosomes that feature few intergenic spaces, introns or non-coding regions. Promoters and untranslated regions are also minimized, leading to a highly condensed and efficient genome [14]. This appears to be a distinctive trait of the diplomonads and the genome is even more compact in the closely related rodent parasite *G. muris* [17]. Although similar, the genomes of each nucleus in an individual parasite are not identical and the differences between the four different genomes are termed **allelic sequence heterozygosity** (ASH). The proportion of heterozygous bases within a genome typically ranges between 0.25–0.74 % for most *Giardia* isolates [18–20] but can be extremely low (<0.01%) [9,11]. The majority of these heterozygous sites contain two different bases but some feature three or four, capturing the diversity present across all of the copies of the genome [20]. Regions of heterozygosity are not distributed evenly throughout the genome and more typically occur in non-coding regions, as might be expected with purifying selection acting on coding regions [20].

Genotyping *Giardia duodenalis*

Initial **isozyme** and **18S ssu-rRNA gene** sequencing demonstrated that two broad groups of *G. duodenalis* infected human patients (eventually termed A and B). These were characterized as assemblages to reflect the fact that the relationships between the groups were undefined [22–24]. Additional genetic data from animal-derived isolates, coupled with various biological differences, allowed a further six distinct assemblages to be differentiated (C–H) (Table 1). Assemblages A and B contain zoonotic isolates that can infect humans and

71 animals, whereas assemblages C–H show specificity to particular animal hosts (Table 1).
72 However, isolates with molecular sequences similar to assemblages C–H have been
73 amplified from humans, suggesting there are either limits to current molecular typing tools
74 or *G. duodenalis* may have a wider zoonotic potential than first assumed [25–27]. As these
75 molecular detections in humans are often from asymptomatic individuals, it is unclear
76 whether the DNA detected in the molecular screens represents infection, carriage or
77 contamination. Antigen-capture assay, immunofluorescent antibody testing (IFAT), direct
78 microscopy and ssu-rRNA quantitative polymerase chain reaction (qPCR) are the standard
79 methods for detecting *G. duodenalis*, although many diagnostic laboratories rely on
80 microscopy for detection due to cost and established pipelines [28]. Microscopy can lack
81 sensitivity when parasitaemia is low (which is common) or where expertise is lacking,
82 indicating a switch to immunological assays or qPCR in diagnostic settings may be preferable
83 to assess accurately the presence of the parasite [29]. However, it should be appreciated
84 that although qPCR sensitively detects *Giardia* nucleic acid, a positive test result does not
85 necessarily confirm the presence of viable parasite cells. Routine qPCR-based detection
86 methods are also unsuitable for determining the relationships between isolates as they do
87 not provide detailed genetic information [30]. Over the years, several molecular markers
88 have been developed to create a **multilocus sequence typing** (**MLST**) panel to investigate
89 the molecular phylogeny of *G. duodenalis.* These involve targeted PCR and subsequent
90 sequencing of genes that are relatively stable but possess some degree of variability for
91 differentiating isolates. The most commonly used regions of the genome are the loci
92 encoding β-giardin (*bg*) [31], triosephosphate isomerase (*tpi*) [32] and glutamate
93 dehydrogenase (*gdh*) [33,34]. Of these three primary markers, *tpi* displays the most
94 polymorphisms in the currently sequenced population in terms of substitutions per

nucleotide site (π = 0.12) and *bg* the least (π = 0.03), with *gdh* intermediate between the two (π = 0.06) [35]. There are also some rarely used loci that are more difficult to amplify than the typical markers, including the internal transcribed spacer (*its1*) and elongation factor 1 (*ef1*) [36,37]. In common with other parasite species, PCR and sequencing of the ssu-rRNA region is also employed for molecular genotyping [35]. Amplification success for this region can be higher than other PCR targets due to being multi-copy, making it highly sensitive and ideal for identification [38]. However, single-copy markers are still commonly employed due to the relatively low discriminatory power of the ssu-rRNA region [35] and unusually high GC content that can lead to issues with specificity [39,40]. Further analysis has shown that assemblage B isolates display greater polymorphism than other strains at the commonly used marker sites, possessing higher ASH within individual parasites and greater allelic diversity in the population [41,42]. Infections with multiple assemblages [43–45] or sub-assemblages [46] can also occur in humans and animals, affecting estimates of heterozygosity. Furthermore, it appears that mixed genotypic infections can affect infection dynamics, such as increasing cyst shedding [44]. However, it is unclear how often mixed *Giardia* infections occur as they may only be detectable after sub-cloning and sequencing at a depth able to detect low levels of ASH [41].

Current methods provide insufficient resolution

MLST approaches using targeted PCR with sequencing of amplicons have largely validated the assemblage model, leading to the proposal that the assemblages should be formally re-defined as species [47]. However, as the current markers do not provide the resolution required to determine accurately relationships between isolates beyond assemblages, their effectiveness in the public health sphere has been limited. While putative sub-assemblages

have been defined within assemblage A (AI–III) [46], the framework describing sub-structuring within assemblage B appears less robust [35]. There are also situations in which there is a lack of concurrence between the different markers that may be due to their low discriminatory power [48,49]. For example, there is only a single nucleotide difference that discriminates between the allele subtypes AI and AII within the *bg* amplicon and two for the corresponding *tpi* subtypes [35]. These issues are clearly demonstrated by recent high-resolution analysis that used an MLST consisting of six markers to examine assemblage A isolates [46]. While delineation into three distinct sub-assemblages was supported, individual markers were less stable and showed conflicting results when examined in isolation. This was attributed to potential recombination within the population. Similar incongruities were noted in a recent study of primate *Giardia* isolates; the current assemblage model could not be reliably reproduced with ssu-rRNA data, likely due to the low resolution of the marker [50]. There are also issues with the reliability of PCR assays targeting single copy genes and it is common for only one gene to amplify. Success rates vary from 11–91% across the different markers depending on the study [38]. Mixed infections and ASH can also make it impossible to infer alleles from direct PCR sequencing and it is necessary to use laborious transformation and cloning protocols before sequencing. These are often difficult and costly to implement in a diagnostic setting. This lack of reliability is likely due in part to the large amounts of contaminating DNA and inhibitors found in faecal material, compounded by the variable number of *G. duodenalis* cysts present [51,52]. However, it is also likely that there is a degree of sequence variability present in the genes affecting primer binding sites and amplification success (Box 1 and Figure 2). It is therefore probable that a great deal of genetic diversity in the *Giardia* population is being overlooked due to the specificity of the primers used and the difficulties in amplifying from

142 faecal material. As such, the currently available marker-based system for understanding the
143 molecular phylogeny of *G. duodenalis* is limited in scope and does not provide a high level of
144 genetic discrimination. This makes it virtually impossible to identify reliably transmission
145 routes, reservoirs and relationships between strains, hampering public health efforts to
146 control giardiasis.

