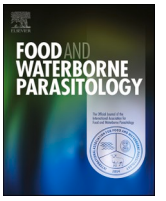
Food and Waterborne Parasitology 39 (2025) e00265

Contents lists available at ScienceDirect

Food and Waterborne Parasitology

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



Detection of parasites in food and water matrices by shotgun metagenomics: A narrative review☆

Paolo Vatta, Simone M. Cacciò *

Unit of Foodborne and Neglected Parasitic Diseases, Department of Infectious Diseases, Istituto Superiore di Sanità, Viale Regina Elena 299, Rome 00161, Italy

ARTICLE INFO

Keywords:
Parasites
Food
Water
Shotgun metagenomics
Bioinformatics
Review

ABSTRACT

Many helminths and protozoa are transmitted to humans through the consumption of contaminated food or water, and this underlines the importance of methods for their detection in these matrices. Due to the difficulties in isolating parasites prior to their identification, indirect detection methods are used, mostly relying upon targeted amplification of nucleic acids via PCR and/or qPCR. With the development of high throughput sequencing technologies, an untargeted detection method, shotgun metagenomics, became available. By sequencing the total DNA extracted from a given source, and through bioinformatics analyses of the sequencing reads, shotgun metagenomics allows profiling the entire microbial community therein present, including eukaryotes and, therefore, parasites. In this article, we reviewed the studies that specifically addressed the detection of parasites in food ($n = 2$) and water matrices ($n = 10$) by shotgun metagenomics. Most studies focused on wastewater samples and reported the detection of many parasites of human and veterinary importance from various areas of the world, highlighting the potential of shotgun metagenomics to provide important data for parasitic pathogens surveillance. After examining the different analytical workflows employed in these studies, which were not developed for detection of eukaryotes (or parasites), we identified two aspects deserving attention. First, that assignment based on short reads matching ribosomal sequences may generate false positives due to high sequence conservation among eukaryotic organisms. Second, that reassessing the relatively small number of reads of eukaryotic origin by a BLAST search can confirm, or deny, identification of parasitic pathogens.

1. Introduction

The introduction of the Next Generation Sequencing (NGS) technologies has had an enormous impact in many areas of research, including infectious diseases, cancer, and the study of the microbiome (Satam et al., 2023). These advancements in sequencing technologies also rapidly changed the way studies of the composition, functions and interactions of microbial communities in different ecosystems are conducted. The focus of these studies has been mostly on bacteria and viruses, whereas the abundant and diverse eukaryotic taxa has received less attention, but understanding the role they play as decomposers, predators, producers and parasites is an area of intense research (Bik et al., 2012).

☆ This article is part of a Special issue entitled: 'EMOP 2024' published in Food and Waterborne Parasitology.

* Corresponding author.
E-mail addresses: paolo.vatta@iss.it (P. Vatta), simone.caccio@iss.it (S.M. Cacciò).

https://doi.org/10.1016/j.fawpar.2025.e00265
Received 27 March 2025; Received in revised form 17 April 2025; Accepted 18 April 2025
Available online 21 April 2025

For the purpose of this review, it is important to distinguish the two DNA-based NGS approaches, namely targeted metagenomics (or metataxonomics), which refers to the sequencing of specific marker genes, usually ribosomal genes, and untargeted metagenomics (or shotgun metagenomics), which refers to the sequencing of total DNA from a sample without selecting for a particular target. The term "metagenome" was first coined to indicate the theoretical collection of all genomes from members of a microbial community from a specific environment (Handelsman et al., 1998). There has been some ambiguity in the use of terms such as metataxonomics, metabarcoding and metagenomics, but this has been clarified in a number of publications (e.g., Marchesi and Ravel, 2015). In this review, we focus specifically on shotgun metagenomics, as defined above.

In its application to water and food matrices, the vast majority of the studies has focused on prokaryotes or viruses, and has been paralleled by the development of a number of bioinformatics tools and workflows for data analysis (e.g., Breitwieser et al., 2019). As an example, wastewater-based epidemiology is becoming an important tool in infectious disease surveillance capacity, and, in this context, NGS technologies allow to detect, characterize, and track pathogens at the community level (Farkas et al., 2024). Similarly, shotgun metagenomics allows to track foodborne pathogens or spoilers in foods and food-handling environments (Sekse et al., 2017), and, along with other approaches like nutrigenomics, proteomics, metabolomics and transcriptomics, are at the base of the field of foodomics (Sadurski et al., 2024).

The detection of eukaryotic organisms in shotgun metagenomics data is influenced by issues related to specificity and sensitivity. Indeed, the relative scarcity of complete genomes, their larger complexity compared to prokaryotic ones, and the fact that eukaryotes represent an extremely small fraction of the sequencing pool obtained from food or water samples, account for the importance of these issues.

