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

Please cite this article as: Moreno, Y., International Journal of Hygiene and Environmental Health (2017), http://dx.doi.org/10.1016/j.ijheh.2017.10.008

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* spp. 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