147 Investigating the epidemiology of *Giardia* infections

148 The lack of high-resolution genotyping tools also limits the ability to answer fundamental
149 biological questions concerning the parasite, many of which have wider effects on
150 understanding transmission and controlling disease. For example, a large number of
151 companion and livestock animals are infected with *G. duodenalis*, including assemblages
152 that can infect humans [53]. Although in some cases disease manifestation can be severe,
153 the typical clinical impact for animals appears to be low and may often not be associated
154 with clinical signs [54]. While companion animals appear more likely to be infected with
155 species-specific assemblages (C/D in dogs, F in cats), they can also be infected with
156 assemblages A and B. However, whether these actually pose a zoonotic risk is inherently
157 difficult to ascertain due to the low resolution of current markers. Several studies have
158 shown that animals and humans can share genotypes [55] and even sub-assemblages
159 [42,56–61], but incomplete MLSTs, in addition to the low resolution and incongruences
160 between markers, mean these data cannot be definitive. Indeed, many studies rely on the
161 use of a single marker, despite this being inadequate to group isolates reliably [35,41,50,62].
162 Instead, identifying *G. duodenalis* transmission between humans and animals has been
163 inferred using classical epidemiological studies and indirect observations. For example,
164 wide-scale vaccination of dogs in a deprived community in Argentina led to a corresponding

165 decrease in the prevalence of *Giardia* in the local children [63]. Additionally, epidemiological
166 analyses in India found a highly significant association between the prevalence of *G.*
167 *duodenalis* in humans, dog ownership and the presence of a *G. duodenalis*-positive dogs in
168 the same household [37]. Similar links were found between dog ownership and human
169 infection with assemblage A in a United Kingdom setting [64]. However, no link has been
170 found in other communities [65] and it likely that the epidemiology and zoonotic risk of
171 *Giardia* infections vary in different locations. This diversity of epidemiological contexts
172 underlines the need for novel high-resolution genotyping methods which can be applied to
173 reveal the particular transmission pathways in action in different areas.

174 ## Population genetic structure of *Giardia*

175 Another important aspect of *G. duodenalis* biology that cannot be resolved with the current
176 molecular tools is the role that sexual recombination plays in creating diversity in natural
177 populations of the parasite. Although seemingly an academic question, this issue is of
178 practical importance as the occurrence of genetic exchange in pathogen populations can
179 have a significant impact on disease epidemiology. For example, in asexual organisms only
180 rare mutations at specific loci or horizontal gene transfers can provide new genotypes that
181 may lead to drug resistance or increased virulence. Conversely, sexual organisms are able to
182 produce new genotypes constantly through meiosis and chromosomal re-assortment,
183 allowing alleles conferring a fitness advantage to spread in the population. This in turn
184 allows pathogens to adapt and exploit new conditions. In diagnostic and public health
185 settings, sexual recombination also affects the ability to track outbreaks and identify
186 transmission networks by disrupting the genotypes involved. Although *G. duodenalis*
187 possesses many of the genes for meiosis [66], sexual reproduction has never directly been

observed and examination of **linkage disequilibrium** among isozymes suggests *G. duodenalis* is asexual [67]. There are hints that the parasite may not be completely asexual and sexual recombination may simply be a rare event [14,21]. In particular, extensive examination of the genetics of *G. duodenalis* from a single household found evidence for the reassortment of alleles between infections, suggesting sexual recombination [68]. Similar reassortment was suggested by a high-resolution study of assemblage A isolates that also indicated possible cross-assemblage recombination with assemblage E [46]. Horizontal gene transfer has also been documented between assemblages A and B [69,70].

An alternative explanation for the low levels of variance and heterozygosity observed in *G. duodenalis* is the utilization of a **parasexual cycle** during reproduction, similar to many fungi [71]. A parasexual cycle differs from meiosis in that it involves the fusing of two diploid parent cells prior to genetic exchange rather than haploid gametes. To return to a diploid state there must be a reduction in chromosomal number after this process. Microscopic evidence has shown that within *G. duodenalis* cysts, nuclear fusion and genetic exchange can occur during the transition from the 4x2N to the 4x4N stage (Figure 1B), although without the loss of chromosomes [72]. This unique parasexual cycle could act to decrease heterozygosity within the *G. duodenalis* population, reducing the negative effects associated with deleterious mutations that accumulate in asexual eukaryotes [73]. It can also lead to the generation of new allele combinations, emulating some of the benefits of true sexual recombination. However, as this is essentially a form of self-fertilization (an extreme form of inbreeding), the system can only slow down the accumulation of mutations rather than eliminate them completely. Additionally, a parasexual cycle cannot explain the apparent recombination observed between *Giardia* isolates across assemblages and sub-assemblages [46,68–70], nor incidences of **lateral gene transfer** from bacteria and the host [14].

Alternatively, *G. duodenalis* may utilize large-scale **gene conversion** (evidenced by long-range **loss of heterozygosity (LOH)**) achieved through homologous recombination events to compensate for the build-up of deleterious mutations, similar to that described for the asexual parasite *Trypanosoma brucei gambiense* [74]. It is currently unclear if the large regions of homozygosity found in *G. duodenalis* genomes are due to loss of heterozygosity. Irrespective of the method, it seems likely that some form of recombination occurs in *G. duodenalis*. Higher-resolution markers or wide-spread sequencing will make it easier to understand how common the phenomenon is and what the effects may be on the molecular epidemiology of *Giardia* infections. This is further supported by a recent study that found evidence for recombination between assemblage A sub-assemblages using six markers [46]. However, it is likely that infrequent recombination would affect estimates of how related individuals are, complicating interpretation.

A need for new genotyping tools

Together, these issues indicate a need for more robust tools for genotyping *G. duodenalis* to understand better the molecular epidemiology of the disease and the biology of the parasite to improve outbreak management. The recent publication of updated reference genomes for *G. duodenalis* [75] and *G. muris* [17] provides more complete scaffolds to build upon and a well-characterized outgroup for comparison, contributing to these aims. However, the development of new tools depends on collecting a large and diverse selection of sequenced isolates to capture the diversity in the field more fully. Previously, sequencing of *G. duodenalis* samples has been restricted due to the limitations of sampling from faecal material and the requirement to adapt strains to axenic culture [14,19–21]. This adds significant cost, is labour intensive and introduces time delays to sequencing efforts. It also

ensures that only culturable strains can be sequenced, introducing potential bias, although assemblage-specific techniques may improve axenic culture techniques in the future. Comparative genomic analysis may provide information that would improve the axenic culture of specific assemblages, such as the recent analysis of assemblages C and D that identified assemblage-specific genes [18]. Several clinical isolates have recently been sequenced without axenic culture by concentrating cysts from clinical samples [76]. This approach may have limited effectiveness in many situations due to the requirement of a large number of starting cysts (often difficult to obtain in a diagnostic setting) and also results in highly variable sequencing quality and coverage. Several new technologies are now reaching maturity that may allow the rapid whole-genome sequencing (WGS) of *G. duodenalis* isolates from the small amount of starting material available in the clinical diagnostic setting. Accurate genomes that represent the individual assemblages would also assist in resequencing efforts in samples with low starting material. For many years this was limited to assemblages A, B and E, although genomes for assemblages C and D have recently been added as a resource for the community [18].