Some analytical workflows have been proposed to facilitate the detection of eukaryotes in complex datasets, including CCMetagen (Marcelino et al., 2020), EukDetect (Lind and Pollard, 2021), CORRAL (Bazant et al., 2023) and MetaPhlAn4 (Blanco-Míguez et al., 2023). However, these workflows have not yet been tested on shotgun metagenomics data from food or water matrices.

Parasites are, of course, eukaryotes and there is a growing interest towards the use of shotgun metagenomics for their detection in these matrices, which often play a fundamental role in the overall epidemiology of this very diverse group of pathogens. The simultaneous detection of parasites, other eukaryotes (e.g., fungi), viruses and bacteria present in food and water matrices, offered by shotgun metagenomics, is an essential aspect in ecological studies.

In this article, we reviewed the studies that specifically addressed the detection of parasites in food and water samples using shotgun metagenomics, to then examine the analytical workflows employed and the results obtained. Finally, we underline critical aspects that, in our opinion, need to be considered when analyzing these complex datasets.

Table 1
List of the studies that used shotgun metagenomics for detection of parasites in food and water matrices.

Matrix and origin	Sequencing platform	Range of n. of reads (Gbp)	Analytical workflow used	Database used	Reference
Spiked smoked salmon (Italy)	IonTorrent and Illumina	$0{,}6–8{,}0 \times 10^7$ (0,82–12,2 Gbp)	MG-RAST	NCBI RefSeq	Sala et al. (2020)
Romaine lettuce (Sweden)	Illumina HiSeq2500	$1{,}7–2{,}2 \times 10^8$ (43–54 Gbp)	Centrifuge	NCBI nucleotide	Ahlinder et al. (2022)
Drinking water (USA)	IonTorrent	$4{,}8 \times 10^6$ to 3×10^7 (0,77–4,8 Gbp)	CosmosID	GenBook®	Brumfield et al. (2020)
Surface water (Rivers, Haiti)	IonTorrent	$1{,}7 \times 10^7$ (3,4 Gbp)	CosmosID	GenBook®	Roy et al. (2018)
Surface water (Little Bighorn river, USA)	MinION	$\sim 4 \times 10^5$ (~1,1 Gbp)	CosmosID and MG-RAST	GenBook®	Hamner et al. (2019)
Surface water (Arid wetlands sediment samples, South Africa)	Illumina NextSeq 2000	$1{,}18–2{,}0 \times 10^7$ (0,18–0,3 Gbp)	NCBI Nephele	NCBI RefSeq	Ogola et al. (2024)
Sewage (urban informal settlement, Kenia)	Illumina HiSeq	$0{,}78–10{,}8 \times 10^7$ (2×10^9–$2{,}7 \times 10^{10}$ Gbp)	MGMapper v2.2	NCBI RefSeq	Hendriksen et al. (2019)
Sewage (Treatment plant, Switzerland)	Illumina HiSeq 4000	$1{,}7–2{,}2 \times 10^7$ (2,5–3,3 Gbp)	MG-RAST	PR^2 (rDNA database)	Freudenthal et al. (2022)
Sewage (Treatment plant, Turkey)	Illumina NextSeq 500	$9{,}4–28{,}5 \times 10^7$ (1,41–4,28 Gbp)	BWA	Protozoan database from NCBI	Gündoğdu et al. (2023)
Sewage (Treatment plant, New Zealand)	Illumina NovaSeq6000	$6{,}9 \times 10^5$ to $1{,}5 \times 10^7$ (0,13–0,22 Gbp)	Kraken v2.07	NCBI RefSeq	Ariyadasa et al. (2023)
Sewage (Treatment plant, South Africa)	Illumina NexSeq	$0{,}71–0{,}93 \times 10^7$ (3,6 Gbp)	Kraken v2.07	Custom eukaryotic database	Mthethwa-Hlongwa et al. (2024)
Sewage (Gutter water, Congo)	Illumina HiSeq 4000	$8{,}4 \times 10^7$ (25 Gbp)	Kaiju	Custom eukaryotic database	Moumen et al. (2024)

2. Application of shotgun metagenomics for detection of parasites in food

We searched Pubmed and Scopus electronic databases in December 2024 using the terms shotgun metagenomics AND food AND (parasite OR protozoa OR helminths). This search returned two publications (Table 1). The first was a proficiency test in which a mock community comprising the Apicomplexan *Cryptosporidium parvum*, six bacterial and three viral pathogens, was used to spike a piece of cold smoked salmon (Sala et al., 2020). The spiked food was then delivered to the participating laboratories and processed using the different wet-lab techniques routinely in use. The metagenomics datasets were all analyzed using a single bioinformatics pipeline (MG-RAST). The results showed that the wet-lab procedures (e.g., method for DNA extraction, library preparation and sequencing platform) had an impact on the relative read abundance of the pathogens comprised in the mock community, although all pathogens, including *C. parvum*, were detected. It should be noted that, to achieve an expected relative read abundance of 0.2 %, an extremely high spike level (one million *C. parvum* oocysts) was used.

