Use of metagenomic microbial source tracking to investigate the source of a foodborne outbreak of cryptosporidiosis

J. Ahlinder [a,*], A.-L. Svedberg [b], A. Nystedt [b], R. Dryselius [c], K. Jacobsson [c], M. Hägglund [d], B. Brindefalk [a], M. Forsman [a], J. Ottoson [e], K. Troell [f,g]

[a] CBRN Defence and Security, Swedish Defence Research Agency, FOI, Umeå, Sweden
[b] Department of Infection control, Region Norrbotten, Luleå, Sweden
[c] Department of Biology, National Food Agency, Uppsala, Sweden
[d] Department of Microbiology, Tumor and Cell Biology, Clinical Genomics Facility, Stockholm, Sweden
[e] Department of Risk and benefit assessment, National Food Agency, Uppsala, Sweden
[f] Department of Microbiology, National Veterinary Institute, Uppsala, Sweden
[g] Department of Medical Biochemistry and Microbiology, Uppsala university, Uppsala, Sweden

ARTICLE INFO

ABSTRACT

Cryptosporidium is a protozoan parasite of global public health importance that causes gastro-enteritis in a variety of vertebrate hosts, with many human outbreaks reported yearly, often from ingestion of contaminated water or food. Despite the major public health implications, little is typically known about sources of contamination of disease outbreaks caused by *Cryptosporidium.* Here, we study a national foodborne outbreak resulted from infection with *Cryptosporidium parvum* via romaine lettuce, with the main goal to trace the source of the parasite. To do so, we combined traditional outbreak investigation methods with molecular detection and characterization methods (i.e. PCR based typing, amplicon and shotgun sequencing) of romaine lettuce samples collected at the same farm from which the contaminated food was produced. Using 18S rRNA typing, we detected *C. parvum* in two out of three lettuce samples, which was supported by detections in the metagenome analysis. Microbial source tracking analysis of the lettuce samples suggested sewage water as a likely source of the contamination, albeit with some uncertainty. In addition, the high degree of overlap in bacterial species content with a public human gut microbial database corroborated the source tracking results. The combination of traditional and molecular based methods applied here is a promising tool for future source tracking investigations of food- and waterborne outbreaks of *Cryptosporidium* spp. and can help to control and mitigate contamination risks.

1. Introduction

Cryptosporidiosis is a diarrhoeal disease caused by protozoa of the genus *Cryptosporidium*. Human infection is predominantly caused by the species *C. hominis* and *C. parvum*. While *C. hominis* infection affects almost only humans, *C. parvum* can infect a wider range of animals. The transmission route is faecal-oral by direct contact with infected persons or animals, or indirectly by ingesting contaminated water or food (Ryan et al., 2018). The most common symptom is watery diarrhoea with sudden onset but abdominal

---

* Corresponding author.
  E-mail address: jon.ahlinder@foi.se (J. Ahlinder).

pain, fever, nausea, dehydration, and weight loss also occur. The incubation time can be up to 14 days with a mean of 7 days (Chalmers and Davies, 2010). Cryptosporidiosis may last longer than gastroenteritis caused by many other agents but is usually self-limited in immunocompetent people (Cacciò and Chalmers, 2016). However, for immunocompromised people, diarrhoea can be severe and even life-threatening, since treatment options are limited (Chalmers and Davies, 2010). Despite the relatively severe and unpleasant diarrhoea, cryptosporidiosis is grossly underreported; due to lack of accurate laboratory diagnosis strategies and surveillance (Cacciò and Chalmers, 2016; Harvala et al., 2016).

Globally, drinking water or recreational water associated cryptosporidiosis outbreaks are more frequently reported than foodborne outbreaks (Gharpure et al., 2009; McKerr et al., 2018; Robertson et al., 2013). The Nordic European countries seem to be over-represented in terms of foodborne outbreaks with several outbreaks from fresh produce reported from Sweden, Norway, Finland and Denmark (Robertson and Chalmers, 2013). Since oocysts are sensitive to desiccation, the cold damp climate could be an explaining factor. However, Robertson and Chalmers (2013) concluded that the reason was more likely that these countries had more directed investigations and reporting in place than in most other parts of the world. Reported Swedish foodborne outbreaks have been caused mainly by *C. parvum* (and sometimes by *C. hominis*) associated with contaminated argula salad (Insulander et al., 2013), parsley (from a béarnaise sauce) (Insulander et al., 2008), unpasteurised juice with spinach, kale (National Veterinary Institute, 2020) and two un-determined, but related to food, which were linked by the *gp60* subtype (Gherasim et al., 2012). However, the source of contamination could not be determined in any of these outbreaks. The ability to trace foodborne pathogens, or faecal indicators, to their point of origin is important for the food industry, food regulatory agencies, and public health since this information can help to better un-derstand sources of contamination and take the right corrective actions to prevent transmission (Fu and Li, 2014).