Whole genome sequencing of *Giardia*

Central to efforts to sequence *G. duodenalis* clinical isolates are affordable and relatively simple sequencing platforms that can be inserted into diagnostic pathways with little disruption. These technologies have led to the average cost of sequencing falling from $1,000 per megabase in 2009 to $0.01 in 2019. Costs are predicted to fall further with the drive to perform WGS routinely for certain pathogens to generate epidemiological data and to assist the management of outbreaks. Indeed, wider deployment of Illumina NovaSeq and third generation long-read sequencing have recently been used to update the *G. duodenalis*

reference genome [75]. The long-term aim would be user-friendly sequencing machines that could be deployed at the benchtop and used by non-specialist scientists in diagnostic laboratories. While we are still some way from this goal, and even upcoming “black box” technologies (such as Seeplex or Filmarray) only detect pathogens rather than genotype them, on-site rapid sequencing has shown promise in improving the management of bacterial and viral outbreaks by enhancing throughput, reproducibility and sensitivity. It has also led to a rapid expansion in the number of detectable genotypes and new strategies to understand the molecular epidemiology of these diseases [77]. This in turn has led to an improvement in identifying infectious agents and sources, tracking outbreaks and monitoring drug resistance markers in infected individuals who do not respond to treatment. We suggest similar efforts should be made to build a substantial collection of sequenced samples from multiple centres across the globe to capture the diversity of *G. duodenalis* in clinical, veterinary and environmental samples, leading to better management of clusters/outbreaks, reservoirs and drug resistance. To avoid the bottleneck of adapting strains to culture, several approaches have the potential to be developed to allow sequencing of isolates directly from faecal samples. For example, researchers in a recent study used a combination of cytometric sorting and single-cell whole-genome amplification to sequence assemblage C and D isolates from dogs, neither of which have been successfully cultured [18]. This revealed numerous genes that may be linked to host specificity and highlighted important differences in heterozygosity between the assemblages. Another promising technology is exome capture, an approach using biotinylated DNA or RNA ‘baits’ to capture DNA fragments from a target genome. This has successfully been used to identify and sequence material with a large amount of contaminating DNA, including enteric pathogens from faeces [78].

Making sense of the genomic data

If the issues of concentrating and purifying cysts to obtain sufficient quantity and quality of *G. duodenalis* DNA for sequencing are overcome, the next concern that limits the development of high-resolution genotyping markers becomes collating and analysing the large amounts of genomic data. A centralized global database will be required to develop a standardized set of markers efficiently, either by adapting current resources like GiardiaDB [79] or developing a dedicated system that receives MLST or other forms of sequence data. Several such databases have emerged that collate data from bacterial and viral sources, such as Enterobase [80], PubMLST and the European Nucleotide Archive that facilitate standards used for MLST genotyping. Similar efforts have been established in the past for *Giardia* species, for example ZOOPNET [55], however only now is the technology maturing sufficiently to meet the ambitions of the community for research and clinical applications. While a diagnostic panel of single nucleotide polymorphisms would provide the highest resolution for discriminating *G. duodenalis* genotypes, the most widely deployable output in the first instance will be additional MLST loci that expand on the current markers to increase reliability and resolution. These would ideally target genes or regions without indels that would cause frame shift mutations, making them more amenable to direct sequencing and avoiding cloning procedures. Direct sequencing is able to identify heterozygous positions across each of the four genomes present in a single *Giardia* isolate, providing extra discriminatory information [41]. Indeed, it may be preferable to target heterozygous regions to identify potential recombination events occurring between generations. Alternatively, if it is shown that LOH occurs in *G. duodenalis* to reduce deleterious alleles, identifying and contrasting such regions would also serve as a means to establish relationships between strains. In addition, a selection of genes that are under a range of selection pressures would

306 be ideal to provide different temporal resolution. This may entail using markers with
307 relatively high rates of mutation to track close relationships, while using slower evolving
308 genes to elucidate more ancestral relationships. The power of an expanded MLST panel with
309 higher discriminatory power has recently been demonstrated using a combination of six
310 markers, revealing evidence for recombination and zoonotic transfer in assemblage A
311 isolates [46]. However, increasing the number of markers further and improving reliability
312 across all of the different assemblages would allow the collective effort of the *Giardia*
313 community to quantify the degree of zoonotic transmission in different epidemiological
314 contexts and to identify environmental or animal reservoirs of infection.

315 ## The application of genomics to clinical isolates

316 The ability to link closely related *G. duodenalis* isolates within a short period of time would
317 allow potential outbreaks to be rapidly identified and effectively managed and could also be
318 used to identify drug-resistant strains. This approach is already in use for other pathogens
319 including tuberculosis, *Salmonella* spp. and *E. coli* 0157 [81]. Due to selective testing
320 protocols largely based on patient travel history, there is potential for under-detection of
321 the parasite in clinical samples and under-reporting of domestically-acquired cases, an issue
322 recently highlighted in Scotland [82]. Compared to other pathogens, limited resources are
323 directed towards *Giardia* surveillance activities and for this reason it may be hypothesized
324 that public health systems would lack the power to detect small-scale endemic outbreaks
325 should they occur. This is particularly the case if these outbreaks had low case numbers and
326 were not associated with a clear 'point source', such as a water contamination event. For
327 these reasons, having the capacity to detect outbreaks routinely as part of a clinical
328 genomics laboratory service would represent a major step forward for public health [12].

PCR and Sanger sequencing-based MLST approaches have the benefit of being more easily inserted into current laboratory pathways [83,84], are rapid, cost-effective and are also more likely to be adopted in lower-to-middle income countries that lack the capacity to perform large amounts of sequencing. As sequencing technology reaches greater penetration, clinical diagnostic services could begin to incorporate high-throughput sequencing into their pipelines while maintaining backwards compatibility with established MLST systems [84]. The cooperation of low, middle- and higher-income countries will not only be essential to identify both endemic outbreaks and reservoirs but also distinguish cases caused by 'foreign' genotypes of *Giardia* that have been imported through international travel. Working in such a broadly collaborative manner will undoubtedly raise issues in the sharing of public health data, the policies for which can vary widely between countries. Fortunately, efforts such as the Global Alliance for Genomics and Health are working to facilitate such programmes and their recommendations have been adopted by a number of health services worldwide [85].

Using genomics to understand the biology of *Giardia*

Although the collection of large amounts of sequencing data and the development of more robust sets of MLSTs will directly impact the management of giardiasis, these data will also contribute to answering several long-standing questions concerning the biology of the parasite that have implications for the disease. For example, large numbers of genomic or high-resolution MLST sequences would reveal the degree to which allelic recombination occurs between generations of parasites, especially if isolates were closely linked in terms of geographical location and time of sampling. The differences between generations would also demonstrate whether recombination was occurring between individuals or a

parasexual cycle was being utilized [72]. Determining the amount of recombination occurring in the field is important as it directly impacts our understanding of transmission networks and the likelihood of positive mutations becoming fixed in the population. These data would also confirm which assemblages are true species and no longer share genetic information, and those that have host-specific adaptations which limit their zoonotic potential. For example, several genes have been identified in assemblage C isolates that are suggested to be involved in host-specificity [18]. Finally, the capacity to genotype large numbers of isolates accurately has the potential to reveal associations between parasite genes and phenotypes. This will allow forward genetic techniques to be used in *G. duodenalis* for the first time, making it easier to link genotype to phenotype. Similarly, the ongoing refinement of single-cell genomics and transcriptomics also provides a tool to examine important biological questions in *Giardia* [86]. This would include identifying genes that distinguish between drug resistance and treatment failure [87] and identifying genotypes involved in more severe sequelae [11]. These approaches along with other advances in functional analysis will, to some extent, compensate for the lack of a reverse genetic framework for *Giardia* which has stifled research in this area [88]. Fortunately, in-roads are being made with the development of CRISPR/Cas9-mediated gene knockdown protocols, although the capacity for complete knockout remains elusive [89,90].