The second study (Ahlinder et al., 2022) focused on a foodborne outbreak in Sweden linked to the consumption of romaine lettuce contaminated with *Cryptosporidium parvum*. Vegetables were processed using a Stomacher and the eluates used for DNA extraction by the PowerSoil®DNA Isolation Kit (Mo Bio Laboratories Inc.). The presence of *C. parvum* in the suspected food vehicle was demonstrated using a combination of PCR-based typing, amplicon sequencing and shotgun sequencing. The shotgun metagenomics reads from three romaine lettuce samples were mapped against the reference genome of *C. parvum* (Iowa II strain) using BWA (Table 1). This showed a mapping frequency of 0.7, 1.0 and 1.8 % in the three samples. It should be noted that the libraries were sequenced using a single lane on an Illumina HiSeq2500 platform, which generated a very large number of reads (43–54 Gbp).

3. Application of shotgun metagenomics for detection of parasites in water matrices

We searched Pubmed and Scopus electronic databases in December 2024 using the terms shotgun metagenomics AND water AND (parasite OR protozoa OR helminths). This search returned ten publications (Table 1).

3.1. Drinking water

Brumfield et al. (2020) used shotgun metagenomics to analyze samples of tap, drinking fountain, sparkling natural mineral, and non-mineral bottled water, and determined their microbiological content. Water samples were concentrated by filtration and DNA extracted from the filter membranes using the ZymoBIOMICS™ DNA Miniprep Kit (Zymo Research, Irvine, CA, USA). Libraries were sequenced on an Ion S5 XL Semiconductor Sequencer (Ion Torrent, Thermo Fisher Scientific, USA) to generate between $4{,}8 \times 10^6$ and 3×10^7 reads (Table 1). Raw reads were analyzed using the CosmosID Metagenomics Cloud Application, and microbes were identified at the species, subspecies, and/or strain level. Their relative abundance was also quantified. Concerning eukaryotes, *Acanthamoeba palestinensis* was detected in all drinking water samples, except artesian and bottled reprocessed tap water, while *Acanthamoeba mauritaniensis* was detected in bottled spring water and in municipal tap water. These two species are, along with other *Acanthamoeba* species, the causative agents of granulomatous amebic encephalitis and amebic keratitis, and have been associated with cutaneous lesions and sinusitis. Their occurrence in various aquatic environments, but also in air and soil, is well documented (da Silva et al., 2024). No other parasites were identified.

3.2. Surface water

Roy et al. (2018) used shotgun metagenomics to characterize the microbial community in surface water samples collected at two time points from several sites near the original epicenter of the cholera outbreak in the Central Plateau of Haiti. Water samples were concentrated by filtration and the DNA extracted by using the DNeasy PowerWater Sterivex Kit (Qiagen). Libraries were sequenced on an Ion S5 XL Semiconductor Sequencer (Ion Torrent, Thermo Fisher Scientific, USA) to generate an average of $1{,}7 \times 10^7$ reads. Raw reads were analyzed by the CosmosID metagenomics software (Table 1). The analysis showed the presence of *Plasmodium falciparum* at almost all sites, with high abundance at one Hinche River site, and further suggested the presence of *Entamoeba* spp., *Toxoplasma gondii* and *Trypanosoma congolense*. Finally, *Acanthamoeba polyphaga* was detected at two sites.

Hamner et al. (2019) used shotgun metagenomics to characterize the microbial community in water samples collected at the Little Bighorn River in Montana, USA. Water samples were concentrated by filtration and the DNA extracted by using the PowerWater DNA isolation kit (Qiagen). Libraries were sequenced using the MinION sequencing platform (Oxford Nanopore), which generated 397,884 reads with an average read length of 2760 bp. Raw sequences were analyzed by the CosmosID and MG-RAST software (Table 1). The analysis showed the presence of 55 reads supporting the presence of various species of *Acanthamoeba* (*A. castellani*, *A. healyi*, *A. polyphaga*, *A. palestinensis* and *A. quina*) and 78 reads supporting the presence of various species of *Leishmania* (*L. major*, *L. donovani*, *L. arabica*, *L. infantum*, *L. turanica*, *L. aethiopica*).