Outbreak investigation strategies at the farm level are of special concern. More complete assessments of these events and more research into the biology and ecology of pathogen-produce interactions are needed to identify prevention strategies (Lynch et al., 2009). However, to trace-back foodborne outbreaks is a challenge. For example, only in 763 out of 5363 outbreaks investigated was the same causative agent identified in the food vehicle or food chain as in humans in the EU 2012 (EFSA, 2014), and similar trends have been observed more recently (EFSA, 2021). For fresh produce, microbiological confirmation is especially challenging given the short shelf life and rapid distribution and consumption, leading to loss of trace-back evidence. This is especially true for pathogens with a relatively long incubation time such as *Cryptosporidium* spp. (Lynch et al., 2009). The contamination route of the vegetable in question is another cause of concern. Pathogens can originate in soil, introduced through the addition of soil amendments like slurry, manure or biosolids. Pathogens can also be waterborne or introduced via human handling of fresh produce, as well as from ingress of farmed, wild and domestic animals into fields (European Commission, 2017).

Here, we have used a *Cryptosporidium* foodborne outbreak in Sweden, which was reported in late summer 2016, as a case study to trace the cause of contamination on the vegetable vehicle. On the 30th August the disease control unit, responsible for regional health care coordination at the County Council of Norrbotten, contacted the Authorities, e.g. Public Health Agency, Luleå municipality and the National Food Agency, after a suspected foodborne outbreak. Fifty out of 100 teachers and pre-school teachers had fallen ill with frequent diarrhoeal incidences after an event in Luleå on the 17th August. The incubation time (four to seven days) together with symptoms (watery diarrhoea) indicated cryptosporidiosis. A week after the outbreak was detected (6th September) there were three laboratory confirmed cases of cryptosporidiosis reported in the national reporting system of communicable diseases, SmiNet.

In this paper, we characterize and trace the origin of the foodborne cryptosporidiosis outbreak at the production farm. To do so, we first performed traditional outbreak investigation methods, such as case interviews, sampling of patients, and back-tracing to identify romaine lettuce produced at a specific farm as the likely source of the outbreak. Then, to complement the outbreak investigation, additional sampling of romaine lettuce and other crops at the farm were performed and sequenced, with both amplicon and shotgun sequencing techniques, in order to track the faecal source at the site and to characterize the microbial content of the romaine lettuce including detection of *Cryptosporidium*.

2. Methods

2.1. Epidemiological trace-back investigations

Among the food served for lunch for the company of teachers and pre-school teachers – chicken, romaine lettuce, feta, parmesan and spinach pie – romaine lettuce was the most likely source of *Cryptosporidium*. Hence, when two additional cases from two other cities, appeared in the investigation, they were specifically asked about meals in restaurants and consumption of lettuce. One of them had eaten a halloumi salad with romaine lettuce as one of the ingredients in Sundsvall on the 14th August. The other person had a Caesar salad in Kiruna on 18th August. The health inspectors of the respective municipality therefore put their effort in tracing the lettuce, i.e. brands, wholesalers and, in the end, producer/farm.

2.2. Laboratory investigation of human cases

Seven stool samples were analysed for the presence of *Cryptosporidium* oocyst at the regional laboratories, all were found positive. Identification of oocysts was performed using microscopy of acid-fast stained smears. Six of the samples were further analysed for the presence of additional bacterial pathogens (*Shigella*, *E. coli* (Ehec*), Salmonella*, *Yersinia* and *Campylobacter*) using cultivation. In one sample, detection of viruses (Noro-, Adeno-, and Rotaviruses) and *Clostridium difficile* toxin were tested. All test results were negative. Five samples were examined at the Public Health Agency for confirmation and molecular typing.

