International Journal of Hygiene and Environmental Health 221 (2018) 102–111

Contents lists available at ScienceDirect

International Journal of Hygiene and Environmental Health

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

Multiple identification of most important waterborne protozoa in surface water used for irrigation purposes by 18S rRNA amplicon-based metagenomics

Y. Moreno[a,*,1], L. Moreno-Mesonero[a], I. Amorós[a], R. Pérez[a], J.A. Morillo[b], J.L. Alonso[a]

[a] Instituto Universitario de Ingeniería del Agua y Medio Ambiente, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
[b] Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas (EEZA-CSIC), Carretera de Sacramento s/n, La Cañada, Almería, Spain

ARTICLE INFO

Keywords:
18S amplicon sequencing
Waterborne protozoa parasites
Metagenomics
Irrigation water
Public health

ABSTRACT

Understanding waterborne protozoan parasites (WPPs) diversity has important implications in public health. In this study, we evaluated a NGS-based method as a detection approach to identify simultaneously most important WPPs using 18S rRNA high-throughput sequencing. A set of primers to target the V4 18S rRNA region of WPPs such as *Cryptosporidium* spp., *Giardia* sp., *Blastocystis* sp., *Entamoeba* spp, *Toxoplasma* sp. and free-living amoebae (FLA) was designed. In order to optimize PCR conditions before sequencing, both a mock community with a defined composition of representative WPPs and a real water sample inoculated with specific WPPs DNA were prepared. Using the method proposed in this study, we have detected the presence of *Giardia intestinalis*, *Acanthamoeba castellanii*, *Toxoplasma gondii*, *Entamoeba histolytica* and *Blastocystis* sp. at species level in real irrigation water samples. Our results showed that untreated surface irrigation water in open fields can provide an important source of WPPs. Therefore, the methodology proposed in this study can establish a basis for an accurate and effective diagnostic of WPPs to provide a better understanding of the risk associated to irrigation water.

1. Introduction

Parasite diversity has important implications in several research fields including ecology, evolutionary biology and epidemiology (Tanaka et al., 2014). Furthermore, protozoan pathogens are among the major risks of waterborne infections.

The contamination of drinking and bathing water with protozoan pathogens and the usage of sewage water for agricultural purposes poses a serious threat to millions of people worldwide (Plutzer and Karanis 2016). Among the waterborne protozoan pathogens (WPPs), *Giardia* and *Cryptosporidium* are the most common causes of major diarrheal outbreaks globally (Karanis et al., 2007). There is a plethora of information regarding these two pathogens, including their distribution/detection in water and related outbreaks (Karanis et al., 2007). Molecular taxonomic methods have identified *Cryptosporidium hominis* (which infects humans) and *Cryptosporidium parvum* (which infects cattles, humans and other mammals) as the most commonly detected species of *Cryptosporidium* in surface and wastewater (Paziewska et al., 2007; Smith et al., 2006). Giardiasis in humans and many other mammals is caused by *Giardia intestinalis* (Nguyen et al., 2016b). *Cryptosporidium* and *Giardia* have low infective doses and a marked resistance to environmental and water treatment stresses, which assists their dissemination, and have the potential to be transmitted from non-human to human hosts (zoonosis) and *vice versa*, enhancing the reservoir of (oo)cysts markedly (Smith et al., 2007).

In contrast, insufficient information is available for others WPPs such as *Cyclospora cayetanensis*, *Toxoplasma gondii*, *Isospora belli*, *Blastocystis hominis*, *Balantidium coli*, *Entamoeba histolytica* and other free-living amoebae (FLA) (Plutzer and Karanis 2016).

Blastocystis is a prevalent single-celled enteric parasite of unresolved clinical significance (Stensvold et al., 2007). *Blastocystis* is an emerging pathogen in terms of its association with disease and zoonotic potential (Thompson and Smith 2011). *Blastocystis* pathogens in humans comprise a group of at least 9 genetic subtypes (ST1-ST9) (Bart et al., 2013). Among them, ST1 and ST2 are the most common in water sources. Transmission of *Blastocystis* is suggested to occur the faecal-oral route, e.g. via contaminated water or food (Bart et al., 2013). Out of the zoonotic amoebae, *E. histolytica* is the causative agent of amoebiasis and undoubtedly of outmost clinical significance; it results in ~100,000 human deaths annually (Nakada-Tsuki and Nozaki 2016). It and has

* Corresponding author at:Research Institute of Water and Environmental Engineering IIAMA, Ed 8G, Acceso D, Planta 2, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain.
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijheh.2017.10.008
Received 1 September 2017; Received in revised form 4 October 2017; Accepted 18 October 2017

been the aetiological agent in 10 reported waterborne outbreaks (Baldursson and Karanis 2011).

Humans become infected with *Toxoplasma gondii* mainly by ingesting uncooked meat containing viable tissue cysts or by ingesting food or water contaminated with oocysts from the feces of infected cats (Jones and Dubey 2010). *T. gondii* was the etiological agent in 10 reported outbreaks (Plutzer and Karanis 2016).

FLA are ubiquitous protozoa that may behave as parasites under certain conditions (Plutzer and Karanis 2016). Four FLA belonging to *Acanthamoeba* spp., *Naegleria fowleri*, *Balamuthia mandrillaris*, and *Sappinia* spp. are known to cause infections in humans and animals, leading to severe brain pathologies (Schuster and Visvesvara 2004, Retana-Moreira et al., 2014, Baig 2015) or keratitis (*Acanthamoeba*) (Schuster and Visvesvara 2004) but their prevalence is generally low (Delafont et al., 2013). Sometimes, these amoebae can also bear pathogenic bacteria (Delafont et al., 2013) or *Cryptosporidium* oocysts (Scheid and Schwarzenberger, 2011). Among FLA, the water transmission of pathogenic strains of *Acanthamoeba* spp. and *Naegleria s*pp. is of great relevance (Karanis et al., 2007).

High quality information on the prevalence and detection of less frequent waterborne protozoa, such as *C. cayetanensis*, *T. gondii*, *I. belli*, *B. coli*, *B. hominis*, *E. histolytica* and other FLA, is not available (Plutzer and Karanis 2016). Over the past twenty years there have only been few improvements in the neglected waterborne protozoa monitoring and detection (Plutzer and Karanis 2016). Current opinion suggests that molecular techniques are the most promising methods for sensitive, accurate, and simultaneous detection of protozoan parasites in comparison to conventional staining and microscopy methods, which much benefit the water industry and public health (Fletcher et al., 2012). The introduction of molecular techniques, particularly those based on the amplification of nucleic acids, has provided researchers with highly sensitive and specific assays for the detection and quantification of protozoans. The use of sequencing data generated by massively parallel sequencing, also called´next generation sequencing‘ (NGS), is now commonplace in many fields of biological research (Hino et al., 2016). In the field of parasitology, pyrosequencing has been used for detecting and genotyping multiple infections of *T. gondii* (Sreekumar et al., 2005), genotyping of *Blastocystis isolates* (Stensvold et al., 2007) and studying the biodiversity and distribution of the genus *Acanthamoeba* (Fiore-Donno et al., 2016). Hino et al. (2016) introduced a novel method to assess the biodiversity of parasites −especially those in the host alimentary tract- using an 18S rRNA-based metagenomic approach (Parasitome analysis method).Tanaka et al. (2014) performed eukaryotic 18S rRNA-based metagenomics using an Illumina MiSeq sequencer and the analysis of the sequences using the QIIME software to assess biodiversity of helminth parasites in the alimentary tract of wild rats. These authors identified sequences in the 18S Illumina data from the rats that were assigned to taxa which included parasitic protozoa like *Trichomonas*, *Giardia* sp., *Trypanosoma* sp. and *Acanthamoeba* spp.