Ogola et al. (2024) performed a shotgun metagenomics study of sediment of permanent and seasonal arid freshwater wetlands across north-eastern South Africa. DNA was extracted from the samples using the DNeasy® PowerSoil Pro Kit (Qiagen, Germany). Libraries were sequenced on an Illumina NextSeq 2000 platform (2x150bp reads), to generate between 1,18 to $2{,}0 \times 10^7$ reads (Table 1). Raw reads were classified using Kraken2 v2.1.1 against the standard NCBI database, and each read was assigned to the lowest common ancestor using exact kmer matching. Subsequently, phylum-, genus-, and species-level taxonomic profiles were generated using Bayesian re-estimation of kraken2 classifications via bracken. Finally, the protistan OTU sequences extracted underwent analysis using the phyloseq v1.30.0 and vegan v2.3–5. At the phylum level, Apicomplexa, Euglenazoa, Amoebazoa,

Bacillariophyta, Cryptophyta and Microsporidia emerged as the dominant phyla. At the genus level, *Plasmodium*, *Leishmania*, and *Besnoita* were the predominant taxa, with *Theileria*, *Dictyostelium*, *Toxoplasma*, *Trypanosoma*, *Neospora*, *Thalassiosira*, *Phaeodactylum*, *Babesia* and *Hemiselmis* also detected. Finally, at the species level, *Leishmania mexicana, L. major*, *Besnoita besnoiti*, *Plasmodium vivax, P. berghei*, *P. chabaudi, P. malariae*, *P. relictum*, *P. coatneyi*, *P. yoelii*, *P. cynomolgi*, *P. vinckei*, *Toxoplasma gondii*, *Trypanosoma brucei*, *Neospora caninum*, and *Theileria parva* were identified.

3.3. Wastewater

Hendriksen et al. (2019) conducted a temporal metagenomics analysis of urban sewage from Kibera, an informal settlement in Nairobi, Kenya. Samples ($n = 42$) were collected during the dry season (June to August). Samples were centrifuged and the pelleted material used for DNA extraction using the QIAamp Fast DNA Stool mini kit (Qiagen). Libraries were sequenced on the Illumina HiSeq platform to generate between 0,78 and 10,8 $\times$ 10^7 reads (Table 1). Raw reads were analyzed using MG mapper v2.2 to detect and quantify bacterial and associated antimicrobial resistance (AMR) determinants, as well as viral and parasitic pathogens. A database of genome sequence from GenBank and other resources was used for the read mapping procedure. The parasitic pathogens identified were *Plasmodium* spp., *Cryptosporidium* spp., *Giardia* spp., *Blastocystis* spp. and *Ascaris* spp.

Freudenthal et al. (2022) used published shotgun metagenomics to assess the entire wastewater microbiome, including eukaryotes, from ten wastewater treatment plants in Switzerland. The authors also analyzed meta-transcriptomics data from the same samples, which allowed comparison of abundance (rDNA reads) and activity (rRNA reads) of the microbial community. Raw reads were analyzed via MG-RAST, and eukaryotic taxa (protists, fungi, and microscopic metazoa), were identified by searching the PR^2 ribosomal database, with BLASTn used to filter the results using an e value of 1e-^{50} and a similarity threshold of $\geq$80 %. The results obtained showed a surprising diversity and abundance of unicellular parasites, particularly in the inflow of the treatment plants. Taxonomic identification was limited at the genus level, and the parasites *Acanthamoeba*, *Ascaris*, *Blastocystis*, *Cryptosporidium*, *Dientamoeba*, *Entamoeba*, *Giardia*, *Leishmania*, *Tetratrichomonas* and *Trichomonas* were identified.

Ariyadasa et al. (2023) employed shotgun metagenomics to analyze the viral, archaeal, and eukaryotic microflora in wastewater samples taken throughout a treatment plant (raw influent, effluent, oxidation pond water, and oxidation pond sediment) in Aotearoa (New Zealand). Water samples were concentrated by filtration and DNA extracted using the PowerSoil Pro DNA extraction kit (Qiagen, Germany). Libraries were sequenced on the Illumina NovaSeq6000 platform, using 2 $\times$ 150-bp paired-end sequencing, to generate between 6,9 $\times$ 10^5 and 1,5 $\times$ 1 reads (Table 1). Raw reads were assigned taxonomically by Kraken v2.0.7 using the NCBI RefSeq database (including human). Assignments were further analyzed using phylosEq. (v1.3.4) and vegan (v2.5–7). The number of reads assigned to protozoa ranged from 163 to 9871. A particularly high abundance of Apicomplexan parasites was reported, including various species of *Plasmodium* (*P. vivax, P. cynomolgi, P. knowlesi, P. relictum,* and *P. yoelii*), *Babesia* (*B. bovis* and *B. bigemina*), *Theileria* (*T. annulata* and *T. parva*), as well as *Toxoplasma gondii* and *Cryptosporidium parvum*. The intestinal parasite *Giardia*, and organisms belonging to the phyla Cercozoa, Evosea, Euglenozoa, Microsporidia, and Fornicata, were also detected.