The current genetic contribution to drug resistance is unclear [91] and appears to be largely linked to transcriptional changes mediated by epigenetic factors [87]. However, there are numerous polymorphisms in several of the genes believed to be involved [92]. This suggests that if associating polymorphisms are identified in these candidate genes then there is the capacity to identify and distinguish cases of true drug-resistance from treatment failure for alternative reasons. The application of WGS to clinical isolates would also reveal whether *de*

novo positively selected mutations, including LOH events, arise *in vivo* and what role they may play in drug resistance. This knowledge would lead to an improvement in patient treatment by allowing alternative drug regimens to be followed immediately rather than waiting for treatment failure. If an effective alternative treatment can be utilized quickly in such cases, this would reduce the selective power of the ineffective treatment and limit the spread of resistant genotypes, therefore benefiting wider public health. To date, no allelic variants of genes have been identified that associate with different clinical outcomes of *G. duodenalis* infection, despite symptoms ranging from asymptomatic carriage to long-term IBS [11]. Preliminary data is largely ambiguous, with conflicting genotypes associating with the development of symptoms [2,93]. Again, the use of a publicly available pathogen database (or expanding current resources such as GiardiaDB) that integrates data from forward genetic screens and association studies would facilitate the identification of the genes involved. However, this would require a degree of clinical information being made available alongside the genetic information, complicating data sharing across jurisdictions. It will also be important to determine the molecular profile of isolates from asymptomatic cases, raising further ethical and logistical issues. The input that host and parasite genetics have in determining outcome is important to establish as these asymptomatic cases may represent a large and overlooked reservoir of infection for susceptible individuals, again impacting public health.

Concluding Remarks

In summary, there has been a long-held view that new genotyping markers are required for *G. duodenalis* to address numerous issues (Outstanding Questions Box). New sequencing technologies based on genome capture and single cell sequencing mean that it is now

possible to achieve these aims using clinical samples. However, successfully expanding the MLST framework for *G. duodenalis* will require cooperation across the research, medical and veterinary communities to develop a consistent set of standards and methods to avoid replicating effort and maximising return. It will also benefit from establishing a centralised database to collate and process data to deliver tangible outcomes that benefit public health. This does not necessarily require the generation of new tools, as current resources such as GiardiaDB may be expanded to perform a wider role. The routine application of WGS to clinical samples in the public health sphere would allow a genomics-led approach to outbreak detection, which contrasts to the 'response mode' approach currently taken where only large-scale outbreaks identified by other surveillance activities are genetically characterized. Clinical genomics would also allow drug-resistant isolates to be comprehensively genotyped, determining whether resistant lineages are circulating and whether *de novo*, positively selected mutants play a role in this poorly understood phenomenon. If successful, these approaches will greatly improve the global effort to reduce *Giardia* infections effectively and minimize outbreaks, and also answer long-standing questions concerning the biology of these unique eukaryotes.

Acknowledgements

The figures use art adapted from Servier Medical Art under a Creative Commons Attribution 3.0 Unported License https://creativecommons.org/licenses/by/3.0/).

References

1 Torgerson, P.R. *et al.* (2015) World Health Organization Estimates of the Global and Regional Disease Burden of 11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. *PLoS Med* 12, e1001920

2 Jerez Puebla, L.E. *et al.* (2017) Molecular analysis of *Giardia duodenalis* isolates from symptomatic and asymptomatic children from La Habana, Cuba. *Parasite Epidemiology and Control* 2, 105–113

3 Muadica, A.S. *et al.* (2020) Molecular Diversity of *Giardia duodenalis*, *Cryptosporidium* spp. and *Blastocystis* sp. in Asymptomatic School Children in Leganés, Madrid (Spain). *Microorganisms* 8, 466
4 Oliveira-Arbex, A.P. *et al.* (2016) Genotyping of *Giardia duodenalis* isolates in asymptomatic children attending daycare centre: evidence of high risk for anthroponotic transmission. *Epidemiol. Infect.* 144, 1418–1428
5 Ramírez, J.D. *et al.* (2015) Molecular diagnosis and genotype analysis of *Giardia duodenalis* in asymptomatic children from a rural area in central Colombia. *Infection, Genetics and Evolution* 32, 208–213
6 Rogawski, E.T. *et al.* (2017) Determinants and Impact of *Giardia* Infection in the First 2 Years of Life in the MAL-ED Birth Cohort. *Journal of the Pediatric Infectious Diseases Society* 6, 153–160
7 Fink, M.Y. and Singer, S.M. (2017) The Intersection of Immune Responses, Microbiota, and Pathogenesis in Giardiasis. *Trends in Parasitology* 33, 901–913
8 Allain, T. and Buret, A.G. (2020) Pathogenesis and post-infectious complications in giardiasis. In *Advances in Parasitology* 107pp. 173–199, Elsevier
9 Beatty, J.K. *et al.* (2017) *Giardia duodenalis* induces pathogenic dysbiosis of human intestinal microbiota biofilms. *International Journal for Parasitology* 47, 311–326
10 Barash, N.R. *et al.* (2017) *Giardia* Alters Commensal Microbial Diversity throughout the Murine Gut. *Infect. Immun.* 85, e00948-16, /iai/85/6/e00948-16.atom
11 Hanevik, K. *et al.* (2009) Development of functional gastrointestinal disorders after *Giardia lamblia* infection. *BMC Gastroenterol* 9, 27
12 Efstratiou, A. *et al.* (2017) Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011–2016. *Water Research* 114, 14–22
13 Adam, E.A. *et al.* (2016) Giardiasis outbreaks in the United States, 1971–2011. *Epidemiol. Infect.* 144, 2790–2801
14 Morrison, H.G. *et al.* (2007) Genomic Minimalism in the Early Diverging Intestinal Parasite *Giardia lamblia*. *Science* 317, 1921–1926
15 Karnkowska, A. *et al.* (2016) A Eukaryote without a Mitochondrial Organelle. *Current Biology* 26, 1274–1284
16 Bernander, R. *et al.* (2001) Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cell Microbiol* 3, 55–62
17 Xu, F. *et al.* (2020) The compact genome of *Giardia muris* reveals important steps in the evolution of intestinal protozoan parasites. *Microbial Genomics* DOI: 10.1099/mgen.0.000402
18 Kooyman, F.N.J. *et al.* (2019) Whole-genome sequencing of dog-specific assemblages C and D of *Giardia duodenalis* from single and pooled cysts indicates host-associated genes. *Microbial Genomics* 5,
19 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, e1000560
20 Ankarklev, J. *et al.* (2015) Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. *BMC Genomics* 16, 697
21 Jerlström-Hultqvist, J. *et al.* (2010) Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC Genomics* 11, 543
22 Mayrhofer, G. *et al.* (1995) Division of *Giardia* isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with *Giardia muris*. *Parasitology* 111, 11–17
23 Homan, W.L. *et al.* (1992) Comparison of *Giardia* isolates from different laboratories by isoenzyme analysis and recombinant DNA probes. *Parasitol Res* 78, 316–323
24 Keulen, H. *et al.* (1995) A Three Nucleotide Signature Sequence in Small Subunit rRNA Divides Human *Giardia* in Two Different Genotypes. *J Eukaryotic Microbiology* 42, 392–394