Outbreak incidents raise the question of whether the least frequent etiological agents of outbreaks are really less frequent in water (Plutzer and Karanis 2016). In some European countries, agricultural water consumption may represent up to 80% of the total water use. Current guidelines for the microbiological quality of water used to irrigate are based on the presence of coliforms bacteria and *Escherichia coli*. It remains unclear how these indexes correlate with the presence of specific human pathogens. The transmission of parasites and the role of emergent and new pathogens are not fully understood. Therefore, the development of new and rapid approaches is necessary to evaluate the role that irrigation water could have in the transmission of existing, new and emerging pathogens to the human population.

In this study a set of primers to be used in amplicon-based metagenomics have been designed and tested in a constructed DNA mock community and an inoculated sample to establish an optimized bioinformatic pipeline by which most of the WPPs could be detected. Moreover, this study focused on using the massive capacity of metagenomics to facilitate multiple WPPs detection in water samples used for irrigation, employing the same established set of primers and bioinformatic pipeline, to protect public health.

2. Experimental procedures

2.1. Reference genomic DNA

Genomic DNA (gDNA) from *C. hominis* was obtained from the *Cryptosporidium* Reference Unit (Chalmers R., National Public Health Service Microbiology, Swansea, UK). Genomic DNAs from *G. intestinalis* ATCC 30888D (Portland 1 strain), *C. parvum* ATCC PRA-67D (Iowa strain), *E. histolytica* ATCC 30459D (isolated from *E. histolytica* Schaudinn), *T. gondii* ATCC 50174D (strain RH) and *B. hominis* ATCC 50608D (strain BT1) were obtained from the American Type Culture Collection. Genomic DNAs were quantified using the Quant-iT™ dsDNA HS and BR Assay kits (Invitrogen, Thermo Fisher Scientific, USA) and the Qubit® 2.0 fluorometer following the manufacturer's instructions (Invitrogen).

2.2. Design and evaluation of 18S targeted primers

The specific sequences of the new primers designed in this work target the V4 18S rRNA hypervariable region of eukaryotes. The forward primer EUKAF and the reverse primer EUKAR (Table 1) were selected after examining the *in silico* alignment of 18S sequences obtained from the GenBank Database (www.ncbi.nlm.nih.gov/genbank/) (Fig. S1, Table S1), which specifically included sequences of *Cryptosporidium, Acanthamoeba* and *Giardia* species. The alignment was conducted by Clustal Omega program available online (http://www.ebi.ac.uk/Tools/msa/clustalo/). Thus, theorical amplicon lengths were checked to be compatible with Illumina MiSeq specifications (2 × 300 nt pair end reads) (Table 2). Self and cross-dimers were tested using OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html).

Specificity and taxonomic coverage of the primers was performed *in silico* by TestPrime 1.0 using SILVA database SSU 128 and the RefNR sequence collection (https://www.arb-silva.de/search/testprime/) (Klindworth et al., 2013). TestPrime runs an *in silico* PCR on SILVA databases and assigns a score depending on the mismatches between the primer and the sequence. Zero or 2 mismatches were allowed in this analysis.

Specificity was also assessed *in vitro* by conventional PCR with a battery of reference DNAs from the protozoan species described above. PCR conditions were optimized in order to amplify both GC-moderate and GC-rich regions of protozoan DNA. Different Taq polymerases and reaction buffers were tested for this purpose: Accuprime GC-rich DNA polymerase (Invitrogen, UK) and KAPA HiFi HotStart plus GC buffer (KAPABiosystems, USA). Furthermore, several chemicals such as betaine, acetamide, DMSO, glycerol and BSA were added to the mix as PCR enhancers of the reaction as according to Kramer and Coen (2001). PCR reaction consisted in a final volume of 25 µL containing 5X reaction buffer, 7.5 mM dNTPs mix, 5 µM of each primer and 1 U of Taq polymerase. Cycling conditions were 95 °C for 5 min followed by 28 cycles of amplification (denaturation at 98 °C for 30 s, annealing at

Table 1
18S rRNA primers used in this work. Yeast 5′ position is based on *Saccharomyces cerevisiae,* GenBank accession number Z75578. GC range and Tm range were calculated using the online tool http://www.biophp.org/minitools/melting_temperature/demo.php

Primer ID	Sequence (5′-3′)	Yeast 5′ position	GC range (%)	Tm range (°C)
EUKAF	GCC GCG GTA ATT CCA GCT C	571	63.2	55.4
EUKAR	CYT TCG YYC TTG ATT RA	980	29.4	37.4–47,1

Table 2
Genome characteristics of Waterborne protozoa species included in the assay and theorical amplicon size obtained by the primers EUKAF and EUKAR designed.

ID	Organism	Amplicon size (bp)	Genome size	Ploidy	rDNA copies	references
ATCC PRA-67D	*Cryptosporidium parvum*	374	9.11Mb	1N	5 (H)	Torres-Machorro et al. (2010)
U07408	*Cryptosporidium hominis*	377	9.16Mb	1N	5 (H)	Drumo et al. (2012)
ATCC 30010	*Acanthamoeba castellanii*	540	33Mb	25N	24(H);600	Torres-Machorro et al. (2010)
ATCC 30459D	*Entamoeba histolytica*	396	24Mb	4N	200	Torres-Machorro et al. (2010)
ATCC 50174D	*Toxoplasma gondii*	409	63Mb	1N	110	Torres-Machorro et al. (2010)
ATCC 50237	*Vermamoeba vermiformis*	415	ND	4N	1330	Kuiper et al. (2006)
ATCC 50608D	*Blastocystis hominis*	451	18.8Mb	ND	17	Poirier et al. (2011)
ATCC 30888D	*Giardia intestinalis*	265	12Mb	4-8N	60(H); 300 cyst	Torres-Machorro et al. (2010)

appropriate temperature for 30 s and extension at 72 °C for 30 s) and a final elongation step of 72 °C for 5 min. Annealing temperatures tested were 55 °C, 57 °C and 60 °C. Amplicon sizes were visualized on a 1.5% agarose gel.

2.3. Mock community DNA

A mock community was generated as a control sample to assess the suitability of the primers, sequencing conditions and bioinformatics analysis. According to different molecular weights of each specie DNA and the ploidy of the organisms (Table 2), different amounts of reference DNAs (representing in all cases 10^3 genomic units) from the protozoa species *G. intestinalis, C. parvum, E. histolytica, T. gondii, A. castellanii* and *B. hominis* were mixed into a single tube. This mock community was sequenced and analyzed as described below. Moreover, qPCRs described below were used to check sequencing results in this community.

2.4. Water samples

In order to evaluate the suitability of the high-throughput method designed in this study on environmental samples, we carried out the same analysis as in the mock sample to identify pathogenic protozoa in 3 surface irrigation water samples which mix with wastewater from surrounding farms (samples A1-A3). All samples were collected in sterile bottles and processed within 24 h of collection. For DNA isolation, a total of 1.5 L of each sample were concentrated by centrifugation at 2500 g for 15 min and finally resuspended in 1 mL of phosphate buffered saline (PBS 1X buffer; Thermo Fisher Scientific, USA).

Simultaneously, 500 mL of the same samples were also concentrated by centrifugation at 2500 g for 15 min to evaluate the presence of *Cryptosporidium* oocysts and *Giardia* cyst by the standardized method IMS-IFA as described below.

Moreover, one of the irrigation water samples (A1) was inoculated with DNA of *Giardia*, *Cryptosporidium* and *Acanthamoeba* reference strains included in this work at the same concentration as in the mock sample (sample IS) in order to test the good performance of PCR prior to sequencing in this type of samples.