DNA from the human faecal samples was extracted using a QIAmp DNA mini kit (Qiagen, Hilden, Germany) according to the

manufacturer's instructions. Prior to extraction, oocyst disruption was done by beadbeating using a BulletBlender (Techtum, Sweden). *Cryptosporidium* species identification was made by investigation of the 18S rRNA gene by PCR, followed by restriction fragment length polymorphism (RFLP) and bi-directional sequencing of the PCR amplicon (Xiao et al., 2001, 1999).

To further characterize the isolates, partial sequences of the 60 kDa glycoprotein (*gp60*) gene were amplified (Alves et al., 2003) and PCR products were subsequently sequenced. Editing and analysis of sequences were done using CLC Main Workbench (Qiagen, Aarhus, Denmark, version 6.9.1) and compared with sequences in the GenBank database using the Basic Local Alignment Search Tool (BLAST; NCBI www.ncbi.nlm.nih.gov/blast/BLAST.cgi).

2.3. Environmental trace-back investigation

Prior to the cryptosporidiosis outbreak, an ongoing project at the National food agency, with aim to characterize and investigate the microbial quality of both ready-to-eat vegetables and the irrigation water for the crops on major production farms in Sweden, conducted a large scale sampling effort (Livsmedelsverket, 2019). In August 2016, sampling of both romaine lettuce, which originated from the same batch as the lettuce connected to the outbreak, and other vegetables was performed at the same farm which produced the contaminated romaine lettuce. All crops were harvested in triplicates from separate locations at the farmland. The farm is situated in Skåne in Southern Sweden and is certified under a standard for quality-assured production through third-party certification (Kvalitetssystem, 2016) using stricter guidelines than the global good agriculture practices (Global GAP, 2016). The county board of Skåne inspected the farm after the outbreak but could neither find any remarks on the production nor on the handling of the produce. The lettuce was free-land cultured and over-head irrigated using ramps. The irrigation water came from two bores (deep wells) with high water quality: *E. coli*, enterococci or coliform bacteria normally not detected in 100 ml. On one occasion during the project (8th August 2016), 1 MPN *E. coli* 100 ml^{-1} was detected in the irrigation water in one of the ramps (data not shown). According to the company manager, only chemical fertilizers were used on the fields. Despite being situated in one of the more animal dense counties in Sweden, there are neither animal farms, nor wastewater outlets in the close vicinity to the groundwater source or the fields from where the romaine lettuce was harvested. To our knowledge, no confirmed cases of cryptosporidiosis were reported locally to the farm nor among the farm workers.

2.4. Microbiological investigation of vegetables

Since the authorities were notified 13 days after the event in Luleå, the possibility to track the contaminated food for sampling was deemed minimal and no microbiological investigation on food was initiated. However, samples of romaine lettuce from the same farm and batch, had previously been analysed for faecal contamination within the irrigation water project (Livsmedelsverket, 2019). Extracted DNA and eluate from these lettuce samples was saved in the freezer which gave an opportunity to further investigate the cause of the outbreak by looking for *Cryptosporidium* specific DNA and performing faecal source tracking based on amplicon and shotgun sequencing, respectively.

The vegetables harvested in the project included three or four replicates of broccoli, iceberg lettuce, leek, cauliflower, chive, parsley and romaine lettuce collected at five farms in the south of Sweden. The romaine lettuce were sampled pre-harvest 8th August 2016, originating from the same batch as the romaine lettuce heads causing cryptosporidiosis in the case example, and sent to the

Table 1
Results from microbiological analyses of vegetables from the farm (sampled six days before the onset of the outbreak) including 16S rRNA microbial source tracking and metagenome based analyses.

Sample	Coliform bacteria (CFU/g)	*E. coli* (CFU/g)	Intestinal enterococci (CFU/g)	*Clostridium perfringens* (CFU/g)	Bacterial pathogens	*Cryptosporidium*	Source-tracking 16S	Human gut associated species
Romaine lettuce A	190	<10	< 100	10	Neg 25 g	*C. parvum*	Sewage (1.8%), Gull (0.1%), Unknown (0.5%)	239
B	270	270	< 100	< 10	Neg 25 g	*C. parvum*	Unknown (1.6%)	542
C	< 10	<10	< 100	< 10	Neg 25 g	Negative	Unknown (0.9%)	140
Cauliflower A	<10	<10	< 100	<10	Neg 25 g	ND	–	–
B	200	10	< 100	<10	Neg 25 g	ND	Unknown (0.1%)	–
C	<10	<10	< 100	<10	Neg 25 g	ND	–	–
Leek A	43,200	<10	< 100	<10	Neg 25 g	ND	–	–
B	<10	<10	< 100	<10	Neg 25 g	ND	–	–
C	20	<10	< 100	<10	Neg 25 g	ND	–	–
Lettuce A	60	<10	< 100	<10	Neg 25 g	ND	–	–
B	20	<10	< 100	<10	Neg 25 g	ND	–	–
C	30	<10	< 100	<10	Neg 25 g	ND	–	–

