Food and Waterborne Parasitology 21 (2020) e00098

Contents lists available at ScienceDirect

Food and Waterborne Parasitology

journal homepage: www.elsevier.com/locate/fawpar

Assessment of next generation amplicon sequencing of the beta-giardin gene for the detection of *Giardia duodenalis* assemblages and mixed infections

Jenny G. Maloney, Aleksey Molokin, Monica Santin *

Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, United States Department of Agriculture, 10300 Baltimore Ave, Beltsville, MD 20705, United States

ARTICLE INFO

Article history:
Received 10 July 2020
Received in revised form 9 October 2020
Accepted 12 October 2020

Keywords:
Giardia
Next generation sequencing
Mixed infections
Beta-giardin
Assemblages

ABSTRACT

Giardia duodenalis is an enteric protozoan parasite commonly found in humans and many other animals around the world. The parasite is grouped into genetically related strains called assemblages which display differing degrees of host specificity. Although mixed assemblage infections have been documented the full extent of the occurrence and importance of mixed infections remains to be characterized as current sequencing technologies lack the sensitivity to readily detect mixed infections. Here we have developed a next generation amplicon sequencing (NGS) protocol and analysis pipeline for detecting *Giardia* assemblages using the beta-giardin gene. NGS was validated using 37 isolates that included *Giardia muris* and six assemblages (A-F) of *Giardia duodenalis* obtained from seven different hosts. NGS was compared to traditional PCR and direct Sanger sequencing for its ability to detect *Giardia* species, assemblages, and mixed assemblage infections. We demonstrate that NGS works as well as PCR and Sanger sequencing for assemblage detection as the same assemblage was observed in all samples by both methods. NGS has the further benefit of detecting mixed assemblage infections, low abundance assemblages, and intra-assemblage variation in samples which would have been missed using direct Sanger sequencing alone. NGS represents a powerful new tool for exploring *Giardia* infections not only in infected hosts but also in environmental specimens which may aide in understanding *Giardia* epidemiology.

1. Introduction

The protozoan *Giardia* is one of the most common enteric parasites in humans in the world. Infection can cause a variety of intestinal symptoms including diarrhea, bloating, and malabsorption or may be asymptomatic (Feng and Xiao, 2011). It is estimated to affect nearly 300 million people a year worldwide with 1.2 million cases a year occurring in the U.S. (Scallan et al., 2011). *Giardia* is also a frequent parasite in animals including food animals, pets, and wildlife (Feng and Xiao, 2011). Both humans and animals can shed large numbers of cysts during infection with fecal cyst shedding occurring for up to months at a time (Hoar et al., 2009; Pickering et al., 1984; Rendtorff, 1954; Santín et al., 2009). The cyst form is environmentally hardy and can persist for long periods in cool, moist environments facilitating transmission. The transmission of *Giardia* is via the fecal oral route, and

* Corresponding author at: Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD, United States.
E-mail address: monica.santin-duran@usda.gov. (M. Santin).

https://doi.org/10.1016/j.fawpar.2020.e00098

human infections occur through a number of routes including direct contact with infected humans and animals and indirectly by ingestion of cysts in contaminated food or water (Cacciò and Ryan, 2008).

Giardia is one of the most commonly reported etiological agents in waterborne outbreaks worldwide (Efstratiou et al., 2017). A recent summary of all published waterborne outbreaks associated with protozoan parasites reported 905 outbreaks, and *Giardia* was implicated in a third of those (293; 32.4%) (Efstratiou et al., 2017). *Giardia* cysts have been reported in diverse types of fresh produce as well as in unpasteurized milk and shellfish (Dixon, 2015; Ryan et al., 2019). However, few *Giardia* foodborne outbreaks have been reported and the total number of outbreaks is likely underestimated (Ryan et al., 2019).

Currently, there are eight valid species of *Giardia*, namely: *G. agilis*, *G. ardeae*, *G. cricetidarum*, *G. duodenalis* (synonyms *G. intestinalis* and *G. lamblia*), *G. microti*, *G. muris*, *G. peramelis*, and *G. psittaci* (Ryan et al., 2019). Among those species, *G. duodenalis* is the only one that causes human infections. It is a species complex composed of eight assemblages (A-H) that are morphologically indistinguishable but that show consistent genetic uniqueness. The most commonly used genes to determine assemblages are small subunit (*SSU*) rRNA, beta-giardin (*bg*), triosephosphate isomerase (*tpi*), or glutamate dehydrogenase (*gdh*) (Feng and Xiao, 2011). These assemblages display differing degrees of host specificity, Assemblages A and B show an unrestricted host range and are found in humans as well as in a wide range of animals, while the remaining assemblages (C-H) have stricter host-specificity with Assemblages C and D found in canids, Assemblage E in hoofed animals, Assemblage F in felines, Assemblage G in rodents, and Assemblage H in sea mammals. Assemblages A and B are frequently reported infecting humans and animals, and they can be transmitted zoonotically. Zoonotic transmission is thought to mostly occur via hooved livestock and wild ruminants and less commonly involve companion animals that are mainly infected with assemblages C, D, and F (Feng and Xiao, 2011). Although Assemblages A and B are the most common assemblages reported in humans, Assemblages C, D, E, and F have also been documented sporadically in humans further confusing the degree of zoonotic transmission potential of *Giardia* (Cacciò et al., 2018).