2.5. DNA extraction

Total DNA from water samples was extracted using the UNEX protocol (Hill et al., 2015). The UNEX procedure includes pre-treatment with proteinase K, lysis of the microbes with the UNEX buffer, and inclusion of PCR facilitator (acetylated BSA in real-time mastermix) (Hill et al., 2015). One volume of 125 µL of the proteinase K enzyme (20 mg/mL stock concentration) was added to a volume of 1 mL of the lysis buffer-water sample concentrate mixture. After this addition, the sample was kept at room temperature for 15 min and transferred to a lysing matrix E tube (each impact-resistant 2.0 mL matrix E tube contains 1.4 mm ceramic spheres, 0.1 mm silica spheres, and one 4 mm glass) (MP Biomedicals, USA). To facilitate disruption of (oo)cysts, a homogenization step was included using the FastPrep-24® instrument (MP Biomedicals). Samples, in lysing matrix E tube, underwent two rounds of vigorous mixing at 6.5 m/s for 60 s as recommended by Shields et al. (2013) before starting UNEX protocol. Samples were placed on ice for two minutes between each round of mixing. After the disruption step the lysing matrix tube E was centrifuged at 10,000 x g for 1 min. The supernatant was passed through a nucleic acid-binding silica column (GeneJET Genomic DNA Purification Column, Fermentas, Thermo Fisher Scientific, USA) for washing and elution. The eluate was spinned through a PVPP column (OneStep™ PCR Inhibitor Removal kit, Zymo Research, CA, USA) at 8000 x g for 1 min. Final DNA was stored at −20 °C. DNA from water samples was sequenced and analyzed as described below.

2.6. Sequencing

Illumina sequencing was carried out on a MiSeq platform at FISABIO Sequencing and Bioinformatics Service (Valencia, Spain). The amplicon sequencing protocol targets the V4 region of the 18S gene with the primers designed surrounding conserved regions. DNA amplicon libraries were generated as described by Illumina guide in the documentation provided by Illumina (http://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf) with the modifications above described and using the enzyme KAPA HiFi HotStart with GC buffer (5X) (KAPABIOSYSTEMS, USA). Using a limited cycle PCR, Illumina sequencing adaptors and dual-index barcodes were added to each amplicon. The Illumina overhang adapter sequences to be added to primer- specific sequences were the forward 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG- and the reverse 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG. Libraries were then normalized and pooled prior to sequencing. Samples containing indexed amplicons were loaded onto the MiSeq reagent cartridge and onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual indexes reads was performed (2 × 300 bp run).

2.7. Bioinformatic analysis

Illumina data were analyzed using QIIME 1.9.1 (Caporaso et al., 2010) applying additional scripts available in Microbiome Helper virtual box (Comeau et al., 2017). As a first step, forward and reverse reads were merged using PEAR v0.9.19 (Zhang et al., 2014). FastQC (Andrews 2010) was used to confirm that reads were correctly stitched. Subsequently, stitched reads were filtered by length and quality score (reads with less than 200 bp or a minimum of Q30 over at least 90% of the read were removed) using FASTX-Toolkit v0.0.14 (Gordon 2009). Reads with any ambiguous base ("N") were also filtered out. After an additional examination with FastQC the low quality tails of the merged sequences were removed. Potential chimeric sequences were screened out using VSEARCH v1.11.1. (Rognes et al., 2016). The remaining sequences were processed using the QIIMÉs open reference script, applying the methods SortMeRNA v2.0 (Kopylova et al., 2012.) and SUMACLUST v1.0.00 (Mercier et al., 2013) for the reference-based and

Table 3
Cycling conditions of the WPPs qPCR.

Target gene	Primer name: Sequence (5′→3′)	amplicon size (bp)*	qPCR conditions[a]	Ref
Acanthamoeba 18SrRNA gene	Acant900-F: CCCAGATCGTTTACCGTGAA	180	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 8 s at 63 °C and 7 s at 72 °C	Qvarnstrom et al. (2006)
	Acant1100-R: TAAATATTAATGCCCCCAACTATCC			
	Acant1000-P:6-FAM-CTGCCACCGAATACATTAGCATGG-BHQ1			
Cryptosporidium COWP gene	COWP-F: CAAATTGATACCGTTTGTCCTTCTG	151	Denaturation 10 min at 95°, 40 cycles of 10 s at 95 °C, 8 s at 66 °C and 6 s at 72 °C	Guy et al., 2003; Alonso et al. (2011)
	COWP-R: GGCATGTCGATTCTAATTCAGCT			
	COWP-P:6-FAM-TGCCATACATTGTTGTCCTGACAAATTGAAT-TAMRA			
Giardia ß-giardin gene	P241-F: CATCCGCGAGGAGGTCAA	74	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 8 s at 58 °C and 7 s at 72 °C	Guy et al., 2003; Alonso et al. (2011)
	P241-R: TCCAATCTGGGCATAAGATTTG			
	P241-P:6-FAM-AAGTCCGCCGACAACATGTACCTAACGA-TAMRA			
G. duodenalis	P434P1-F: AGTCCACGAGATTGGCAACA	74	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 8 s at 58 °C and 7 s at 72 °C	Guy et al., 2003; Alonso et al. (2010)
genotype A	P434P1-R: GCAGGTTAAAGATTTTCACG			
ß-giardin gene	P434P1-P: 6-FAM-AATTATTCGAGATGGCGCCCACG-BHQ1			
G. duodenalis	P434H3-F: AAGGTGCCTACAAGCGAAGT	74	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 8 s at 58 °C and 7 s at 72 °C	Guy et al., 2003; Alonso et al. (2010)
genotype B	P434H3-R: TTCGTTTAGGGCAAGGTAGC			
ß-giardin gene	P434H3-P: 6-FAM-ATCGTGGGGTCGCCCGTACG-BHQ1			
Toxoplasma	Toxo-F: AGTGACAAGAAATAACAACACTGG	315	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 15 s at 65 °C and 14 s at 72 °C	Lalonde and Gajadhar (2011)
18S rDNA gene	Toxo-R: CCTGCTTGAAACACTCTAATTTTC			
Blastocystis	Blasto FWD-F5: GGTCCGGTGAACACTTTGGATTT	118	Denaturation 10 min at 95 °C, 40 cycles of 10 s at 95 °C, 8 s at 58 °C and 5 s at 72 °C	Stensvold et al. (2012)
18S rDNA gene	Blasto R-F2: CCTACGGAAACCTTGTTACGACTTCA			
gene	Blasto-P: 6-FAM-TCGTGTAAATCTTACCATTTAGAGGA-BHQ1			

[a] Toxoplasma and Blastocystis qPCR conditions were optimized for LightCycler 2.0 platform in this study.

de novo clustering steps, respectively. OTUs were defined at the 97% genetic similarity cut-off. The PR2 protist ribosomal database pr2_gb203_version_4.5.fasta (Gillou et al., 2013) was used as the reference.

2.8. Quantitative polymerase chain reaction (qPCR)

All samples were screened for the presence of *Cryptosporidum*, *Giardia*, *Acanthamoeba*, *Toxoplasma* and *Blastocystis*. The specific primers, TaqMan probes and qPCR conditions for the detection and quantification of *Acanthamoeba* spp., *Cryptosporidium* spp., *Giardia* spp., *G. intestinalis* genotype A and *G. intestinalis* genotype B, *Toxoplasma* and *Blastocystis* are indicated in Table 3. Amplifications were made in duplicate. qPCRs and data analysis were performed using the LightCycler 2.0 PCR system (Roche, Barcelona, Spain), with the qPCR cycling conditions described in Table 3. Standard curves were prepared from a 10-fold serial dilutions of DNA purified from cysts and oocysts for *Cryptosporidium*, *Giardia* (Guy et al., 2003) and *Toxoplasma* (Lalonde and Gajadhar, 2011) qPCRs and from 10-fold serial dilutions of a synthetic DNA containing the specific sequence of the amplicon generated from *Acanthamoeba* and *Blastocystis* qPCRs. The quantification cycle (Cq) value along with a standard curve were used to calculate the quantification. Detection limits of each qPCR reaction were stablished by the authors cited in Table 3 as follow: *Acanthamoeba* spp (1 amoeba/assay; 16 amoeba/L of water); *Cryptosporidium* spp (1 oocyst or 4 COW gene copies/assay; 16 oocysts/L of water); *Giardia* spp. (16 β-giardin copies/assay; 1 cyst/L of water); *Toxoplasma* (10 oocysts/assay; $1.66{\cdot}10^2$ oocysts/L of water); *Blastocystis* (2.5 parasites/assay; 32 parasites/L of water).