Gündoğdu et al. (2023) used shotgun metagenomics to assess the presence of protozoa in wastewater samples from an urban sewage treatment plant serving 1 $\times$ 10^6 individuals, a rural sewage treatment plant serving 2 $\times$ 10^4 individuals, and a Hospital in Kayseri, Turkey. Water samples were concentrated by filtration and DNA extracted using the Power-Soil DNA Isolation Kit (Qiagen). Libraries were sequenced on the Illumina NextSeq 500 platform to generate between 9,4 and 28,5 $\times$ 10^7 reads (Table 1). Raw reads from each sample were aligned against a database composed of 80 human parasite genomes retrieved from the National Center for Biotechnology Information (NCBI) GenBank database. Remarkably, 73 and 75 of the 80 protozoan species were detected in the inlet and outlet of the urban sewage treatment plant, respectively, while 65 species were detected in both the inlet and outlet of the rural sewage treatment plant, and 61 species in the outlet from the Hospital. The parasites identified included various subtypes of *Blastocystis* spp., several species of *Leishmania* (*L. donovani*, *L. enriettii*, *L. gerbilli,* and *L. panamensis*), *Plasmodium* (*P. falciaparum* and *P. ovale*) and *Cryptosporidium* (*C. parvum*, *C. hominis*, *C. meleagridis*, *C. muris* and *C. andersoni*). Notably, *Enterocytozoon bieneusi*, *Toxoplasma gondii*, *Giardia intestinalis* and *Cyclospora cayetanensis* were also detected. Therefore, a very large diversity of vector-borne and oral-fecal transmitted parasites characterized the wastewater samples from Turkey.

Mthethwa-Hlongwa et al. (2024) used a combination of 18S rRNA short amplicon and shotgun metagenomics sequencing approaches to profile protozoan microbial communities in treated and untreated wastewater samples collected from different regions of South Africa. Water samples were concentrated by filtration and DNA extracted using the phenol-chloroform method. Libraries were sequenced on the Illumina NexSeq platform to generate between 0,71 and 0,93 $\times$ 10^7 reads (Table 1). Raw reads from shotgun metagenomics (performed on a subset of untreated wastewater samples) were classified taxonomically using Kraken 2 and a custom database for Eukaryotes. On average, about 1 % of the reads was assigned to Eukaryotes, and parasites belonging to the phyla Apicomplexa, Euglenozoa, Evosea, Discosea, Cercozoa, Parabasalia, and Fornicata were identified. At the species level, many parasites were identified, including *Cryptosporidium parvum*, *C. hominis*, *C. muris*, *Eimeria acervulina*, *Entamoeba histolytica, E. nuttalli, Giardia intestinalis*, and *Toxoplasma gondii*. Therefore, a high diversity of oral-fecal transmitted parasites characterized the untreated wastewater samples from South Africa.

Moumen et al. (2024) used shotgun metagenomics to assess the microbial diversity of gutter water from the city of Pointe-Noire, Republic of Congo. Water samples were concentrated by filtration and DNA was extracted using the DNeasy PowerWater kit. Libraries were sequenced on an Illumina HiSeq 4000 platform, using 2 $\times$ 150-bp paired-end sequencing, to generate 8,4 $\times$ 10^7 reads per sample (Table 1). Raw reads were classified taxonomically using Kaiju and three associated databases, including nr-euk, a non-redundant database that includes fungi and microbial eukaryotes, the Reference Viral Database, rvdb, and a plasmid database. A very large diversity of vector-borne and oral-fecal transmitted parasites was found, including 17 species of *Plasmodium*, 9 species of

Cryptosporidium, 8 species of *Eimeria*, along with *Babesia microti* and *B. bovis*, *Theleria euqi* and *T. parva*, *Toxoplasma gondii*, *Besnoitia besnoiti*, *Neospora caninum*, *Cyclospora cayetanensis*, *Giardia intestinalis* and *Trichomonas vaginalis*.

4. Discussion

Although the number of studies that specifically focus on the detection of parasites in food and water matrices using shotgun metagenomics is still limited (Mthethwa et al., 2021), we believe that a critical assessment of the methods used and of the results obtained, will be informative.

As for other methodologies, sensitivity and specificity are the main parameters also in the context of shotgun metagenomics. In the case of Eukaryotes, and of parasites, the issue of sensitivity is very important, as only a very small fraction of sequencing reads in shotgun experiments originate from these organisms. Indeed, in the literature we reviewed, the fraction assigned to Eukaryotes was always <1 % (Table 1), and, of course, that potentially originating from parasites is even smaller. On the other hand, >99 % of the (classifiable) reads obtained from the samples analyzed was of bacterial origin. This means that short reads of bacterial origin can be assembled into longer contiguous sequences (contigs) or even into draft genomes (or MAGs, Metagenome Assembled Genomes), which facilitate downstream analyses leading to the identification of bacteria at the species or strain level (Lindner et al., 2024). Arguably, this approach is barely applicable to eukaryotic pathogens.