Currently there is no standardized method for the molecular detection of *Giardia* assemblages from fecal and environmental samples. Molecular characterization in epidemiology studies of *G. duodenalis* isolates relies mostly on sequencing fragments of *SSU* rRNA, *gdh*, *tpi*, and *bg* genes (Cacciò et al., 2018). Real-time assemblage-specific PCR assays have also been developed for the detection of assemblages A, B, and, E that are valuable in the detection of infections with mixed assemblages (Almeida et al., 2010; Elwin et al., 2014; Van Lith et al., 2015). Several studies using one or more of these genes to detect *Giardia* assemblages have demonstrated that mixed assemblage infections are common in human, animals, and water (Adell-Aledón et al., 2018; Geurden et al., 2008; Gómez-Muñoz et al., 2012; Iijima et al., 2018; Lalle et al., 2005; Ramo et al., 2017; Sprong et al., 2009; Waldram et al., 2017). In fact, mixed infections were found to range from 2% to 21% in molecular surveys of *Giardia* in humans from Australia, the United Kingdom, India, Italy, and Ethiopia (Cacciò and Ryan, 2008). Furthermore, it is recognized that the occurrence of mixed infections may be underestimated using current detection methods as a study using assemblage specific primers demonstrated that a large number of mixed infections are not detected using conventional PCR and sequencing (Geurden et al., 2008). Given the potential importance of mixed assemblage infections in understanding the epidemiology of *Giardia*, an improved method for the determination of mixed infections is needed. Understanding mixed infections and intra-assemblage variation within a sample is key from an epidemiology and public health point of view to understand potential sources of infections and transmission and from a clinical point of view to understand the wide range of clinical presentations and response to treatments during infection.

Next generation amplicon sequencing (NGS) allows for the generation of an immense number of individual sequences from a single sample offering one potential tool for overcoming the limitations of conventional PCR and sequencing for detecting mixed assemblage infections and intra-assemblage variations. The giardin genes are unique to *Giardia* making assays targeting those genes highly discriminatory. The *bg* gene is commonly used in molecular characterization of *Giardia* isolates as it can differentiate all eight assemblages of *G. duodenalis* and has the ability to capture intra-assemblage variability (Sprong et al., 2009). A high sensitivity for amplification of *bg* has been reported in multilocus studies using *tpi*, *bg*, and *gdh* (Covacin et al., 2011; Soliman et al., 2011; Broglia et al., 2013). In addition, the substitution rate for *bg* was reported to be 0.03 substitutions per nucleotide (Feng and Xiao, 2011) which offers appropriate variation for differentiation of *Giardia* species, *G. duodenalis* assemblages, and has the potential for detecting within assemblage variations. In this study, we have developed an NGS protocol and analysis pipeline for *bg* that was then applied to DNAs obtained from 37 *Giardia* positive samples from multiple hosts previously identified as *G. duodenalis* assemblages A-F and *G. muris*. We further compare NGS and direct Sanger sequencing results to demonstrate the suitability of NGS for species and assemblage detection and the ability of NGS to detect mixed assemblage infections and intra-assemblage variations.

2. Materials and methods

2.1. Specimens

A panel of 37 DNAs from fecal samples previously identified as *G. duodenalis* and *G. muris* positive was used in this study (Table 1). Species and assemblage of each of the DNAs included in the panel were previously determined using PCR and direct sequencing of the *SSU* rRNA gene with primers previously described (Hopkins et al., 1997). A total of 1 *G. muris* and 6, 5, 2, 6, 13, and 4, specimens assigned to *G. duodenalis* assemblages A, B, C, D, E, F, respectively were used.

2.2. PCR and Sanger sequencing

A ~500 bp fragment of the *bg* gene was amplified using a nested PCR protocol with primers previously described (Lalle et al., 2005). The PCR mix consisted of buffer containing 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.5 μM for each forward and reverse

Table 1
Information of *Giardia* specimens used in this study (host/geographic origin) and identification of *Giardia* spp. and *G. duodenalis* assemblages using the beta-giardin gene by Sanger and next generation amplicon sequencing (NGS) (at 98% and 100% clustering).