National Food Agency in a cooling bag. The day after harvest, 10 g from each subsample (i.e. lettuce head) was shredded and put in 90 ml of peptone buffer. This was then treated for one minute at normal speed in a Steward stomacher for analysis of bacterial indicators. *E. coli* and coliform bacteria were analysed by plate spreading on Chromocult® Coliform ES-agar (Merck; Darmstadt, Germany). Intestinal enterococci were enumerated using NMKL No. 68 (NMKL, 2011) and anaerobic sulphite reducing bacteria according to NMKL No. 56 (NMKL, 2015). Further, 25 g of each vegetable sample was analysed by standard enrichment methods for the presence of *Salmonella, Campylobacter*, shigatoxin-producing *E. coli* and ESBL-producing *E. coli*. None of these organisms were detected in any of the samples (Table 1).

2.5. DNA extraction and molecular analyses

Nucleic acids were extracted from 2 ml of the peptone buffer from the Stomacher protocol using a modification of the protocol of Meta-G-Nome DNA Isolation Kit (Epicentre Biotechnologies, Madison, Wisconsin, United States) as follows: 2 ml of eluate was centrifuged at 13,000 $\times g$ for 60 min at 4 °C; then, 1.9 ml was removed, and the pellet stored at -20 °C until extraction. The samples were resuspended in 200 μl TE-buffer and mixed by vortexing.

Cryptosporidium samples for molecular analysis were prepared using the aforementioned DNA extractions and two additional preparations. The additional preparations were made from stored (-20 °C for 10 months) aliquots of 14 ml peptone water from the stomacher preparations, two for each romaine lettuce head. The samples were used for either immunomagnetic separation, IMS, or direct extraction. The IMS was done using the Dynabeads™ GC-Combo kit (Applied Biosystems) according to manufacturer's protocol with a starting volume of 10 ml and with a final volume of 50 μl purified material. The other 14 ml sample was first centrifuged 2700 $\times g$ for 10 min, the supernatant was removed to 1 ml and the sample was transferred to a microcentrifuge tube and centrifuged at full speed for 3 min. The supernatant was removed, and the remaining sample was extracted with PowerSoil®DNA Isolation Kit (Mo Bio Laboratories Inc.) according to the manufacturer's protocol and eluted in 80 μl water.

2.6. Cryptosporidium analyses

Molecular detection and identification of *Cryptosporidium* species was done by amplification of the 18S rRNA gene by PCR, followed by bi-directional sequencing of the PCR amplicons. A nested PCR protocol for partial amplification of the 18S rRNA gene was set up with all three DNA preparations as template. The reaction was set up using primers by Xiao et al. (1999) and reaction conditions and settings as described in Björkman et al. (2018). Two μl of the first reaction was used as template for the second reaction using the same reaction conditions except that annealing temperature was raised to 63 °C. Subtyping was performed through partial amplification of the *gp60* gene using the same PCR mixture as for 18S but with primers described in Alves et al. (2003) (Table 1). The *gp60* annealing temperature was 52 °C in the first PCR and 55 °C in the second. Achieved sequences were compared with sequences deposited in GenBank using Basic Local Alignment Search Tool (BLAST, NCBI, http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

2.7. Microbial source tracking: Amplicon sequencing and shotgun sequencing

DNA was isolated using the Meta-G-Nome DNA Isolation Kit (Epicentre Biotechnologies, Madison, Wyoming, United States) and amplified using the "No. 5 Hot Mastermix 2.5x kit" (5 PRIME, Hilden, Germany) with bacteria/archaeal primers 515F/806R, specific for the V4 region of the 16S rRNA gene (Caporaso et al., 2012). The forward and reverse primers were modified to include a 12-nucleotide Golay barcode. All samples were amplified in triplets and pooled after PCR amplification (94 °C for 3 min, 35 cycles; 94 °C for 45 s; 50 °C for 1 min; 72 °C for 1.5 min, and 10 min to finish). The PCR product was run on a 1% agarose gel and the DNA concentration was estimated with a Qubit fluorometer (Invitrogen). The amplicons were pooled at equimolar concentrations and purified with the "Ultra Clean PCR Clean-Up kit" (MoBio; Carlsbad, CA, USA) following the supplier's instructions. Finally, the DNA concentration of the pooled amplicon product was measured with a Qubit fluorometer and adjusted to 2 nM. The library was denaturised and diluted as described by Illumina (MiSeq System User Guide, Part # 15027617 Rev. C), before it was loaded into a MiSeq (Illumina) using 500 bp paired-end reagent kits.