The second issue is specificity, which is crucial in our opinion. All analytical pipelines that process metagenomics shotgun data by assigning the reads need a reference (or, at least, a draft) genome to perform this task (Fig. 1, Panel A). Not surprisingly, the number of bacterial genomes (2,537,464 genomes, of which 20,677 are reference genomes, as per February 2025) is vastly larger than that of eukaryotes (49,000 genomes, of which 17,813 are reference genomes) and of parasites (about 1500 genomes, of which 200 are reference genomes). Moreover, bacterial genomes are smaller in size than eukaryotic genomes, and therefore have a reduced complexity (e.g., repetitive sequences and gene families are rare). Consequently, detection and taxonomic classification of prokaryotes from shotgun metagenomics is less challenging than for eukaryotes (and parasites).

Nevertheless, shotgun metagenomics data from surface and wastewater samples collected from different environments and geographical origin (Table 1) revealed the presence of many parasites, including insect- and tick-borne transmitted species (notably *Plasmodium* species, but also *Theileria*, *Babesia* and *Leishmania*) as well as species transmitted via the oral-fecal route (notably *Cryptosporidium*, *Toxoplasma, Entamoeba* and *Giardia*). These findings were consistent with data from epidemiological surveys conducted in the same areas, which showed a high prevalence of parasitic infections in the human population. However, other results, such as the detection of *Leishmania amazonensis* in water samples in Turkey (Gündoğdu et al., 2023) and of *Plasmodium falciparum* and *P. vivax* in water samples in New Zealand (Ariyadasa et al., 2023), were more difficult to reconcile with known epidemiologic data, since these parasites were never detected in humans in the areas of investigation.

Without questioning the validity and interest of the results, we believe that the studies we reviewed overlooked two aspects. The

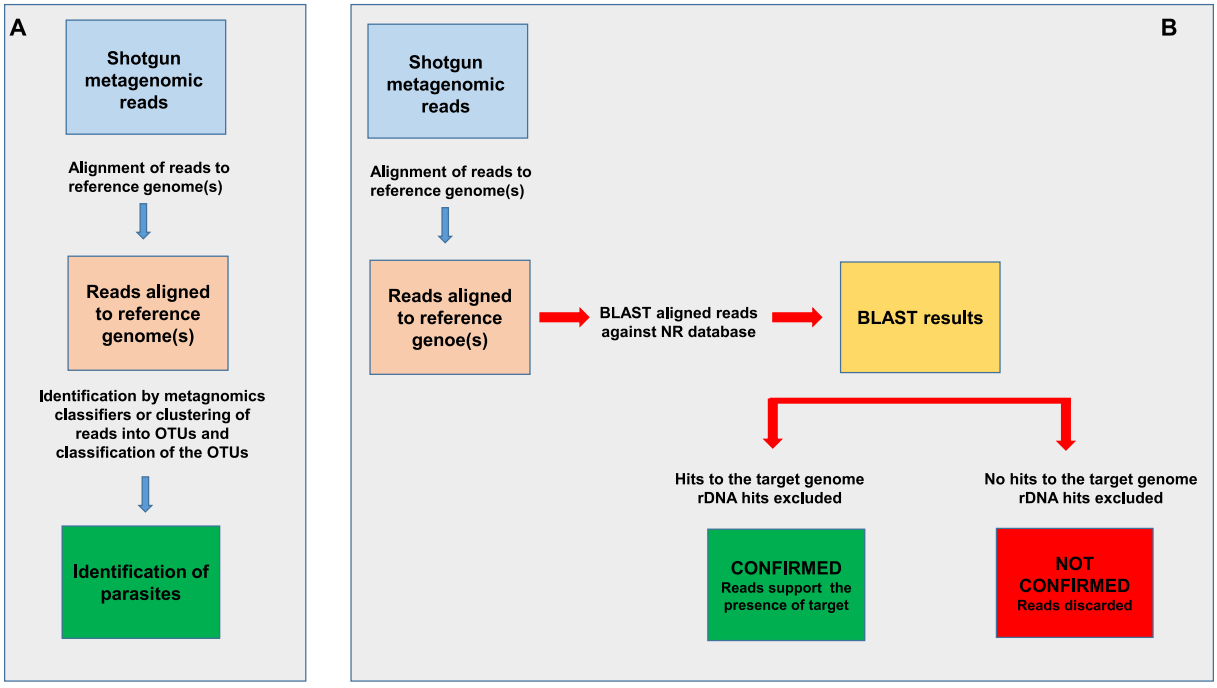


Fig. 1. Panel A: schematic representation of a typical workflow for identification of parasites from shotgun metagenomics data. Panel B: schematic representation of the proposed integrations to the workflow.

first is the possible bias arising from the taxonomical assignment of short reads representing fragments of the eukaryotic ribosomal DNA (18S, 5S, and 28S) genes. It has been shown that an accurate detection of parasites from complex samples is achievable by sequencing full-length rRNA molecules, an approach for which the term ribosomalomics has been proposed (Wylezich and Höper, 2021). On the other hand, short ribosomal fragments may exhibit high sequence similarity between different species, increasing the risk of spurious matches and therefore wrong assignments. The issue of false positives has been discussed in the context of bacterial metagenomics research (Bradford et al., 2024), also as a possible source of errors in estimating the relative abundance due to the recruitment of reads from non-target genomes (Castro et al., 2018).