Specimen ID	Host	Location	*Giardia* species or *G. duodenalis* assemblage by Sanger	*Giardia* species or *G. duodenalis* assemblage by NGS (% of sample) (GenBank Accession No.)	
				Clustering 98%	Clustering 100%
1	Cattle	USA	A	A (100%) (MT713315)	A (100%) (MT713315)
2	Cattle	USA	A	A (100%) (MT713315)	A (100%) (MT713315)
3	Cattle	USA	A	A (62.0%) (MT713315)/E (38.0%) (MT713318)	A (65.6%) (MT713315)/E (34.4%) (MT713318)
4	Cattle	USA	A/E	A (64.9%) (MT713315)/E (35.1%) (MT713313)	A (65.1%) (MT713315)/E[a] (20.3%/14.7%) (MT713313/MT713317)
5	Goat	USA	A/E	A (72.5%) (MT713315)/E (27.5%) (MT713306)	A (74.3%) (MT713315)/E (25.8%) (MT713306)
6	Goat	USA	A	A (85.3%) (MT713315)/E (14.7%) (MT713306)	A[a] (78.5%/8.4%) (MT713315/MT713339)/E (13.1%) (MT713306)
7	Chinchilla	USA	B	B (100%) (MT713309)	B[a] (51.5%/48.5%) (MT713309/MT713324)
8	Human	USA	B	B (100%) (MT713304)	B[a] (61.3%/38.7%) (MT713304/MT713323)
9	Human	Colombia	B	B (100%) (MT713321)	B[a] (45.4%/27.3%/23.7%/3.7%) (MT713321/MT713326/MT713323/MT713336)
10	Human	Colombia	B	B (100%) (MT713321)	B (100%) (MT713321)
11	Human[b]	USA	B	B (100%) (MT713310)	B (100%) (MT713310)
12	Dog	USA	C	C (100%) (MT713302)	C[a] (88.4%/11.6%) (MT713302/MT713331)
13	Dog	USA	C	C (100%) (MT713305)	C[a] (71.7%/28.3%) (MT713305/MT713302)
14	Dog	USA	D	D (100%) (MT713308)	D[a] (43.4%/30.4%/12.7%/8.0%/5.4%) (MT713308/MT713303/ MT713319/MT713334/MT713335)
15	Dog	Colombia	D	D (100%) (MT713320)	D[a] (72.5%/27.6%) (MT713320/MT713303)
16	Dog	USA	D	D (100%) (MT713319)	D (100%) (MT713319)
17	Dog	USA	D	D (100%) (MT713319)	D (100%) (MT713319)
18	Dog	USA	D	D (83.3%) (MT713303)/C (16.7%) (MT713305)	D[a] (54.4%/29.1%) (MT713303/MT713320)/C[a] (11.2%/2.8%/2.1%) (MT713305/MT713340/MT713302)
19	Dog	USA	D	D (99.7%) (MT713320)/C (0.2%) (MT713302) E (0.1%) (MT713311)	D[a] (68.1%/19.9%/11.9%) (MT713320/MT713327/MT713308)/C (0.1%) (MT713302)
20	Cattle	USA	E	E (100%) (MT713311)	E (100%) (MT713311)
21	Cattle	USA	E	E (100%) (MT713311)	E (100%) (MT713311)
22	Cattle	USA	E	E (100%) (MT713307)	E[a] (48.7%/25.1%/14.1%/11.9%) (MT713307/MT713317/MT713328/MT713313)
23	Cattle	USA	E	E (100%) (MT713317)	E[a] (44.3%/22.4%/13.5%/9.3%/7.7%/2.7%) (MT713317/MT713311/MT713329/MT713332/ MT713318/MT713337)
24	Cattle	USA	E	E (100%) (MT713311)	E[a] (57.5%/34.8%/7.7%) (MT713311/MT713322/MT713325)
25	Cattle	USA	E	E (100%) (MT713318)	E (100%) (MT713318)
26	Cattle	USA	E	E (100%) (MT713311)	E[a] (70.5%/19.3%/10.2%) (MT713311/MT713322/MT713325)
27	Cattle	USA	E	E (100%) (MT713317)	E[a] (77.0%/23.0%) (MT713317/MT713311)
28	Cattle	USA	E	E (100%) (MT713311)	E[a] (60.3%/32.8%/6.9%) (MT713311/MT713322/MT713325)
29	Cattle	USA	E	E (100%) (MT713322)	E[a] (47.9%/38.2%/13.9%) (MT713311/MT713322/MT713330)
30	Cattle	USA	E	E (100%) (MT713317)	E[a] (56.8%/33.8%/9.4%) (MT713317/MT713311/MT713333)
31	Goat	USA	E	E (100%) (MT713306)	E (100%) (MT713306)
32	Goat	USA	E	E (100%) (MT713318)	E (100%) (MT713318)
33	Cat	USA	F	F (100%) (MT713316)	F (100%) (MT713316)
34	Cat	USA	F	F (100%) (MT713316)	F (100%) (MT713316)
35	Cat	USA	F	F (100%) (MT713316)	F (100%) (MT713316)
36	Cat	USA	F	F (100%) (MT713312)	F (100%) (MT713312)
37	Mouse[b]	USA	*G. muris*	*G. muris* (100%) (MT713314)	*G. muris*[a] (86.8%/13.2%) (MT713314/MT713338)

[a] Intra species or assemblage variation.
[b] Isolates (cysts) acquired from Waterborne Inc. (New Orleans, LA, USA); H3 isolate propagated in gerbils for sample #11 and Roberts-Thomson isolate propagated in mice for *G. muris* for sample #37.