2.9. Detection of Giardia and Cryptosporidium by IMS-IFA

Enumeration of *Cryptosporidium* oocysts and *Giardia* cysts from irrigation water was carried out according to USEPA Method 1623 (US Environmental Protection Agency, 2005). A volume of 500 mL water sample was concentrated by centrifugation at 3500 rpm for 15 min, and the pellet was resuspended in 10 mL of distilled water in a Leighton tube. Immunomagnetic separation (IMS) was conducted using the commercially available Dynabeads GC−Combo kit (Life Technologies AS, Oslo, Norway) according to the manufacturer's instructions. The final concentrate from the IMS was dried overnight at room temperature and labeled with fluorescent monoclonal antibody for *Giardia* and *Cryptosporidium* immunofluorescence assay (IFA) according to the manufacturer's protocol (Merifluor®; Meridian Bioscience Inc., OH, USA) A blue filter (excitation, 480 nm; emission, 520 nm) was used to detect fluorescein isothiocyanate–conjugated MAb-labeled (oo) cysts.

3. Results

3.1. Designed primers

The primers EUKAF and EUKAR were designed to amplify the variable region V4 of 18S rRNA gene since after alignment of the database sequences good matches between the pathogenic protozoa of this study were found. Furthermore, other authors suggested that this region is rich in taxonomic information (Hugerth et al., 2014). These new primers matched 86,2% of the total pool of 67,380 full-length 18S eukaryotic SILVA sequences which included 100% coverage of those belonging to the most important pathogenic protozoa, such as *Cryptosporidium*, *Giardia*, *Acanthamoeba*, *Vermamoeba*, *Blastocystis*, *Toxoplasma*, *Balantidium*, and *Cyclospora* species when 2 mismatches were allowed (Table 4). A total of 4 degenerated residues were included in the EUKAR primer in order to increase the primer coverage.

PCR conditions were optimized by the amplification of the reference protozoan DNAs included in the study. An annealing temperature of 57 °C showed to be the optimal to amplify the DNA of all the genera cited above. Amplification of both *Giardia* and *Cryptosporidium* 18S specific fragments was only possible using KAPA HiFi HotStart polymerase with GC buffer (Fig. S2). The rest of additives tested in the amplification reaction, only enhanced *Giardia* detection in detriment of the other genera assayed. Amplicon sizes obtained were optimal for MiSeq sequencing.

Table 4
Theorical primers coverage for the protozoa in the SILVA database.

		0 mismatches	2 mismatches
Kingdom		%	%
Archaea	Total	0,0	0,0
Bacteria	Total	0,0	0,0
Eukaryota	Total	52,1	86,2
	Acanthamoeba	88,8	91,0
	Balamuthia	83,3	100,0
	Balantidium	100,0	100,0
	Blastocystis	98,2	100,0
	Cryptosporidium	92,3	94,2
	Cyclospora	0,0	100,0
	Echinamoeba	100,0	100,0
	Giardia	63,6	90,9
	Naegleria	0,0	85,7
	Toxoplasma	100,0	100,0
	Vermamoeba	85,7	100,0
	Vannella	94,7	100,0

3.2. Generation of 18S rRNA amplicon sequencing data

A duplicated DNA mock community was used to evaluate the utility of the amplicon sequencing approach developed to study a mixture of pathogenic protozoa. After quality filtering, trimming and detection of PCR-chimeras, the molecular analysis of protozoa mock community yielded 288,316 high-quality reads (428.04 ± 14.21 bp) which were clustered in 57 eukaryotic OTUs with assigned taxonomy. Taxonomy was assigned at 97% similarity against PR2 database. All species included in the mock were recovered in the sequenced data (Table 5). No differences were obtained between the numbers of OTUs obtained at the different cut-offs tested (57 OTUs). However, the relative abundance of the species differed from expected since the amount of DNA of each species in the mock was equivalent to 103 cells/mL (Table 2). Approximately 260,933 of the reads obtained belonged to *E. histolytica* and 24,783 reads belonged to *Blastocystis*. spp. On the other hand, other taxa were underrepresented, 766 reads corresponded to *A. castellanii*, 776 to *G. intestinalis* and only 273 to *C. hominis*. Although rRNA copies of 18S vary in the different taxa included in the mock community sample, it does not correspond with the bias yielded.

In order to test the potential inhibition of the PCR prior to sequencing with the designed primers, one of the irrigation water samples was inoculated with DNA of *Giardia*, *Cryptosporidium* and *Acanthamoeba* reference strains included in this work (sample IS). Results showed that this PCR was not affected by the samples so the number of reads obtained from the inoculated sample was similar to those obtained from the mock community sample (Table 6). Furthermore, detection of

Table 5
Results obtained after analysis of the Illumina data from the Mock reference community DNA.

	Mock sample		
Taxonomy	n° OTUS	Reads	%
Cryptosporidium parvum	2	273	0,09
Toxoplasma gondii	2	752	0,26
Entamoeba histolytica	31	260936	90,50
Acanthamoeba castellanii	2	766	0,26
Giardia intestinalis	6	675	0,23
Blastocystis sp.	9	24794	8,60

Table 6
Comparison of results obtained after analysis of the Illumina data from the mock reference community DNA and the sample A1 inoculated with *Acanthamoeba*, *Giardia* and *Cryptosporidium* DNA at the same concentration (IS) and qPCR.

Sample	*Acanthamoeba*		*Giardia*		*Cryptosporidium*	
	Reads	qPCR	Reads	qPCR	Reads	qPCR
A1	0	+	4	+	0	−
IS	826	+	1419	+	272	+
M	765	+	659	+	279	+

Acanthamoeba, *Giardia* and *Cryptosporidium* in IS sample was also confirmed by specific qPCRs (Table 6).

A total of 9034 raw sequences were generated from the irrigation water samples after Illumina MiSeq sequencing (A1: 5784; A2: 1396; A3: 1854). After quality analysis and joining paired–end- reads, 6428 sequences remained (A1: 3796; A2: 1010; A3: 1622). A total of 123 OTUs with assigned taxonomy were obtained at 97% similarity threshold. Forty eight OTUs were no taxonomically assigned. Nine OTUs of these 123, belonged to the pathogenic species included in the study. *Blastocystis* sequences were clustered in 4 OTUs, 2 belonged to *B. hominis* specie and the other 2 to *Blastocystis* spp. *Acanthamoeba*, *Giardia* and *Entamoeba* sequences were clustered in 1 OTU each. Relative abundances of some potentially pathogenic protozoa present in the samples were very low in all samples (Fig. 1), representing the total sequence percentages from 0.026 to 1.396%. All sequences were identified at species level. The most abundant were *E. histolytica* and *Blastocystis* sp., which were detected in all samples (Table 7). *T. gondii* and *A. castellanii* sequences were only recovered from 1 sample, sample A1 and A2 respectively, and in very low percentage out of the total abundance. *G. intestinalis* was also detected in 2 samples (A1, A3) with low relative abundance. No sequence of *Cryptosporidium* was obtained in any of the samples.

Furthermore, some representative sequences of each OTU were checked by nucleotide BLAST alignment tool at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and showed matchings higher than 97%. Among the *Blastocystis* sequences, ST2 and ST4 subtypes were identified and included into different OTUs (data not showed). A total of 4 OTUs of *Blastocystis* were recovered.

In parallel, by qPCR analysis, the presence of *Blastocystis* spp. in samples A1 and A2, *A. castellanii* in samples A1 and A2; *T. gondii* in the sample A1 and *G. intestinalis* genotype A and B in the sample A1 were detected. No samples yielded positive results for the presence of *Cryptosporidium* spp. by qPCR (Table 7). None of the qPCR analysis was inhibited and all inoculated species were amplified.

The number of *Cryptosporidium* oocysts and *Giardia* cysts detected by immunofluorescence microscopy in water samples are showed in Table 7. Oocysts and cysts were detected in all the samplings. *Giardia* cyst counts ranged between 180–768 cysts/L, while *Cryptosporidium* oocyst counts were lower, ranging from 4 to 8 oocysts/L (Table 7).