The second aspect is the lack of a confirmatory step where the reads assigned to one or more parasites are subjected to a BLAST search against the non-redundant GenBank database, in order to reassess their identity with an associated statistical value (*p*-value). As the number of reads assigned to the targeted parasites that should undergo a BLAST search is relatively limited (tens to few hundreds), and the analysis can be automated and executed on a local computer, we advocate the use of this strategy (Fig. 1, panel B).

In conclusion, while the application of shotgun metagenomics for detection of parasites in food and water matrices generates important data, there are still challenges to be considered and possibilities to achieve a trustworthy taxonomic classification of sequencing reads.

CRediT authorship contribution statement

Paolo Vatta: Writing – original draft, Methodology, Investigation. **Simone M. Cacciò:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

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.

Acknowledgements

We thank Valeria Michelacci and Federica Gigliucci (Department of Food Safety, Nutrition and Veterinary Public Health, Istituto Superiore di Sanità, Italy), and Davide Sassera (IRCCS Fondazione Policlinico San Matteo, Pavia, Italy) for critical reading of the manuscript.

References

Ahlinder, J., Svedberg, A.L., Nystedt, A., Dryselius, R., Jacobsson, K., Hägglund, M., et al., 2022. Use of metagenomic microbial source tracking to investigate the source of a foodborne outbreak of cryptosporidiosis. Food Waterborne Parasitol. 26, e00142. https://doi.org/10.1016/j.fawpar.2021.e00142.

Ariyadasa, S., Taylor, W., Weaver, L., McGill, E., Billington, C., Pattis, I., 2023. Nonbacterial microflora in wastewater treatment plants: an underappreciated potential source of pathogens. Microbiol. Spectr. 11 (3), e0048123. https://doi.org/10.1128/spectrum.00481-23.

Bazant, W., Blevins, A.S., Crouch, K., Beiting, D.P., 2023. Improved eukaryotic detection compatible with large-scale automated analysis of metagenomes. Microbiome 1 (1), 72. https://doi.org/10.1186/s40168-023-01505-1.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., Thomas, W.K., 2012. Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol. Evol. 27 (4), 233–243. https://doi.org/10.1016/j.tree.2011.11.010.

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., et al., 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat. Biotechnol. 41 (11), 1633–1644. https://doi.org/10.1038/s41587-023-01688-w.

Bradford, L.M., Carrillo, C., Wong, A., 2024. Managing false positives during detection of pathogen sequences in shotgun metagenomics datasets. BMC Bioinformatics 25 (1), 372. https://doi.org/10.1186/s12859-024-05952-x.

Breitwieser, F.P., Lu, J., Salzberg, S.L., 2019. A review of methods and databases for metagenomic classification and assembly. Brief. Bioinform. 20 (4), 1125–1136. https://doi.org/10.1093/bib/bbx120.

Brumfield, K.D., Hasan, N.A., Leddy, M.B., Cotruvo, J.A., Rashed, S.M., Colwell, R.R., et al., 2020. A comparative analysis of drinking water employing metagenomics. PLoS One 15 (4), e0231210. https://doi.org/10.1371/journal.pone.0231210.

Castro, J.C., Rodriguez-R, L.M., Harvey, W.T., Weigand, M.R., Hatt, J.K., Carter, M.Q., et al., 2018. imGLAD: accurate detection and quantification of target organisms in metagenomes. Peer J. 6, e5882. https://doi.org/10.7717/peerj.5882.

da Silva, T.C.B., Chaúque, B.J.M., Benitez, G.B., Rott, M.B., 2024. Global prevalence of potentially pathogenic free-living amoebae in sewage and sewage-related environments-systematic review with meta-analysis. Parasitol. Res. 123 (3), 148. https://doi.org/10.1007/s00436-024-08164-7.

Farkas, K., Williams, R.C., Hillary, L.S., Garcia-Delgado, A., Jameson, E., Kevill, J.L., et al., 2024. Harnessing the power of next-generation sequencing in wastewater-based epidemiology and global disease surveillance. Food Environ. Virol. 17 (1), 5. https://doi.org/10.1007/s12560-024-09616-0.

Freudenthal, J., Ju, F., Bürgmann, H., Dumack, K., 2022. Microeukaryotic gut parasites in wastewater treatment plants: diversity, activity, and removal. Microbiome 10 (1), 27. https://doi.org/10.1186/s40168-022-01225-y.

Gündoğdu, A., Charyyeva, A., Nalbantoğlu, Ö.U., 2023. Metagenomic profiling of human protozoan parasites in wastewater and hospital effluents. J. Clin. Pract. Res. 45 (5), 435–446. https://doi.org/10.14744/cpr.2023.10820.