primers, 1.25 μl BSA (0.1 g/10 ml), 2.5 U *Taq*, and 2.5 μl template in a 25 μl total reaction volume. The PCR reaction was performed as follows: denaturation at 94 °C for 5 min followed by 40 cycles of 94 °C for 30 s, 65 °C for 30 s (primary PCR) or 55 °C for 30 s (secondary PCR), 72 °C for 1 min, and a final elongation step of 72 °C for 7 min. Following amplification, PCR products were analyzed using a QIAxcel (Qiagen, Valencia, CA). Amplicons were purified using Exonuclease I/Shrimp Alkaline Phosphatase (ExoSAP-IT™ Express, Applied Biosystems, Foster City, CA), and sequenced in both directions using the same primers as the secondary PCR in 10 μl reactions, Big Dye™ chemistries, and an ABI 3130xl sequence analyzer (Applied Biosystems, Foster City, CA). Sequence chromatograms of each strand were aligned and examined with Lasergene software (DNASTAR, Inc., Madison, WI).

2.3. Next generation amplicon sequencing library preparation

PCR amplification of the same *bg* gene fragment used for Sanger sequencing was performed with the exception that KAPA HiFi HotStart ReadyMix (KAPABioSystems, Cape Town, South Africa) was used in both the primary and secondary PCR, and the primers for the secondary PCR were modified to contain the Illumina overhang adapter sequences on the 5′ end; forward primer ILMN_GiardinF 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAACGAGATCGAGGTCCG-3′ and reverse primer ILMN_GiardinR 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCGACGAGCTTCGTGTT-3′(adapter sequence underlined). The Illumina 16S Metagenomic Sequencing Library Preparation protocol (Part# 15044223 Rev. B) was used for the remaining steps of library preparation. Final libraries were quantified by Quant-iT dsDNA Broad-Range Assay Kit (Thermo Fisher, Waltham, MA) on a SpectraMax iD5 (Molecular Devices, San Jose, CA) prior to normalization. A final pooled library concentration of 8 pM with 10% PhiX control was sequenced on an Illumina MiSeq using 600 cycle v3 chemistry (300 base pair, paired-end reads) (Illumina, San Diego, CA) following manufacturer's recommendations.

2.4. Bioinformatic analysis

Demultiplexing of raw bcl files was performed using Illumina MiSeq Reporter software. Paired end reads were then processed and analyzed using an in-house pipeline which included the BBTools package v38.22 (Bushnell, 2014), VSEARCH v2.8.0 (Rognes et al., 2016), and BLAST+ 2.7.1 similar to the pipeline previously reported (Maloney et al., 2019). Briefly, read pairs were merged, filtered for quality and length, denoised, and checked for chimeric sequences. Clustering and the assignment of centroid sequences to operational taxonomic units (OTU) was performed within each sample at both a 98% and 100% identity threshold. Low abundance OTUs composed of less than 100 sequences were filtered out and another round of chimera filtering was performed within each sample. OTUs were then blasted against a reference database consisting of all *Giardia* sequences available from NCBI with the BLAST+ blastn command. BLAST results were filtered to remove hits below an alignment length of 400. All OTUs were assigned a *Giardia* species and *G. duodenalis* assemblage based on the best match from the BLAST results. Negative controls were included in sequencing runs and after bioinformatic analysis no *Giardia* OTUs were observed in the controls. All raw fastq files were deposited to the NCBI sequence read archive under the accession number SRR12137346-SRR12137382. The nucleotide sequences for OTUs obtained in this study have been deposited in GenBank under the accession numbers MT713302-MT713338.

3. Results

3.1. Sanger sequencing

After Sanger sequencing the *bg* gene of the 37 isolates included in this study two samples (#4 and #5) had the appearance of a mixed infection with multiple double peaks leading to ambiguous base calls present in the sequence chromatogram (data not shown). The other 35 samples appeared to contain a single unambiguous sequence.

3.2. Next generation amplicon sequencing

A total of 8,935,898 paired end reads was generated from the 37 *Giardia* positive samples included in this study. Following end trimming and pair merging 3,288,494 reads were retained. After quality filtering, denoising, and the removal of chimeric sequences, 2,580,156 merged reads remained for use in operational taxonomic unit (OTU) generation. Two clustering identity thresholds, 98% and 100%, were used in this study to determine species, assemblage level, and intra-assemblage level sequence differences within individual samples. Clustering at 98% yielded a single OTU/*Giardia* species/assemblage in 31 of the samples. Multiple OTUs/*Giardia* assemblages/assemblage variants were found in 6 of the samples with up to 3 OTUs in a single sample (Table 1). Clustering at 100% yielded a single OTU/*Giardia* species/assemblage in 15 samples and multiple OTUs/assemblages/assemblage variants in 22 samples with up to 6 OTUs in a single sample (Table 1).

3.3. Comparison of Sanger sequencing and next generation amplicon sequencing

Six assemblages of *G. duodenalis* (A-F) and *G. muris* were successfully sequenced and identified using both Sanger and NGS methods. In all samples the assemblage found by Sanger sequencing matched the most abundant assemblage observed in the NGS results (Table 1). Additionally, in six samples NGS could detect the presence of mixed assemblage infections, while mixed infections were only suspected in two samples based on the direct Sanger sequencing chromatograms.