4. Discussion

18S rRNA amplicon sequencing is showed as an effective and

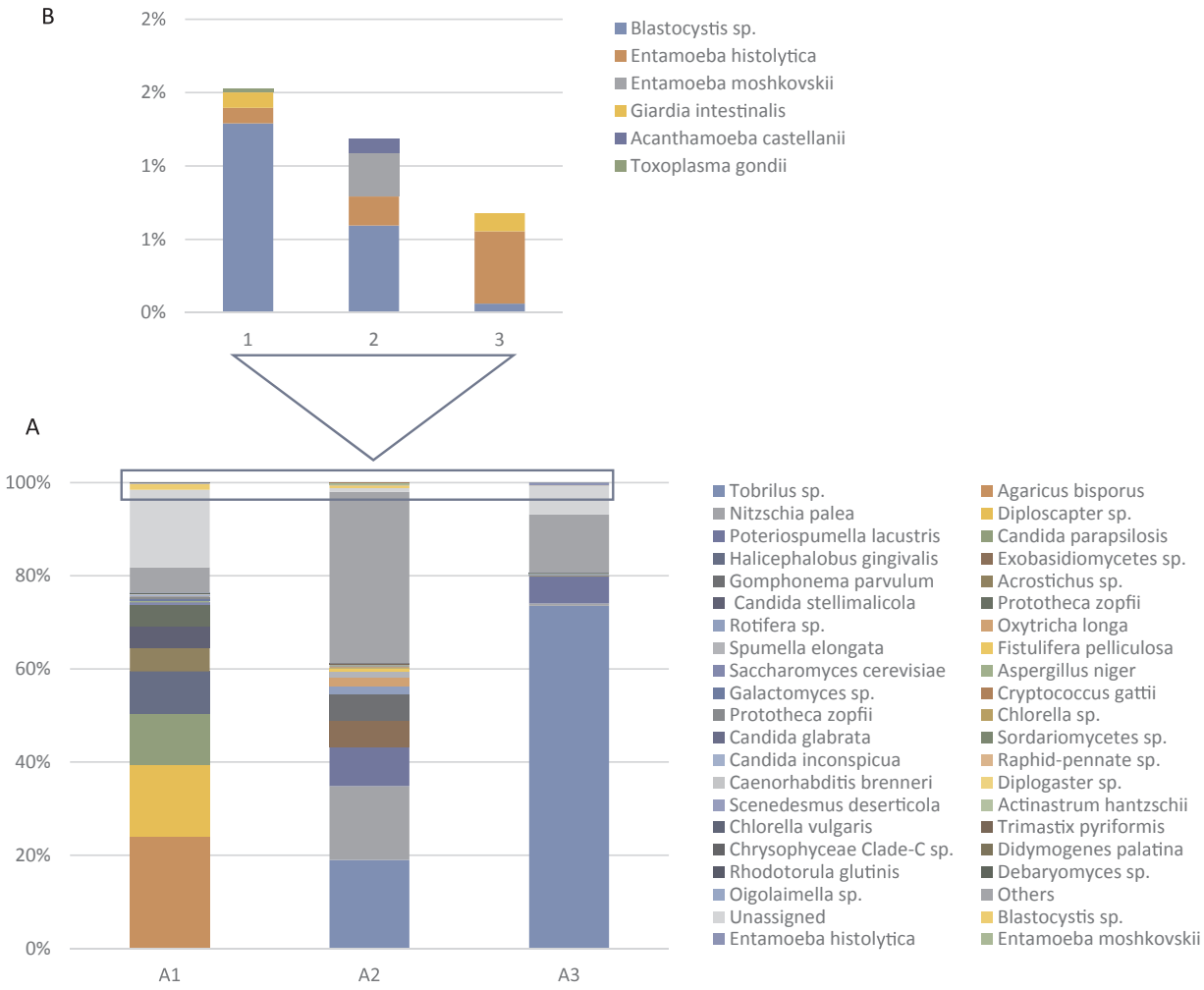

Fig. 1. Relative abundances of Eukariotic groups in 18S data set (A) and of the waterborne protozoa species (B) from the irrigation water samples.

Table 7
WPP identified in irrigation water samples.

Species	A1				A2				A3			
	[a]Reads	%	qPCR	[b] IFA	Reads	%	qPCR	IFA	Reads	%	qPCR	IFA
Toxoplasma gondii	1	0,026	+	+	0	0	−	ND	0	0	−	ND
Entamoeba histolytica	4	0,105	ND	ND	2	0,198	ND	ND	8	0,493	ND	ND
Cryptosporidium spp.	0	0	−	8	0	0	−	4	0	0	−	0
Acanthamoeba castellanii	0	0	+	+	1	0,099	+	ND	0	0	+	ND
Giardia intestinalis	4	0,105	+	710	0	0	−	80	2	0,123	−	0
Blastocystis sp.	49	1291	+	ND	6	0,594	+	ND	1	0.062	+	ND

[a] Reads are the number of sequences obtained from Illumina 18S amplicon sequencing and the percentage represents the relative abundance of each sample.
[b] number of (oo) cysts.

sensitive method to study the eukaryotic diversity (Tanaka et al., 2014; Kim et al., 2017). However, only a few studies using NGS technologies for pathogenic protozoa identification have been published (Vermeulen et al., 2016; Cooper et al., 2016). The current study is focused on the development and evaluation of a 18S rRNA amplicon-based sequencing approach as an emerging detection method to identify the most important WPPs in irrigation samples.

18S rRNA gene have different hypervariable regions (V1 to V9) flanked by conserved regions suitable to design primers for PCR sequencing. A number of recent studies have compared variable regions along the entire 18S rRNA gene for all eukaryotes and eukaryotic plankton and highlighted conserved regions that may be best suited for amplifying hypervariable regions (Bradley et al., 2016). Pawlowsky et al. (2012) recommended the V4 region for protists studies due to the high taxonomic resolution. In this study, we have designed EUKAF and EUKAR primers to amplify the V4 region, which is long enough to allow the differentiation of the pathogenic protozoa included in the study and compatible with Illumina MiSeq specifications (2 × 300 nt) as well. Furthermore, other authors suggested that this region is rich in taxonomic information to study eukaryotic diversity (Hugerth et al., 2014). Amplicon lengths obtained with these primers were very similar for *C. parvum*, *E. histolytica*, *T. gondii* and *B. hominis*, therefore the biases in the amplification due to the different fragment size could be avoided. Anyway, the performance of the amplification in this region seems not to be related to the amplicon size since the one of *G. intestinalis* was the smallest and this fact did not enhance its amplification in the mock community sample. PCR-amplification of GC-rich templates is often hampered by the formation of secondary structures like hairpins and higher melting temperatures. Some solutes have been used in PCR in order to increase the specificity and to reduce the formation of those secondary structures. Since *Giardia* 18S rRNA is high in GC content and its amplification with standard polymerase buffers was not achieved, different reagents and PCR conditions to optimize the reaction were tested in this work. Finally, using KAPA HiFi HotStart polymerase with GC buffer, all WPP species included in the study were successfully amplified, including *G. intestinalis*.

Some authors recommended the use of mock communities in order to study the variability of the results due to PCR errors, number of rRNA copies of each protozoa, primers affinity within the mixed community and bioinformatics analysis of data (Wegener, 2015; Bradley et al., 2016). We observed in the mock sample that we recovered the least number of reads from *Cryptosporidium* which is the protozoan with less 18S rRNA copies (5 copies). Nevertheless, the highest numbers of reads were recovered from *Blastocystis*, whose 18S rRNA copies are lower than the ones from *Toxoplasma*, *Acanthamoeba*, *Entamoeba* and *Giardia*. Then, we suppose that PCR may have a bias which enhances the amplification of the genus *Blastocystis*. According to other authors (Geisen et al., 2015), the differences observed between the amount of inoculated DNA in the mock and the sequences recovered from each species showed that absolute or relative quantification of protist communities result extremely difficult. It is therefore hard to conclude that a difference in number of gene copies might modify the genus ratio.