25 Soliman, R.H. *et al.* (2011) Identification of a novel Assemblage B subgenotype and a zoonotic Assemblage C in human isolates of *Giardia intestinalis* in Egypt. *Parasitology International* 60, 507–511
26 Gelanew, T. *et al.* (2007) Molecular characterization of human isolates of *Giardia duodenalis* from Ethiopia. *Acta Tropica* 102, 92–99
27 Zahedi, A. *et al.* (2017) Molecular typing of *Giardia duodenalis* in humans in Queensland – first report of Assemblage E. *Parasitology* 144, 1154–1161
28 Alexander, C.L. *et al.* (2017) An audit of *Cryptosporidium* and *Giardia* detection in Scottish National Health Service Diagnostic Microbiology Laboratories. *Epidemiol. Infect.* 145, 1584–1590
29 Soares, R. and Tasca, T. (2016) Giardiasis: an update review on sensitivity and specificity of methods for laboratorial diagnosis. *Journal of Microbiological Methods* 129, 98–102
30 Laude, A. *et al.* (2016) Is real-time PCR-based diagnosis similar in performance to routine parasitological examination for the identification of *Giardia intestinalis* , *Cryptosporidium parvum* / *Cryptosporidium hominis* and *Entamoeba histolytica* from stool samples? Evaluation of a new commercial multiplex PCR assay and literature review. *Clinical Microbiology and Infection* 22, 190.e1-190.e8
31 Cacciò, S.M. *et al.* (2002) Sequence analysis of the β-giardin gene and development of a polymerase chain reaction–restriction fragment length polymorphism assay to genotype *Giardia duodenalis* cysts from human faecal samples. *International Journal for Parasitology* 32, 1023–1030
32 Sulaiman, I.M. *et al.* (2003) Triosephosphate Isomerase Gene Characterization and Potential Zoonotic Transmission of *Giardia duodenalis*. *Emerg. Infect. Dis.* 9, 1444–1452
33 Cacciò, S.M. *et al.* (2008) Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B. *International Journal for Parasitology* 38, 1523–1531
34 Monis, P.T. *et al.* (1996) Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus. *Parasitology* 112, 1–12
35 Wielinga, C.M. and Thompson, R.C.A. (2007) Comparative evaluation of *Giardia duodenalis* sequence data. *Parasitology* 134, 1795–1821
36 Cacciò, S.M. *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
37 Traub, R.J. *et al.* (2004) Epidemiological and molecular evidence supports the zoonotic transmission of *Giardia* among humans and dogs living in the same community. *Parasitology* 128, 253–262
38 Thompson, R.C.A. and Ash, A. (2016) Molecular epidemiology of *Giardia* and *Cryptosporidium* infections. *Infection, Genetics and Evolution* 40, 315–323
39 Rochelle, P.A. *et al.* (1997) Comparison of primers and optimization of PCR conditions for detection of *Cryptosporidium parvum* and *Giardia lamblia* in water. *Appl. Environ. Microbiol.* 63, 106–114
40 Weiss, J.B. *et al.* (1992) Classification of subgroups of *Giardia lamblia* based upon ribosomal RNA gene sequence using the polymerase chain reaction. *Mol. Biochem. Parasitol.* 54, 73–86
41 Ankarklev, J. *et al.* (2012) Allelic sequence heterozygosity in single *Giardia* parasites. *BMC Microbiol* 12, 65
42 Lalle, M. *et al.* (2005) Genetic heterogeneity at the β-giardin locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes. *International Journal for Parasitology* 35, 207–213
43 Almeida, A. *et al.* (2010) Genotyping of *Giardia duodenalis* Cysts by New Real-Time PCR Assays for Detection of Mixed Infections in Human Samples. *AEM* 76, 1895–1901
44 Kohli, A. *et al.* (2008) *Giardia duodenalis* assemblage, clinical presentation and markers of intestinal inflammation in Brazilian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 102, 718–725

45 Geurden, T. *et al.* (2008) Mixed *Giardia duodenalis* assemblage A and E infections in calves. *International Journal for Parasitology* 38, 259–264
46 Ankarklev, J. *et al.* (2018) A novel high-resolution multilocus sequence typing of *Giardia intestinalis* Assemblage A isolates reveals zoonotic transmission, clonal outbreaks and recombination. *Infection, Genetics and Evolution* 60, 7–16
47 Monis, P.T. *et al.* (2009) Variation in Giardia: towards a taxonomic revision of the genus. *Trends in Parasitology* 25, 93–100
48 Robertson, L.J. *et al.* (2006) Application of Genotyping during an Extensive Outbreak of Waterborne Giardiasis in Bergen, Norway, during Autumn and Winter 2004. *Applied and Environmental Microbiology* 72, 2212–2217
49 Bonhomme, J. *et al.* (2011) Limitations of tpi and bg genes sub-genotyping for characterization of human Giardia duodenalis isolates. *Parasitology International* 60, 327–330
50 Brynildsrud, O. *et al.* (2018) *Giardia duodenalis* in primates: Classification and host specificity based on phylogenetic analysis of sequence data. *Zoonoses Public Health* 65, 637–647
51 Elwin, K. *et al.* (2014) *Giardia duodenalis* typing from stools: a comparison of three approaches to extracting DNA, and validation of a probe-based real-time PCR typing assay. *Journal of Medical Microbiology* 63, 38–44
52 Miller, K.M. and Sterling, C.R. (2007) Sensitivity of Nested PCR in the Detection of Low Numbers of *Giardia lamblia* Cysts. *Applied and Environmental Microbiology* 73, 5949–5950
53 Thompson, R.C.A. (2004) The zoonotic significance and molecular epidemiology of *Giardia* and giardiasis. *Veterinary Parasitology* 126, 15–35
54 Thompson, R.C.A. *et al.* (2008) The public health and clinical significance of *Giardia* and *Cryptosporidium* in domestic animals. *The Veterinary Journal* 177, 18–25
55 Sprong, H. *et al.* (2009) Identification of Zoonotic Genotypes of *Giardia duodenalis*. *PLoS Negl Trop Dis* 3, e558
56 Volotão, A.C. *et al.* (2007) Genotyping of *Giardia duodenalis* from human and animal samples from Brazil using β-giardin gene: A phylogenetic analysis. *Acta Tropica* 102, 10–19
57 Durigan, M. *et al.* (2014) Genetic Diversity of *Giardia duodenalis*: Multilocus Genotyping Reveals Zoonotic Potential between Clinical and Environmental Sources in a Metropolitan Region of Brazil. *PLoS ONE* 9, e115489
58 Caseca Volotão, A.C. *et al.* (2011) Giardiasis as zoonosis: between proof of principle and paradigm in the Northwestern region of São Paulo State, Brazil. *The Brazilian Journal of Infectious Diseases* 15, 382–383
59 Bartley, P.M. *et al.* (2019) Detection of potentially human infectious assemblages of *Giardia duodenalis* in fecal samples from beef and dairy cattle in Scotland. *Parasitology* 146, 1123–1130
60 Lebbad, M. *et al.* (2011) Multilocus Genotyping of Human *Giardia* Isolates Suggests Limited Zoonotic Transmission and Association between Assemblage B and Flatulence in Children. *PLoS Negl Trop Dis* 5, e1262
61 Lebbad, M. *et al.* (2010) From mouse to moose: Multilocus genotyping of *Giardia* isolates from various animal species. *Veterinary Parasitology* 168, 231–239
62 Wielinga, C. *et al.* (2011) Multi-locus analysis of *Giardia duodenalis* intra-Assemblage B substitution patterns in cloned culture isolates suggests sub-Assemblage B analyses will require multi-locus genotyping with conserved and variable genes. *International Journal for Parasitology* 41, 495–503
63 Serradell, M.C. *et al.* (2016) Vaccination of domestic animals with a novel oral vaccine prevents *Giardia* infections, alleviates signs of giardiasis and reduces transmission to humans. *npj Vaccines* 1, 16018
64 Minetti, C. *et al.* (2015) Case-Control Study of Risk Factors for Sporadic Giardiasis and Parasite Assemblages in North West England. *J. Clin. Microbiol.* 53, 3133–3140