A mixture of assemblages A and E was observed in two cattle (#3 and #4) and two goat samples (#5 and #6) at both clustering levels (Table 1). Samples #4 and #5 were the only samples suspected to contain multiple assemblages by direct Sanger sequencing with multiple peaks and ambiguous base calls present in the sequencing chromatogram (data not shown). The best BLAST match for samples #3 and #6 from the Sanger sequence was assemblage A, while NGS detected both assemblages but with a predominance of assemblage A (Table 1). The other two mixed infections were observed in dogs. For both samples (#18 and #19), the Sanger sequence contained assemblage D and the chromatogram did not indicate the presence of a mixed infection, while mixtures of C and D (#18) and C, D and E (#19) were observed in these samples using NGS. The assemblage E observed in sample #19 was present at a low relative abundance and only detected using the 98% clustering threshold.

There were 21 (98% clustering) and 39 (100% clustering) unique *Giardia* nucleotide sequences observed among the 37 samples analyzed in this study. Multiple unique sequences were observed for all assemblages except for assemblage A and *G. muris* at the 98% clustering threshold, while multiple unique sequences of all assemblages and *G. muris* were observed at the 100% clustering threshold (Table 2). With the exception of assemblage F, within-sample sequence variation was observed for all assemblages and *G. muris* at the 100% clustering threshold with between two and six sequence variants of the same assemblage observed in individual samples (Table 1). Assemblage E had the most unique sequences at both clustering thresholds (Table 2). This variability occurred both within and between samples with up to six variants of assemblage E identified in a single sample using 100% clustering (#23, Table 1).

4. Discussion

Giardia duodenalis is one of the most common intestinal parasites of humans, yet the molecular epidemiology of this parasite is hampered by the inability of conventional sequencing techniques to detect mixed assemblage infections. Conventional PCR followed by direct Sanger sequencing tends to only detect the most abundant parasite population which is preferentially amplified in the PCR. In the present study, NGS was applied to 37 *Giardia* positive samples representing six assemblages (A-F) of *G. duodenalis* and *G. muris*. NGS was compared to conventional PCR and direct Sanger sequencing to determine the suitability of NGS for both the detection of *G. duodenalis* assemblages and the detection of mixed-assemblage infections.

We found a high level of agreement between Sanger sequencing and NGS for assemblage identification. All 37 samples were found to contain the same assemblage using both technologies (Table 1). And by NGS, the most abundant assemblage identified in samples with mixed assemblages was the assemblage detected by Sanger. This agreement supports the use of NGS for assemblage identification as no disparities between techniques was observed that might indicate a bias for an assemblage or the introduction of erroneous sequence data due to the use of NGS. In fact, NGS successfully produced sequences with full length (511 bp) alignments to BLAST matches ranging from 99.6 to 100% identity to *G. duodenalis* nucleotide sequences available in the GenBank nucleotide database further supporting the ability of this method to produce high quality sequences for assemblage level identification for use in molecular epidemiology studies.

The advent of next generation sequencing technologies provides a potential new tool for exploring and understanding the occurrence and importance of mixed parasite populations in both individual and water/food samples. This topic has been explored for several other enteric protist parasites including *Blastocystis*, *Cryptosporidium*, and *Eimeria* (Grinberg et al., 2013; Maloney et al., 2019; Paparini et al., 2015; Vermeulen et al., 2016). These studies report varying degrees of within host parasite diversity, however one recurring theme is the superior ability of NGS to detect this type of diversity compared to conventional PCR and Sanger sequencing. Indeed, the same trend was observed in this study. Mixed assemblage infections were observed in six samples, two which were suspected from the Sanger sequence chromatogram and four which would have been missed entirely using only Sanger sequencing. Future studies which aim to characterize the degree of *Giardia* species and assemblage mixing could clearly benefit from the use of this technology.

The mixed assemblage infections observed in this study occurred in cattle, goats, and dogs. Cattle and goats were found to be infected with both assemblage A and assemblage E (Table 1). This finding is in line with molecular epidemiological surveys in cattle worldwide that have reported E as the predominant assemblage followed by lower frequencies of assemblage A (Ryan and Zahedi, 2019). Mixtures of A and E have been previously reported in cattle and goats (e.g. Geurden et al., 2008, 2012; Minetti et al., 2014; Santin et al., 2012; Tzanidakis et al., 2014; Yin et al., 2018). Likewise the combination of assemblages C and D observed in one of the dog samples was not surprising as these two assemblages are commonly reported in canines (Cacciò et al., 2018). Mixed assemblage C and D infections have been reported in dogs from Italy, Poland, and the United States (Paoletti et al., 2015; Piekarska et al., 2016; Scorza and Lappin, 2017). However, these mixed infections were only detected using multiple genetic markers with amplification preference for the different assemblages or presumed from the appearance of the Sanger chromatograms. This issue further highlights an added advantage of NGS which can unambiguously detect mixed infections using a single genetic marker. One mixed infection in a dog contained a combination of assemblage C, D, and E. In this infection assemblage D represented the majority of sequences present, over 99% (Table 1). The ability of NGS to detect a mixture of assemblages in this sample demonstrates the improved sensitivity of this method as two other assemblages present in very low

Table 2
Number of samples identified containing each *G. duodenalis* assemblage or *G. muris* with number of unique nucleotide sequence of the beta-giardin gene identified using next generation amplicon sequencing at both 98% and 100% clustering thresholds.