Hamner, S., Brown, B.L., Hasan, N.A., Franklin, M.J., Doyle, J., Eggers, M.J., et al., 2019. Metagenomic profiling of microbial pathogens in the little Bighorn River, Montana. Int. J. Environ. Res. Public Health 16 (7), 1097. https://doi.org/10.3390/ijerph16071097.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5 (10), R245–R249. https://doi.org/10.1016/s1074-5521(98)90108-9.

Hendriksen, R.S., Lukjancenko, O., Munk, P., Hjelmsø, M.H., Veranim, J.R., Ng'eno, E., et al., 2019. Pathogen surveillance in the informal settlement, Kibera, Kenya, using a metagenomics approach. PLoS One 14 (10), e0222531. https://doi.org/10.1371/journal.pone.0222531.

Lind, A.L., Pollard, K.S., 2021. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. Microbiome 9 (1), 58. https://doi.org/10.1186/s40168-021-01015-y.

Lindner, B.G., Gerhardt, K., Feistel, D.J., Rodriguez-R, L.M., Hatt, J.K., Konstantinidis, K.T., 2024. A user's guide to the bioinformatic analysis of shotgun metagenomic sequence data for bacterial pathogen detection. Int. J. Food Microbiol. 30 (410), 110488. https://doi.org/10.1016/j.ijfoodmicro.2023.110488.

Marcelino, V.R., Clausen, P.T.L.C., Buchmann, J.P., Wille, M., Iredell, J.R., Meyer, W., et al., 2020. CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. Genome Biol. 21 (1), 103. https://doi.org/10.1186/s13059-020-02014-2.

Marchesi, J.R., Ravel, J., 2015. The vocabulary of microbiome research: a proposal. Microbiome 3, 31. https://doi.org/10.1186/s40168-015-0094-5.

Moumen, B., Samba-Louaka, C., Kimpamboudi, V.A.M., Boumba, A.M., Ngoma, H.S., Samba-Louaka, A., 2024. Metagenomic data from gutter water in the city of Pointe-Noire, Republic of Congo. Data Brief 55, 110655. https://doi.org/10.1016/j.dib.2024.110655.

Mthethwa, N.P., Amoah, I.D., Reddy, P., Bux, F., Kumari, S., 2021. A review on application of next-generation sequencing methods for profiling of protozoan parasites in water: current methodologies, challenges, and perspectives. J. Microbiol. Methods 187, 106269. https://doi.org/10.1016/j.mimet.2021.106269.

Mthethwa-Hlongwa, N.P., Amoah, I.D., Gomez, A., Davison, S., Reddy, P., Bux, F., et al., 2024. Profiling pathogenic protozoan and their functional pathways in wastewater using 18S rRNA and shotgun metagenomics. Sci. Total Environ. 20 (912), 169602. https://doi.org/10.1016/j.scitotenv.2023.169602.

Ogola, H.J.O., Ijoma, G.N., Edokpay I, J.N., 2024. Exploring the dichotomy: shotgun metagenomics reveals diversity of beneficial and pathogenic protist community in arid wetlands of northeastern South Africa. Sci. Total Environ. 946, 174306. https://doi.org/10.1016/j.scitotenv.2024.174306.

Roy, M.A., Arnaud, J.M., Jasmin, P.M., Hamner, S., Hasan, N.A., Colwell, R.R., et al., 2018. A metagenomic approach to evaluating surface water quality in Haiti. Int. J. Environ. Res. Public Health 15 (10), 2211. https://doi.org/10.3390/ijerph15102211.

Sadurski, J., Polak-Berecka, M., Staniszewski, A., Waśko, A., 2024. Step-by-step metagenomics for food microbiome analysis: a detailed review. Foods 13 (14), 2216. https://doi.org/10.3390/foods13142216.

Sala, C., Mordhorst, H., Grützke, J., Brinkmann, A., Petersen, T.N., Poulsen, C., et al., 2020. Metagenomics-based proficiency test of smoked salmon spiked with a mock community. Microorganisms 8 (12), 1861. https://doi.org/10.3390/microorganisms8121861.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., et al., 2023. Next-generation sequencing technology: current trends and advancements. Biology (Basel) 12 (7), 997. https://doi.org/10.3390/biology12070997. Erratum in: Biology (Basel). 2024 Apr 24;13(5):286. doi: 10.3390/biology13050286.

Sekse, C., Holst-Jensen, A., Dobrindt, U., Johannessen, G.S., Li, W., Spilsberg, B., et al., 2017. High throughput sequencing for detection of foodborne pathogens. Front. Microbiol. 8, 2029. https://doi.org/10.3389/fmicb.2017.02029.

Wylezich, C., Höper, D., 2021. Meta-Ribosomalomics: RNA sequencing is an unbiased method for parasite detection of different sample types. Front. Microbiol. 12, 614553. https://doi.org/10.3389/fmicb.2021.614553.