65 de Lucio, A. *et al.* (2017) No molecular epidemiological evidence supporting household transmission of zoonotic *Giardia duodenalis* and *Cryptosporidium* spp. from pet dogs and cats in the province of Álava, Northern Spain. *Acta Tropica* 170, 48–56
66 Ramesh, M.A. *et al.* (2005) A Phylogenomic Inventory of Meiotic Genes. *Current Biology* 15, 185–191
67 Tibayrenc, M. *et al.* (1990) A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences* 87, 2414–2418
68 Cooper, M.A. *et al.* (2007) Population Genetics Provides Evidence for Recombination in *Giardia*. *Current Biology* 17, 1984–1988
69 Teodorovic, S. *et al.* (2007) Unusually Low Levels of Genetic Variation among *Giardia lamblia* Isolates. *Eukaryotic Cell* 6, 1421–1430
70 Lasek-Nesselquist, E. *et al.* (2009) Genetic Exchange Within and Between Assemblages of *Giardia duodenalis*. *Journal of Eukaryotic Microbiology* 56, 504–518
71 Clutterbuck, A.J. (1996) Parasexual recombination in fungi. *J. Genet.* 75, 281–286
72 Poxleitner, M.K. *et al.* (2008) Evidence for Karyogamy and Exchange of Genetic Material in the Binucleate Intestinal Parasite *Giardia intestinalis*. *Science* 319, 1530–1533
73 Gabriel, W. *et al.* (1993) Muller's Ratchet and Mutational Meltdowns. *Evolution* 47, 1744–1757
74 Weir, W. *et al.* (2016) Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *eLife* 5, e11473
75 Xu, F. *et al.* (2020) A chromosome-scale reference genome for *Giardia intestinalis* WB. *Sci Data* 7, 38
76 Hanevik, K. *et al.* (2015) Whole genome sequencing of clinical isolates of *Giardia lamblia*. *Clinical Microbiology and Infection* 21, 192.e1-192.e3
77 Allicock, O.M. *et al.* (2018) BacCapSeq: a Platform for Diagnosis and Characterization of Bacterial Infections. *mBio* 9, e02007-18, /mbio/9/5/mBio.02007-18.atom
78 Jones, M.R. and Good, J.M. (2016) Targeted capture in evolutionary and ecological genomics. *Mol Ecol* 25, 185–202
79 Aurrecoechea, C. *et al.* (2009) GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Research* 37, D526–D530
80 Zhou, Z. *et al.* (2020) The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* 30, 138–152
81 Lipworth, S. *et al.* (2019) SNP-IT Tool for Identifying Subspecies and Associated Lineages of *Mycobacterium tuberculosis* Complex. *Emerg. Infect. Dis.* 25, 482–488
82 Currie, S.L. *et al.* (2017) Under-reporting giardiasis: time to consider the public health implications. *Epidemiol. Infect.* 145, 3007–3011
83 Ibarz Pavón, A.B. and Maiden, M.C.J. (2009) Multilocus Sequence Typing. In *Molecular Epidemiology of Microorganisms* 551 (Caugant, D. A., ed), pp. 129–140, Humana Press
84 Rossen, J.W.A. *et al.* (2018) Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clinical Microbiology and Infection* 24, 355–360
85 Rahimzadeh, V. *et al.* (2016) An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health. *Biopreservation and Biobanking* 14, 256–259
86 Onsbring, H. *et al.* (2020) An efficient single-cell transcriptomics workflow for microbial eukaryotes benchmarked on *Giardia intestinalis* cells. *BMC Genomics* 21, 448
87 Ansell, B.R.E. *et al.* (2015) Drug resistance in *Giardia duodenalis*. *Biotechnology Advances* 33, 888–901

88 Jex, A.R. *et al.* (2020) Recent advances in functional research in *Giardia intestinalis*. In *Advances in Parasitology* 107pp. 97–137, Elsevier
89 Lin, Z.-Q. *et al.* (2019) Development of CRISPR/Cas9-mediated gene disruption systems in *Giardia lamblia*. *PLoS ONE* 14, e0213594
90 McInally, S.G. *et al.* (2019) Robust and stable transcriptional repression in *Giardia* using CRISPRi. *MBoC* 30, 119–130
91 Müller, J. *et al.* (2019) Resistance formation to nitro drugs in *Giardia lamblia*: No common markers identified by comparative proteomics. *International Journal for Parasitology: Drugs and Drug Resistance* 9, 112–119
92 Saghaug, C.S. *et al.* (2019) Genetic variation in metronidazole metabolism and oxidative stress pathways in clinical *Giardia lamblia* assemblage A and B isolates. *IDR* Volume 12, 1221–1235
93 Sahagún, J. *et al.* (2007) Correlation between the presence of symptoms and the *Giardia duodenalis* genotype. *Eur J Clin Microbiol Infect Dis* 27, 81–83