Assemblage/species	No. of samples positive	No. of unique nucleotide sequences at 98% clustering	No. of unique nucleotide sequences at 100% clustering
A	6	1	2
B	5	4	8
C	4[a]	2	4
D	6	4	7
E	16[b]	7	14
F	4	2	2
G. muris	1	1	2

[a] Two from mixed infection.
[b] Five from mixed infection.

read abundance were also identified in this sample. In dogs, assemblages C and D are most frequently reported with sporadic reports of A and B, indicating that dogs may be host to a variety of assemblages (Adell-Aledón et al., 2018; Cacciò et al., 2018). The identification of assemblage E in dogs is unusual, but it has been previously reported in five dogs in China and a dog in Canada (Julien et al., 2019; Li et al., 2015). Furthermore, given the low relative abundance of assemblage E sequences in this sample it is also possible that the presence of assemblage E in this sample does not represent an active infection but may have been passing through this host. It should be noted that assemblage E was only observed in this sample at the 98% clustering threshold. Because this assemblage was present in such a low abundance in this sample, clustering at 100% identity split the OTU into multiple variants of assemblage E all of which fell below the minimum established threshold of 100 sequences and were removed by the analysis pipeline.

Being able to distinguish between *Giardia* species and assemblages of *G. duodenalis* has important implications in understanding the epidemiology of this parasite, while the ability to detect intra-assemblage variation could have important uses in source tracking during outbreaks and studying environmental samples which could contain cysts from multiple hosts. An example of this is wastewater or irrigation water studies. These types of samples most likely contain cysts from a wide range of sources, including livestock and wildlife. Increasing our ability to unravel the potential diversity of *Giardia* genetic variants within these samples will be key to understanding public health risks and potential source of contamination. To provide tools for achieving these goals, 98% and 100% clustering thresholds were compared. The 98% clustering threshold was able to detect multiple assemblages of *G. duodenalis* within a single sample with the most abundant sequence variant of each assemblage being reported. At the 98% clustering threshold we observed 21 unique sequences among the samples included in this study with multiple unique nucleotide sequences present for all assemblages except A and *G. muris*. However, no within-sample intra-assemblage variation was observed. In comparison, when a clustering threshold of 100% was employed, within-sample intra-assemblage variation was common (Table 1). The number of total unique sequences among the samples increased to 39, and *G. muris* and all *G. duodenalis* assemblages but F displayed intra-assemblage variation within individual samples (Table 1). In fact, up to six unique sequence variants of assemblage E were observed within a single sample at the 100% clustering threshold while only a single unique sequence was observed using a 98% clustering threshold. Thus, clustering at 98% identity might be ideal for studies which aim to define only the presence of individual assemblages within samples, and a 100% identity clustering threshold could be used when SNP level differences could have important implications for the outcome of a study such as source tracking in outbreaks.

Often the multi-locus genotype studies using PCR amplification coupled with direct Sanger sequencing of the *bg*, *gdh*, *tpi*, and *SSU* rRNA genes identified that an assemblage based on one locus differs from the identified assemblage based on another locus (Feng and Xiao, 2011). These discrepancies have been associated with mixed infections combined with unintended selective amplification by PCR primers. In future studies, it will be interesting to incorporate additional loci using the NGS approach to better comprehend the reason behind these discrepancies.

5. Conclusions

Giardia mixed assemblage infections are often overlooked in molecular epidemiology studies in part due to limitations in current molecular detection techniques which frequently miss this type of infection. However, a better understanding of mixed assemblage infections and intra-assemblage variation could have important implications for understanding zoonotic potential and transmission dynamics of *G. duodenalis* as well as pathogenicity and host response to infection. It can also assist with drug design and vaccine development. There is a need for a more discriminatory typing tool and to fill that gap we have developed an NGS method using the *bg* gene and demonstrated both its ability to accurately detect *G. duodenalis* assemblages from DNA samples and to detect low abundance assemblages and intra-assemblage variation present in mixed infections. The application of this technique to study *Giardia* mixed infections in clinical and environmental samples from around the world will help to shed light on the occurrence and importance of mixed assemblage infections that have been for the most part overlooked.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank G. Solano Aguilar, J.P. Dubey, J.A. Cortés Vecino, D. Bowman, and J. Reed for providing the original *Giardia* cysts to obtain some of the DNA that were used in this study. We thank Nadja George and Josephine Hublin for their technical assistance. The study was supported by USDA-ARS Project No: 8042-32000-100-00-D.