As primer binding is a limiting factor to the amplification of DNA from environmental samples (Hugerth et al., 2014) and the presence of PCR inhibitors in those samples is a recurrent problem, a sample inoculated with reference DNA should be analyzed as a control in all environmental studies. The analysis of 18S amplicons in IS sample (artificially inoculated A1 sample) showed, when compared with A1 sample, that some genus detected in A1 were not found in IS. Furthermore, similar numbers of reads belonging to the inoculated species were recovered from the mock and from sample A1, where the inoculated DNA was in presence of the total 18S community of the water sample. Both Illumina and qPCR analysis from the sample IS showed no PCR inhibition in this type of sample.

There are different bioinformatic approaches to analyze the sequencing data obtained from environmental microbiome. Recently, Comeau et al., 2017 have reported a bioinformatic standard operating procedure to allow a rapid and reliable Illumina data analysis using QIIME 1.9.1. The SILVA 128 database is the most commonly used as a reference to classify 18S Illumina reads in QIIME. However, the protozoan sequences in the SILVA database are limited and are not broad enough to cover waterborne protozoan parasite diversities (Tanaka et al., 2014). Therefore, a more complete protist database such as PR2 provides a more powerful tool to study waterborne protozoan parasite diversities. We used PR2 database to classify taxonomically the sequences obtained, achieving an accurate identification at species level. Since the main objective of this work has been to detect simultaneously the presence of most pathogens present in water samples, singletons have not been eliminated from the data. In this study, we did not aim to perform a comparison among the different samples. Due to the low presence of the WPPs of interest in the samples, the samples have been analyzed individually, without a normalization step, following the same procedure of previous studies also focused on the detection of pathogens (Cooper et al., 2016) in order to increase the probability of detection.

Once optimized this approach in both the mock community and the inoculated sample, we tested it to detect WPP in real water samples. Using this method, we have detected the presence of important WPP such as *G. intestinalis*, *A. castellanii*, *T. gondii*, *E. histolytica* and *Blastocystis* spp. in the analyzed irrigation water samples.

Giardia intestinalis and *Cryptosporidium* spp., are the most commonly reported protozoa associated with enteric infections and are associated mainly with food and waterborne outbreaks. The use of 18S rRNA locus has been recommended as a target for screening *Giardia* and *Cryptosporidium* from stool and environmental samples owing to its high copy numbers within the genome, thus enhancing detection sensitivity (Nguyen et al., 2016a). Despite this, no sequence of *Cryptosporidium* was recovered from Illumina data in any of the environmental samples. Therefore, we are not sure that this 18S rRNA region allows to discriminate between species such as *C. parvum* and *C. hominis* since the theoretical similarity between sequences is higher than 97% (Fig. S1).In spite of the high sensitivity of the technique (1 oocyst) (Guy et al.,

2003), qPCR detection was also negative for this pathogen in all water samples analyzed. As expected, low levels of *Cryptosporidium* oocysts were found in our samples by IFA. Other studies in irrigation waters also showed similar counts. An average of 0.4 oocysts/L were reported in the Saint Lawrence River in Canada (Payment et al., 2000), while in 6 rivers in North America, 10.1% of the samples were positive for ooycsts with an average range of < 0.001-0.069 oocysts/L, with occurrence exceeding 1 oocyst/L during peak rain events (Le Chevallier et al., 2003). Forty-eight percent of the surface water used for irrigation purposes in Mexico were positive for Cryptosporidium oocysts and 50% tested positive for Giardia cysts. Concentration ranges for Cryptosporidium oocysts and Giardia cysts were 17–200 and 17–1633 per 100 L, respectively (Chaidez et al., 2005). Besides the low presence of this pathogen, the poor sensitivity of the molecular detection could be due to the presence of empty oocyst and cyst lacking in DNA (Nguyen et al., 2016a)

18S rRNA *Giardia* sequences were recovered from sample A1 and A3 (Table 4) while qPCR detection was only successful for sample A1. However, *Giardia* cysts were identified by IFA in all samples analyzed with a higher percentage in the sample A1. Carmena et al. (2007) in a study in northern Spain also reported that, in most cases, *Giardia* prevalence was higher than that of *Cryptosporidium*, corroborating the tendency observed in other countries (Rimhanen-Finne et al., 2004; Briancesco and Bonadona, 2005).

Other Eukaryotes identified in the analyzed water samples by Illumina sequencing in this work, such as *Blastocystis* spp., *T. gondii*, *E. histolytica* and other FLA such as *A. castellanii*, are emerging as important causes of illness, with serious implications for travellers to developing regions, immunocompromised populations, and young children. Nevertheless, insufficient information is available due to the lack of efficient methods to detect them from the environment (Plutzer and Karanis, 2016).

Blastocystis spp. was the most represented protozoa among the sequences retrieved from the 18S gene. Among the 17 subtypes of *Blastocystis*, ST1-ST4 collectively account for 90% of human carriage, with ST5-ST9 accounting for the remaining 10% (Plutzer and Karanis 2016). Several regions of the 18S small subunit rRNA have been used by different authors for *Blastocystis* genotyping. In our study, the V4 designed primers target a region of the SSU rRNA gene that would facilitate the detection and identification of different subtypes in irrigation waters. Subtypes ST1 and ST4 have been found as the most common subtypes in previous studies (Bart et al., 2013). In our study, the subtypes ST2 and ST4 have been found among the samples from the partial retrieved sequences.

Entamoeba histolytica sequences were also identified among the OTUs obtained from all analyzed water samples. Its prevalence rates and water surveillance data from Europe and North America are not available although outbreaks have been reported (Plutzer and Karanis 2016). Recently, *E. histolytica* has been reclassified into different species, *E. histolytica*, *E. dispar* and *E. moshkovskii*, being the two latter species non-pathogenic (Plutzer and Karanis, 2016). Although the species *E. histolytica* and *E. moshkovskii* cannot be differentiated by microscopic traditional methods, this metagenomic analysis is able to differentiate both species in the sample A2. This fact is very important, so pathogenic and non-pathogenic species could be distinguished.

Although 18S rRNA amplicons obtained with the designed primers from *A. castellanii* were the longest (> 500 pb), sequences were successfully recovered from the mock sample. But although *A. castellanii* is an ubiquitous protozoa that has been detected very frequently in water samples (Magnet et al., 2013), an only sequence of *A. castellanii* was recovered from sample A2. This result is in accordance to those obtained by specific qPCR from *Acanthamoeba* spp., in the analyzed samples. The only sequence recovered was assigned to genotype T4 by BLAST search against NCBI database (data not showed). T4 genotype has been reported to be the most abundant and widespread *Acanthamoeba* in all types of environments (Fiore-Donno et al., 2016) and is directly related to keratitis (Visvesvara et al., 2007). Just as *Acanthamoeba*, even though the set of primers fully matched with the 18S fragment of *T. gondii* and sequences were recovered from the mock sample, an only sequence was detected in the water sample A1. By specific qPCR only the same sample yielded positive results for *T. gondii* detection.

The role of water and food, particularly fresh produce as a source for these protozoan agents is now well-recognized as documented in traceback-outbreak investigations, which is carried out through the detection of oocysts and cysts in vegetables and water samples, e.g. when irrigated in fields (Amoros et al., 2010; Smith et al., 2006). Water contaminated with untreated sewage effluent is commonly used for agricultural irrigation purposes and poses a significant source of infection to humans, through the consumption of raw produce (Nguyen et al., 2016a).

Unfortunately, many waterborne pathogens are still difficult to detect, and despite advances in molecular diagnostics, such methods are not widely available or used even in developed countries (Fletcher et al., 2012). Pathogenic protozoa are present in water in enough numbers to produce illness but not sufficient to be detected. To our knowledge, this is the first report where the presence of important waterborne protozoa such as *Cryptosporidium*, *Giardia*, *Acanthamoeba*, *Toxoplasma*, *Entamoeba* and *Blastocystis* species were detected simultaneously in irrigation water samples using Illumina 18S rRNA amplicon-based sequencing. This method does not require knowledge of protozoan morphologies in laborious traditional identification methods. Most importantly, 18S Illumina sequencing is able to identify more WPP than traditional methods.