Glossary

18S ssu-rRNA gene: highly conserved gene encoding ribosomal RNA and commonly used for phylogenetic studies

allelic sequence heterozygosity (ASH): genetic differences at a genetic locus as assessed across the four different genomes in an individual *Giardia* isolate

cyst: infective, environmental stage of the parasite

diplomixis: a unique parasexual recombination cycle that occurs between two nuclei of a *Giardia* cell during encystation

diplomonad: group of flagellated protozoa with double cells and two nuclei

dysbiosis: disruption of the gut-microflora

excyzoite: a newly excysted *Giardia* cell with 4×4N ploidy

gene conversion: transfer of genetic material from an intact chromosomal DNA sequence to a homologous sequence which contains double-strand breaks

651 **inter-nuclei heterozygosity:** the degree of polymorphism between the two nuclei of a
652 *Giardia* cell, which is typically lower than would be expected

653 **isozyme (or isoenzyme):** multiple forms of the same enzyme that differ in amino acid
654 sequence and which can be used as the basis for a typing method

655 **lateral gene transfer:** the horizontal movement of genetic material between organisms
656 distinct from the vertical transmission of DNA from parent to offspring

657 **linkage disequilibrium:** the non-random association of alleles at two or more loci in a
658 population

659 **loss of heterozygosity (LOH):** regions that display no heterozygous sites in a genome

660 **multilocus sequence typing (MLST):** method used to characterize individuals genetically
661 based on the sequence at a number of marker loci distributed throughout the genome

662 **parasexual recombination:** a process genetic recombination utilized by some organisms
663 that does not require the production and fusing of haploid gametes

664 **peroxisome:** membrane-bound organelle found in eukaryotic cells involved in oxidation and
665 lipid metabolism

666 **ploidy:** the number of sets of chromosomes within the cell of an organism

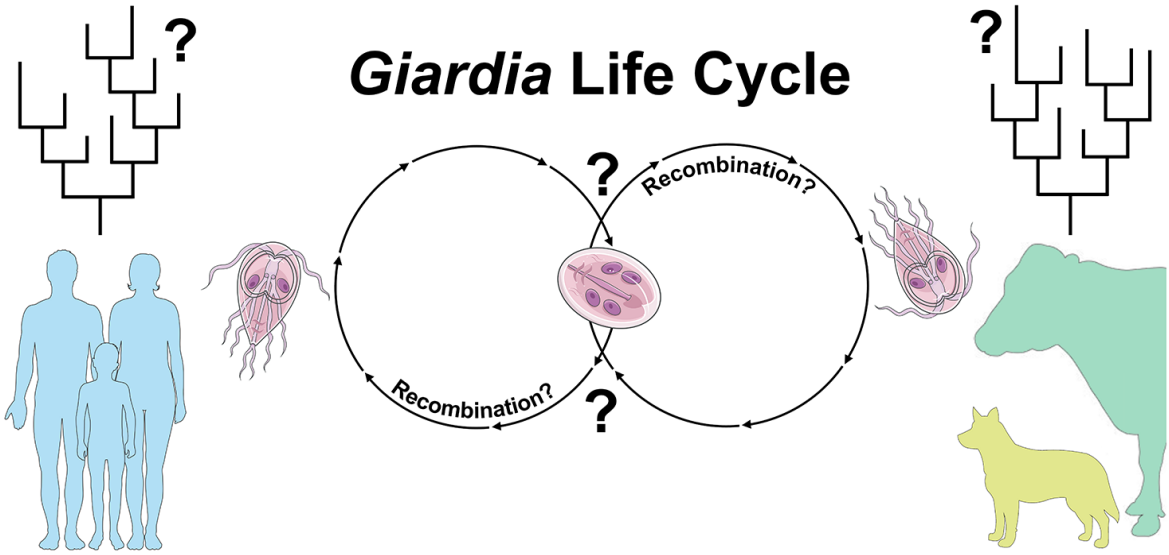
667 **trophozoite:** active 'feeding' form of the parasite located in the small intestine responsible
668 for pathology

669

Text Boxes

Box 1. G376 primer annealing site diversity. The diversity found within the three main genes used to genotype *G. duodenalis* (*bg*, *tpi*, and *gdh*) makes them useful to differentiate assemblages and subtypes. The separate assemblages also display different diversities, allowing a degree of sub-structuring to be observed. However, this diversity may also encompass the primer annealing sites, affecting the amplification success rates for isolates and assemblages. For example, there are 1,598 publicly available *G. duodenalis* sequences that include the annealing site for the commonly used β-giardin primer G376. Within these sequences, there can be up to eight polymorphisms compared to the primer sequence (Figure 2). Assemblage A sequences are the least likely to have polymorphisms, likely reflecting the fact that primers are initially designed using this assemblage. Conversely, assemblages B–H are more likely to contain polymorphisms in the annealing site (most have at least two), making them harder to amplify. In addition, as these public sequences are by definition the products of successful reactions with optimal conditions for promoting amplification, it is reasonable to speculate that many sequences fail to amplify due to polymorphisms in the primer-binding sites and other inhibiting factors.

Figures



Key Figure. The unknown aspects of *Giardia duodenalis* molecular epidemiology. The parasite *Giardia duodenalis* infects a wide range of mammalian hosts and features a relatively simple direct life cycle. Despite being one of the most common enteric parasites in the world, low resolution molecular markers limit research and control efforts in this important organism. For example, the relationships between clinical isolates cannot accurately be established, making it difficult to identify outbreaks and trace sources of disease. Similarly, a role for zoonotic transmission in human disease is also unclear, although it is suspected based on indirect evidence. Finally, certain aspects of *G. duodenalis* biology that would affect efforts to control the disease, such as the frequency of sexual or parasexual recombination, remain ambiguous. In this manuscript, we discuss how new sequencing technologies and strategies may contribute to a new generation of molecular markers for *Giardia* that will aid in addressing these questions and contribute to improving public health.