References

Adell-Aledón, M., Köster, P.C., de Lucio, A., Puente, P., Hernández-de-Mingo, M., Sánchez-Thevenet, P., Dea-Ayuela, M.A., Carmena, D., 2018. Occurrence and molecular epidemiology of *Giardia duodenalis* infection in dog populations in eastern Spain. BMC Vet. Res. 14, 26.

Almeida, A., Pozio, E., Cacciò, S.M., 2010. Genotyping of *Giardia duodenalis* cysts by new real-time PCR assays for detection of mixed infections in human samples. Appl. Environ. Microbiol. 76, 1895–1901.

Broglia, A., Weitzel, T., Harms, G., Cacció, S.M., Nöckler, K., 2013. Molecular typing of *Giardia duodenalis* isolates from German travellers. Parasitol. Res. 112, 3449–3456.

Bushnell, B., 2014. BBMap. https://sourceforge.net/projects/bbmap/.
Cacciò, S.M., Ryan, U., 2008. Molecular epidemiology of giardiasis. Mol. Biochem. Parasitol. 160, 75–80.
Cacciò, S.M., Lalle, M., Svärd, S.G., 2018. Host specificity in the *Giardia duodenalis* species complex. Infect. Genet. Evol. 66, 335–345.
Covacin, C., Aucoin, D.P., Elliot, A., Thompson, R.C., 2011. Genotypic characterisation of *Giardia* from domestic dogs in the USA. Vet. Parasitol. 177, 28–32.
Dixon, B.R., 2015. Transmission dynamics of foodborne parasites on fresh produce. Foodborne Parasites in the Food Supply Web: Occurrence and Control. Elsevier Inc., pp. 317–353.
Efstratiou, A., Ongerth, J.E., Karanis, P., 2017. Waterborne transmission of protozoan parasites: review of worldwide outbreaks - an update 2011–2016. Water Res. 114, 14–22.
Elwin, K., Fairclough, H.V., Hadfield, S.J., Chalmers, R.M., 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. J. Med. Microbiol. 63, 38–44.
Feng, Y., Xiao, L., 2011. Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. Clin. Microbiol. Rev. 24, 110–140.
Geurden, T., Geldhof, P., Levecke, B., Martens, C., Berkvens, D., Casaert, S., Vercruysse, J., Claerebout, E., 2008. Mixed *Giardia duodenalis* assemblage A and E infections in calves. Int. J. Parasitol. 38, 259–264.
Geurden, T., Vanderstichel, R., Pohle, H., Ehsan, A., von Samson-Himmelstjerna, G., Morgan, E.R., Camuset, P., Capelli, G., Vercruysse, J., Claerebout, E., 2012. A multicentre prevalence study in Europe on *Giardia duodenalis* in calves, with molecular identification and risk factor analysis. Vet. Parasitol. 190, 383–390.
Gómez-Muñoz, M.T., Cámara-Badenes, C., Martínez-Herrero, M. del C., Dea-Ayuela, M.A., Pérez-Gracia, M.T., Fernández-Barredo, S., Santín, M., Fayer, R., 2012. Multilocus genotyping of Giardia duodenalis in lambs from Spain reveals a high heterogeneity. Res. Vet. Sci. 93, 836–842.
Grinberg, A., Biggs, P.J., Dukkipati, V.S.R., George, T.T., 2013. Extensive intra-host genetic diversity uncovered in *Cryptosporidium parvum* using next generation sequencing. Infect. Genet. Evol. 15, 18–24.
Hoar, B.R., Paul, R.R., Siembieda, J., Pereira, M.D.G.C., Atwill, E.R., 2009. *Giardia duodenalis* in feedlot cattle from the central and western United States. BMC Vet. Res. 5, 37.
Hopkins, R.M., Meloni, B.P., Groth, D.M., Wetherall, J.D., Reynoldson, J.A., Thompson, R.C.A., 1997. Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. J. Parasitol. 83, 44–51.
Iijima, Y., Itoh, N., Ito, Y., Kimura, Y., 2018. Multilocus genotyping of *Giardia duodenalis* isolates from household cats and pet shop kittens. Vet. Parasitol. 259, 44–48.
Julien, D.A., Sargeant, J.M., Guy, R.A., Shapiro, K., Imai, R.K., Bunce, A., Sudlovenick, E., Chen, S., Li, J., Harper, S.L., 2019. Prevalence and genetic characterization of *Giardia* spp. and *Cryptosporidium* spp. in dogs in Iqaluit, Nunavut, Canada. Zoonoses Public Health 66, 813–825.
Lalle, M., Pozio, E., Capelli, G., Bruschi, F., Crotti, D., Cacciò, S.M., 2005. Genetic heterogeneity at the β-giardin locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes. Int. J. Parasitol. 35, 207–213.
Li, W., Li, Y., Song, M., Lu, Y., Yang, J., Tao, W., Jiang, Y., Wan, Q., Zhang, S., Xiao, L., 2015. Prevalence and genetic characteristics of *Cryptosporidium*, *Enterocytozoon bieneusi* and *Giardia duodenalis* in cats and dogs in Heilongjiang province, China. Vet. Parasitol. 208, 125–134.