5. Conclusions

A set of primers to be used in amplicon-based metagenomics were designed and evaluated in order to facilitate detection and identification of the most important WPPs in a high-throughput manner using a specific bioinformatic pipeline. We found this set of primers present 90–100% coverages of WPPs such as *Cryptosporidium*, *Giardia*, *Acanthamoeba*, *Vermamoeba*, *Blastocystis*, *Toxoplasma*, *Balantidium*, or *Cyclospora* species.

Results showed that untreated surface irrigation water in the open fields can provide a source of WPPs, since by the optimized approach developed in this study, we have recovered sequences from *Giardia*, *Acanthamoeba*, *Blastocystis*, *Toxoplasma* and *Entamoeba* species.

Then, we propose the use of the NGS methodology developed in this study together with the standard identification methods as a powerful approach to detect simultaneously and specifically the most important WPPs present in different environments. However, further analysis of a higher number of samples is required to demonstrate definitively the usefulness of the method.

Acknowledgments

This work was supported through the project funded by the Spanish Ministry of Economy and Competitiveness (MINECO) in the frame of the collaborative international consortium JPIW2013-095-C03-02 of the Water Challenges for a Changing World Joint Programming Initiative (Water JPI) Pilot Call. R. Pérez acknowledges support from MINECO program “Promoción de Empleo Joven e Implantación de la Garantía Juvenil en I + D + i”.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijheh.2017.10.008.

References

Alonso, J.L., Amorós, I., Cuesta, G., 2010. LNA probes in a real-time PCR assay for genotyping of *Giardia duodenalis* in wastewaters. J. Appl. Microbiol. 148, 1594–1601.

Alonso, J.L., Amorós, I., Cañigral, I., 2011. Development and evaluation of a real-time PCR assay for quantification of *Giardia* and *Cryptosporidium* in sewage samples. Appl. Microbiol. Biotechnol. 89, 1203–1211.

Amoros, I., Alonso, J.L., Cuesta, G., 2010. Cryptosporidium *Cryptosporidium* oocysts and *Giardia* cysts on salad products irrigated with contaminated water. J. Food Prot. 73, 1138–1140.

Andrews, S., 2010. FastQC: a Quality-control Tool for High-throughput Sequence Data. Babraham Institute, Cambridge, United Kingdom. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baig, A.M., 2015. Pathogenesis amoebic encephalitis: are the amoebas being credited to an 'inside job' done by the host immune response. Acta Trop. 148, 72–76.

Baldursson, S., Karanis, P., 2011. Waterborne transmission of protozoan parasites: a review of worldwide outbreaks −an update 2004–2010. Water Res. 45, 6603–6614.

Bart, A., Wentink-Bonnema, M.S., Gilis, H., Verhaar, N., Wassenaar, C.J.A., van Vugt, M., Goorhuis, A., van Gool, T., 2013. Diagnosis and subtype analysis of Blastocystis sp. In 442 patients in a hospital setting in the Netherlands. BMC Infect. Dis. 13, 389. http://dx.doi.org/10.1186/1471-2334-13-389.

Bradley, I.M., Pinto, A.J., Guest, J.S., 2016. Design and evaluation of Illumina MiSeq-compatible: 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. 82, 5878–5891.

Briancesco, R., Bonadona, L., 2005. An italian study on *Cryptosporidium* and *Giardia* in wastewater: fresh water and treated water. Environ. Monit. Assess. 104, 445–457.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Gonzalez, A., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336.

Carmena, D., Aguinagald, X., Zigorraga, C., Fernández-Crespo, J.C., Ocio, J.A., 2007. Presence of *Giardia* cysts and *Cryptosporidium* oocysts in drinking water supplies in northern Spain. J. Appl. Microbiol. 102, 619–629.

Chaidez, C., Soto, M., Gortares, P., Mena, K., 2005. Occurrence of Cryptosporidium and Giardia in irrigation water and its impact on the fresh produce industry. Int. J. Environ. Health. Res. 15, 339–345.

Comeau, A.M., Douglas, G.M., Langille, M.G.I., 2017. Microbiome helper: a custom and streamlined workflow for microbiome research. mSystems 2 (1), e00127–16. http://dx.doi.org/10.1128/mSystems.00127-16.

Cooper, M.K., Phalen, D.N., Donahoe, S.L., Rose, K., Slapeta, J., 2016. The utility of diversity profiling using Illumina 18S rRNA gene amplicon deep sequencing to detect and discriminate *Toxoplasma gondii* among the cyst-forming coccidia. Vet. Parasitol. 216, 38–45.

Delafont, V., Brouke, A., Bouchon, D., Moulin, L., Héchard, Y., 2013. Microbiome of free-living amoebae isolated from drinking water. Water Res. 47, 6958–6965.

Drumo, R., Widmer, G., Morrison, L.J., Tait, A., Grelloni, V., D'Avino, N., et al., 2012. Evidence of host-associated populations of Cryptosporidium parvum in Italy. Appl. Environ. Microbiol. 78, 3523–3529.

Fiore-Donno, A.M., Weinert, J., Wubet, T., Bonkowski, M., 2016. Metacommunity analysis of ameboid protists in grassland. Sci. Rep. 6, 19068. http://dx.doi.org/10.1038/srep19068.

Fletcher, S.M., Stark, D., Harkness, J., Ellis, J., 2012. Enteric protozoa in the developed world: a public health perspective. Clinical Microbiol. Rev. 25, 420–429.

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., de Groot, G.A., 2015. Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. Mol. Ecol. 24, 4556–4569. http://dx.doi.org/10.1111/mec.13238.

Gordon, A., 2009. FASTX-Toolkit: FASTQ/A Short-reads Pre-processing Tools. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. http://hannonlab.cshl.edu/fastx_toolkit/.

Guy, R.A., Payment, P., Krull, J., Horgen, P.A., 2003. Real-time PCR for quantification of *Giardia* and *Cryptosporidium* in environmental water samples and sewage. Appl. Environ. Microbiol. 69, 5178–5185.

Hill, V.R., Narayanan, J., Gallen, R.R., Ferdinand, K.L., Cromeans, T., Vinjé, J., 2015. Development of a nucleic acid extraction procedure for simultaneous recovery of DNA and RNA from diverse microbes in water. Pathogens 4, 335–354.

Hino, A., Maruyama, H., Kikuchi, T., 2016. A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing: parasitome analysis method. Parasitol. Int. 65, 572–575.

Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., Wilmes, P., Andersson, A.F., 2014. Sysyematic design gene primers for determining eukaryotic diversity in microbial consortia. PLoS One 9 (4), 1–11.

Jones, J.L., Dubey, J.P., 2010. Waterborne toxoplasmosis–recent developments. Exp. Parasitol. 124, 10–25.

Karanis, P., Kourenti, C., Smith, H., 2007. Waterborne transmission of protozoan parasites: a worldwide review of outbreaks and lessons learnt. J. Water Health. 5, 1–38.

Kim, H., Kim, H., Hwang, H.S., Kim, W., 2017. Metagenomic analysis of the marine coastal invertebrates of South Korea as assessed by Ilumina MiSeq. Anim. Cells Syst. 21 (1), 37–44.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41 (1), e1.

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. http://dx.doi.org/10.1093/bioinformatics/bts611.

Kramer, M.F., Coen, D.M., 2001. Enzymatic amplification of DNA by PCR: standard procedures and optimization. Curr. Protoc. Mol. Biol. 56, 1–4 (15.1:15.1. 1–15.1.14).

Kuiper, M.W., Valster, R.M., Wullings, B.A., Boonstra, H., Smidt, H., van der Kooij, D., 2006. Quantitative detection of the free-living amoeba Hartmannella vermiformis in surface water by using real-time PCR. Appl. Environ. Microbiol. 72, 5750–5756.