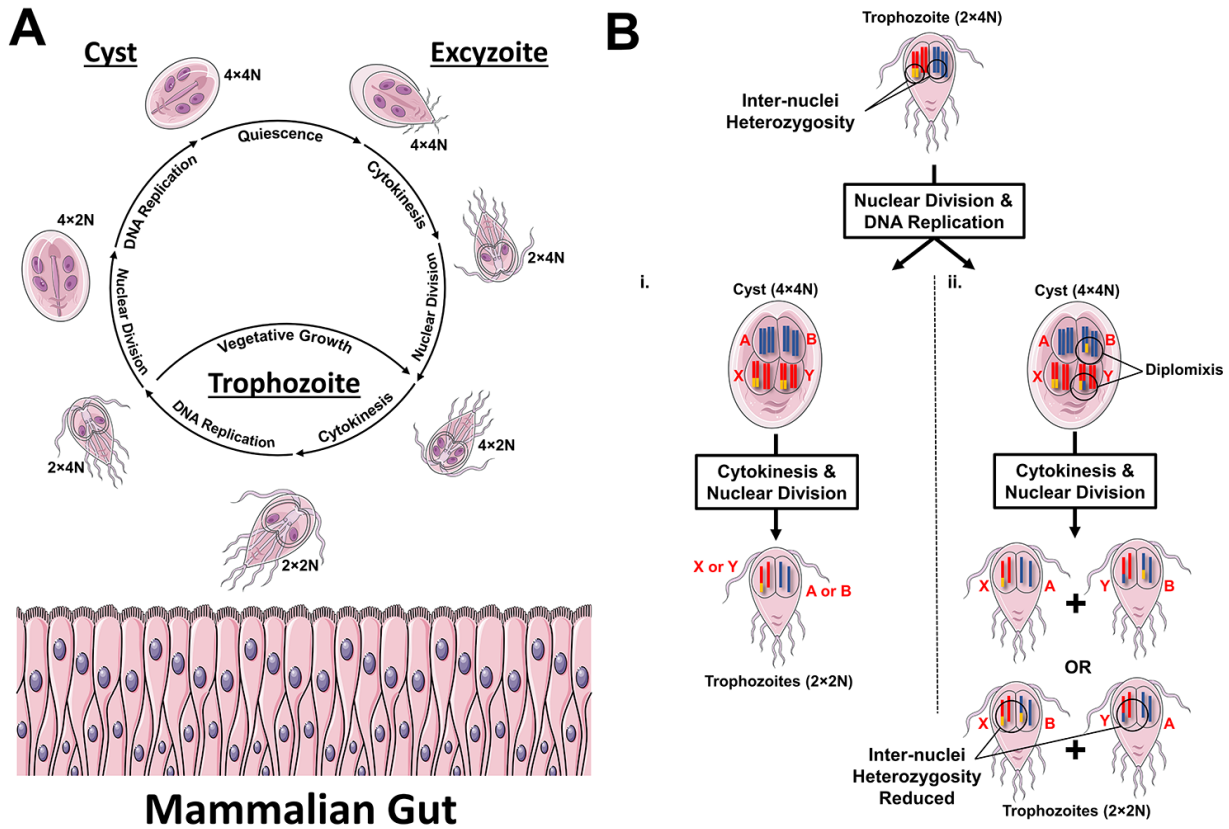
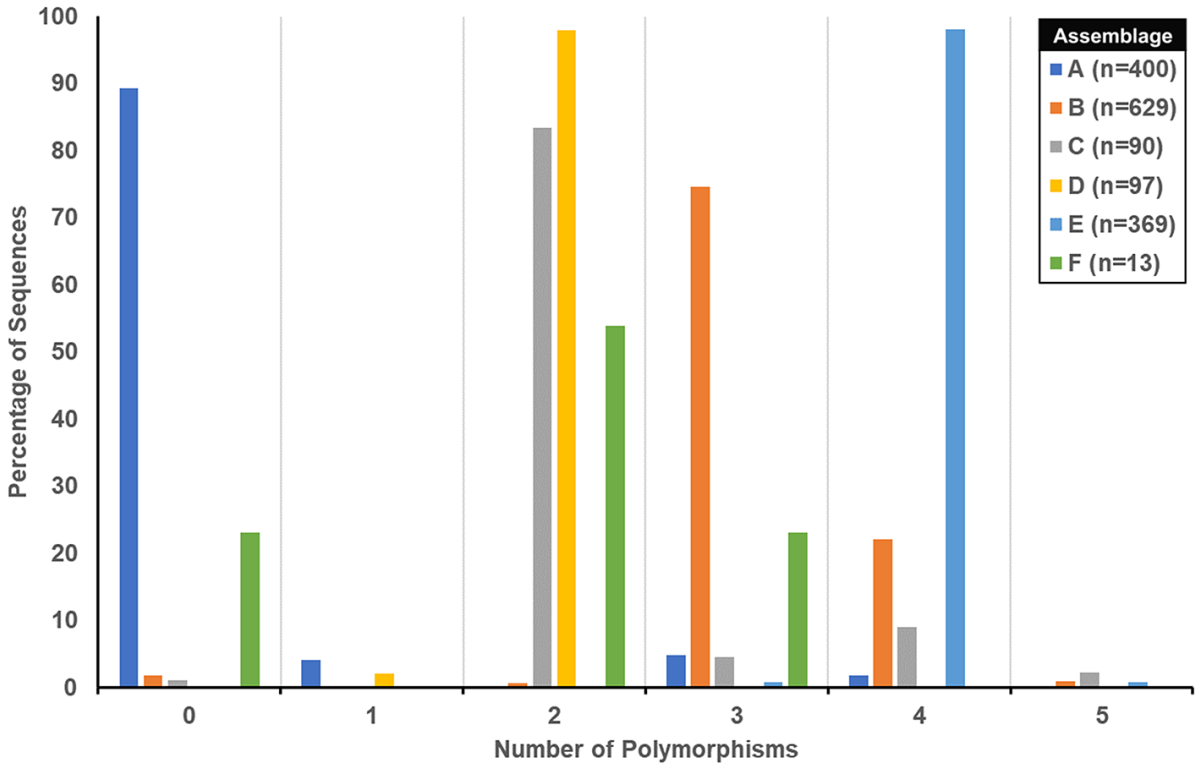


Figure 1. An overview of the *G. duodenalis* life cycle. A) In the mammalian intestinal tract, binucleate trophozoites cycle between 4N and 8N during vegetative growth. Trophozoites swept into the large intestine differentiate into cysts and are released into the environment for direct transmission. During encystation, the two nuclei divide and the DNA replicates, resulting in a ploidy of 16N. After activation in the mammalian stomach, cysts excyst in the intestine to release a 16N **excyzoite** with four nuclei. This **excyzoite** divides twice without DNA replication, resulting in four trophozoites that begin the vegetative cycle in a new host. B) *G. duodenalis* exhibits a unique parasexual cycle (**diplomixis**) that may contribute to lower than expected heterozygosity between the two nuclei of the cell. When the 16N cyst is formed during encystation, genetic exchange can occur between nuclei via homologous recombination. i) Without diplomixis, **inter-nucleus heterozygosity** is maintained in the daughter cells. Consequently, inter-nucleus heterozygosity will continue to increase and the genomes of the two nuclei will diverge. ii) With occasional diplomixis, regions of inter-nucleus heterozygosity can be transferred, reducing heterozygosity in some of the daughter cells and slowing the rate of divergence between the two nuclei. In addition, the process can generate genotypes with new allele combinations, further emulating sexual recombination.



721

722 **Figure 2. β-giardin G376 primer annealing site diversity among assemblages.** For this
723 analysis, 1,598 publicly available *G. duodenalis* sequences were downloaded and the G376
724 primer annealing site identified and aligned. The numbers of differences between each
725 sequenced site and the primer were calculated using the Levenstein distance. The data are
726 presented for each assemblage, showing that assemblage A sequences have few
727 polymorphisms in the annealing site compared to the published primer. In contrast,
728 assemblages B–H have, for the most part, at least two and up to five polymorphisms in total.

729

730

Tables

Table 1. *Giardia duodenalis* host assemblages

***Giardia* Assemblage/Sub-Assemblage**	**Host**	**Proposed Nomenclature** [47]
A I II III	Humans, non-human primates, canines, felines, other mammals	*G. duodenalis*
B	Humans, non-human primates, canines, felines, other mammals	*G. enterica*
C	Canines	*G. canis*
D	Canines	*G. canis*
E	Livestock	*G. bovis*
F	Felines	*G. cati*
G	Rodents	*G. simondi*
H	Marine mammals	