Maloney, J.G., Molokin, A., Santin, M., 2019. Next generation amplicon sequencing improves detection of *Blastocystis* mixed subtype infections. Infect. Genet. Evol. 73, 119–125.
Minetti, C., Taweenan, W., Hogg, R., Featherstone, C., Randle, N., Latham, S.M., Wastling, J.M., 2014. Occurrence and diversity of *Giardia duodenalis* assemblages in livestock in the UK. Transbound. Emerg. Dis. 61, e60–e67.
Paoletti, B., Traversa, D., Iorio, R., De Berardinis, A., Bartolini, R., Salini, R., Di Cesare, A., 2015. Zoonotic parasites in feces and fur of stray and private dogs from Italy. Parasitol. Res. 114, 2135–2141.
Paparini, A., Gofton, A., Yang, R., White, N., Bunce, M., Ryan, U.M., 2015. Comparison of Sanger and next generation sequencing performance for genotyping *Cryptosporidium* isolates at the 18S rRNA and actin loci. Exp. Parasitol. 151-152, 21–27.
Pickering, L.K., Woodward, W.E., DuPont, H.L., Sullivan, P., 1984. Occurrence of *Giardia lamblia* in children in day care centers. J. Pediatr. 104, 522–526.
Piekarska, J., Bajzert, J., Gorczykowski, M., Kantyka, M., Podkowik, M., 2016. Molecular identification of *Giardia duodenalis* isolates from domestic dogs and cats in Wroclaw, Poland. Ann. Agric. Environ. Med. 23, 410–415.
Ramo, A., Del Cacho, E., Sánchez-Acedo, C., Quílez, J., 2017. Occurrence and genetic diversity of *Cryptosporidium* and *Giardia* in urban wastewater treatment plants in north-eastern Spain. Sci. Total Environ. 598, 628–638.
Rendtorff, R.C., 1954. The experimental transmission of human intestinal protozoan parasites: II. *Giardia lamblia* cysts gwen in capsules. Am. J. Epidemiol. 59, 209–222.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Ryan, U., Zahedi, A., 2019. Molecular epidemiology of giardiasis from a veterinary perspective. Adv. Parasitol. 106, 209–254.
Ryan, U., Hijjawi, N., Feng, Y., Xiao, L., 2019. *Giardia*: an under-reported foodborne parasite. Int. J. Parasitol. 49, 1–11.
Santín, M., Trout, J.M., Fayer, R., 2009. A longitudinal study of *Giardia duodenalis* genotypes in dairy cows from birth to 2 years of age. Vet. Parasitol. 162, 40–45.
Santin, M., Dargatz, D., Fayer, R., 2012. Prevalence of *Giardia duodenalis* assemblages in weaned cattle on cow-calf operations in the United States. Vet. Parasitol. 183, 231–236.
Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States-major pathogens. Emerg. Infect. Dis. 17, 7–15.
Scorza, A.V., Lappin, M.R., 2017. Prevalence of selected zoonotic and vector-borne agents in dogs and cats on the pine ridge reservation. Vet. Sci. 4, 43.
Soliman, R.H., Fuentes, I., Rubio, J.M., 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.
Sprong, H., Cacciò, S.M., Van Der Giessen, J.W.B., 2009. Identification of zoonotic genotypes of *Giardia duodenalis*. PLoS Negl. Trop. Dis. 3, e558.
Tzanidakis, N., Sotiraki, S., Claerebout, E., Ehsan, A., Voutzourakis, N., Kostopoulou, D., Stijn, C., Vercruysse, J., Geurden, T., 2014. Occurrence and molecular characterization of *Giardia duodenalis* and *Cryptosporidium* spp. in sheep and goats reared under dairy husbandry systems in Greece. Parasite 21, 45.
Van Lith, L., Šoba, B., Vizcaino, V.V., Svard, S., Sprong, H., Tosini, F., Pozio, E., Cacciò, S.M., 2015. A real-time assemblage-specific PCR assay for the detection of *Giardia duodenalis* assemblages A, B and E in fecal samples. Vet. Parasitol. 211, 28–34.
Vermeulen, E.T., Lott, M.J., Eldridge, M.D.B., Power, M.L., 2016. Evaluation of next generation sequencing for the analysis of *Eimeria* communities in wildlife. J. Microbiol. Methods 124, 1–9.
Waldram, A., Vivancos, R., Hartley, C., Lamden, K., 2017. Prevalence of *Giardia* infection in households of Giardia cases and risk factors for household transmission. BMC Infect. Dis. 17, 486.
Yin, Y.L., Zhang, H.J., Yuan, Y.J., Tang, H., Chen, D., Jing, S., Wu, H.X., Wang, S.S., Zhao, G.H., 2018. Prevalence and multi-locus genotyping of *Giardia duodenalis* from goats in Shaanxi province, northwestern China. Acta Trop. 182, 202–206.