Lalonde, L.F., Gajadhar, A., 2011. Detection and differentiation of coccidian oocysts by real-time PCR and meeting curve analysis. J. Parasitol. 97, 723–730.

LeChevallier, M.W., Di Giovanni, G.D., Clancy, J.L., et al., 2003. Comparison of Method 1623 and Cell Culture-PCR for Detection of Cryptosporidium spp. in Source Waters. Appl. Environ. Microbiol. 62, 971–979. http://dx.doi.org/10.1128/AEM.69.2.971-979.2003.

Magnet, A., Fenoy, S., Galván, A.L., Izquierdo, F., Rueda, C., Vadillo, C.F., Del Águila, C., 2013. A year long study of the presence of free living amoeba in Spain. Wat. Res. 47, 6966–6972.

Mercier, C., Boyer, F., Bonin, A., Coissac, E., 2013. SUMATRA and SUMACLUST: Fast and Exact Comparison and Clustering of Sequences. http://metabarcoding.org/sumatra/.

Nakada-Tsuki, K., Nozaki, T., 2016. Immune response of amebiasis and immune evasion by *Entamoeba histolytica*. Front. Immunol. 7, 175. http://dx.doi.org/10.3389/fimmu.2016.00175.

Nguyen, T.T., Traub, R., Pham, P.D., Nguyen, H.V., Nguyen, K.C., Phun, C.D., Dalsgaard, A., 2016a. Prevalence and molecular characterization of *Cryptosporidium* spp. and *Giardia* spp. in environmental samples in Hanam province Vietnam. Food Waterborne Parasitol. 3, 13–20.

Nguyen, S.T., Fukuda, Y., Nguyen, D.T., Tada, C., Nakai, Y., 2016b. Prevalence and first genotyping of *Giardia duodenalis* in beef calves in Vietnam. Trop. Anim. Health Prod. 48 (4), 837–841.

Pawlowsky, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirku, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukes, J., Mann, D.G., Mitchell, E.A., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G., Smirnov, A.V., Spouge, J.L., Stern, R.F., Stoeck, T., 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 10 (11), e1001419. http://dx.doi.org/10.1371/journal.pbio.1001419.

Payment, P., Berte, A., Prévost, M., Ménard, B., Barbeau, B., 2000. Occurrence of pathogenic microorganisms in the Saint Lawrence River (Canada) and comparison of health risks for populations using it as their source of drinking water. Can. J. Microbiol. 46, 565–576.

Paziewska, A., Bernarska, M., Nieweglowski, H., Karbowiak, G., Bajer, A., 2007. Distribution of *Cryptosporidium* and *Giardia* spp. in selected species of protected and game mammals from North-eastern Poland. Ann. Agric. Environ. Med. 14, 265–270.

Plutzer, J., Karanis, P., 2016. Neglected waterborne parasitic protozoa and their detection in water. Water Res. 101, 318–332.

Poirier, P., Wawrzyniak, I., Albert, A., El Alaoui, H., Delbac, F., Livrelli, V., 2011. Development and Evaluation of a Real-Time PCR Assay for Detection and Quantification of Blastocystis Parasites in Human Stool Samples: Prospective Study of Patients with Hematological Malignancies. J Clin Microbiol 49, 975–983. http://dx.doi.org/10.1128/JCM.01392-10.

Qvarnstrom, Y., Visvesvara, G.S., Sriram, R., Da Silva, A.J., 2006. Multiplex real-time PCR assay for simultaneous detection of *Acanthamoeba* spp *Balamuthia mandrillaris*, and *Naegleria fowleri*. J. Clin. Microbiol. 44, 3589–3595.

Retana-Moreira, L., Abrahams-Sandi, E., Cabello-Vílchez, A.M., Reyes-Batle, M., Valladares, B., Martínez-Carretero, E., Piñero, J.E., Lorenzo-Morales, J., 2014. Isolation and molecular characterization of *Acanthamoeba* and *Balamuthia mandrillaris* from combination shower units in Costa Rica. Parasitol. Res. 113, 4117–4122.

Rimhanen-Finne, R., Vuorinen, A., Marmo, S., Malmberg, S., Hänninen, M.-L., 2004. Comparative analysis of *Cryptosporidium*, *Giardia* and indicator bacteria during sewage sludge hygienization in various composting processes. Lett. Appl. Microbiol. 38, 301–305.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. http://dx.doi.org/10.7717/peerj.2584.

Scheid, P.L., Schwarzenberger, R., 2011. Free-living amoebae as vectors of cryptosporidia. Parasitol. Res. 109, 499–504.

Schuster, F.L., Visvesvara, G.S., 2004. Free-living amoebae as opportunistic and non-opportunistic pathogens of humans and animals. Int. J. Parasitol. 34, 1001–1027.

Shields, J.M., Joo, J., Kim, R., Murphy, H.R., 2013. Assessment of three commercial DNA extraction kits and a laboratory-developed method for detecting *Cryptosporidium* and *Cyclospora* in raspberry wash, basil wash and pesto. J. Microbiol. Meth. 92, 51–58.

Smith, H.V., Caccio, S.M., Tait, A., McLauchlin, J., Thompson, R.C., 2006. Tools for investigating the environmental transmission of *Cryptosporidium* and *Giardia* infections in humans. Trends Parasitol. 22, 160–167.

Smith, H.V., Cacciò, S.M., Cook, N., Nichols, R.A.B., Tait, A., 2007. Cryptosporidium *Cryptosporidium* and *Giardia* as foodborne zoonoses. Vet. Parasitol. 149, 29–40.

Sreekumar, C., Hill, D.E., Miska, K.B., Vianna, M.C., Yan, L., Myers, R.L., Dubey, J.P., 2005. Genotyping and detection of multiple infections of *Toxoplasma gondii* using Pyrosequencing. Int. J. Parasitol. 35, 991–999.

Stensvold, C.R., Traub, R.J., von Samson-Himmelstjerna, G., Jespersgaard, C., Nielsen, H.V., Thompson, R.C.A., 2007. Blastocystis *Blastocystis*: subtyping isolates using pyrosequencing technology. Exp. Parasitol. 116, 111–119.

Stensvold, C.R., Ahmed, U.N., Andersen, L.O., Nielsen, H.V., 2012. Development and evaluation of a genus-specific, probe-based, internal-process-controlled real-time PCR assay for sensitive and specific detection of *Blastocystis* spp. J. Clin. Microbiol. 50, 1847–1851.

Tanaka, R., Hino, A., Tsai, I.J., Palomares-Rius, J.E., Yoshida, A., Ogura, Y., Hayashi, T., Maruyama, H., Kikuchi, T., 2014. Assessment of helminth biodiversity in wild rats using 18S rDNA based metagenomics. PLoS One 9 (10), e110769. http://dx.doi.org/10.1371/journal.pone.0110769.

Thompson, R.C.A., Smith, A., 2011. Zoonotic enteric protozoa. Vet. Parasitol. 182, 70–78.

Torres-Machorro, A.L., Hernández, R., Cevallos, A.M., López-Villaseñor, I., 2010. Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? FEMS Microbiol. Rev. 34, 59–86. http://dx.doi.org/10.1111/j.1574-6976.2009.00196.x.

U.S. EPA, 2005. Method 1622: Cryptosporidiumin Water by Filtration/IMS/FA. EPA 815-R-05-001.

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. Met. 124, 1–9.

Visvesvara, G.S., Moura, H., Schuster, F.L., 2007. Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris*, *Naegleria fowleri*, and Sappinia diploidea. FEMS Immunol. Med Microbiol . 50, 1–26. http://dx.doi.org/10.1111/j.1574-695X.2007.00232.x.

Zhang, J., Kobert, K., Flour, i T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina paired-end reAd mergeR. Bioinformatics 30, 614–620. http://dx.doi.org/10.1093/bioinformatics/btt593.