To process the raw sequence data, the dada2 pipeline was used. Samples were demultiplexed by using deML with default settings (Renaud et al., 2015). Read length truncation parameters were decided based on Phred quality scores plot for each sequencing run and varied between 140 and 180 for forward reads and between 120 and 135 for reverse reads. The obtained sequence reads were then assigned a taxonomic group by using the reference database Silva, version 138, based on 99% sequence identity (Quast et al., 2013). Low abundance amplicon sequence variants (ASVs) were filtered out from the data, following recommendations by Bokulich et al. (2013) on Illumina-generated amplicon data: ASVs with a number of sequences less than 0.001% of the total number of sequences were discarded with a minimum of detection in three samples required for them to be retained. To carry out an microbial source tracking (MST) analysis, SourceTracker version 1.0.0 (Knights et al., 2011) was utilized. In short, the MST analysis is based on defining bacterial fingerprints for known source environment, and compare those to the bacterial profile of a possibly contaminated sample (i.e. a sink sample). The rarefying depth of the MST library was set to 10,000 to reduce the required computational time. However, sink or test samples were not rarefied to maximize the taxa resolution as the goal was to find low abundance ASVs of possible faecal origin. Default values of all other parameters in SourceTracker were used. The MST library was setup as described in Hägglund et al. (2018), with some modifications: the wild bird faecal environment was partitioned into waterfowl ($n = 15$) and gull ($n = 13$) sources, and 17 crop samples (four broccoli, two cauliflower, four chive, three iceberg lettuce, three leek, and one parsley, with negative detection on *E. coli*, enterococci and coliforms) as background environment of the contaminated romaine lettuce samples. By defining the background

community composition, the natural variation was effectively incorporated in the MST library to avoid inflation of false-positive source detection. In addition, 12 negative control samples (i.e. no-template PCR) were added to account for possible lab contamination, following Pullerits et al. (2020). One cauliflower sample (replicate B) was included in the MST analyses to check for systematic contamination at the sampling site.

To further increase the resolution, shotgun sequencing of the three romaine lettuce samples was performed. Sequencing libraries were prepared using a Kapa HyperPlus Single-Index Adapter Kit with TruSeq adapters (Illumina, San Diego, CA, United States), with sample-specific multiplex adaptor according to the manufacturer's instructions. The libraries were sequenced using a single lane in a HiSeq2500 System (Illumina) at the SciLifeLab National Genomics Infrastructure Platform at KTH, Stockholm. Raw data were filtered (i.e. removing adapters and low quality sequences) using Trimmomatic (Bolger et al., 2014). Reads were taxonomically classified using.

Centrifuge (Kim et al., 2016) based on the NCBI nucleotide (nt) database (Agarwala et al., 2018). The taxonomic composition of the romaine lettuce samples were visualized using the phyloseq R package (McMurdie and Holmes, 2013). Trimmed reads were mapped against the reference genome of *C. parvum* Iowa II strain (accession number NC_006980) using BWA (Li and Durbin, 2009). To create human gut associated bacterial signatures, the taxonomic composition of the romaine lettuce samples, at the species level, was compared to the GMrepo database containing annotated human gut metagenomes (Wu et al., 2020). Low abundant species with less than 500 reads assigned were removed. The Venn diagram of the human gut associated species present in the samples was made using the ggVenn R package.

3. Results

3.1. Epidemiological investigation

In total, there were 50 estimated cases from three cities falling ill between 23rd and 25th of August 2016. The incubation times were between four and nine days. Based on the interviews and information from the health inspectors' visits on the respective restaurants, romaine lettuce produced by a company in Southern Sweden was considered to cause the outbreak. Since the evidence was considered as convincing, no further epidemiological trace-back, e.g. case-control study was performed.
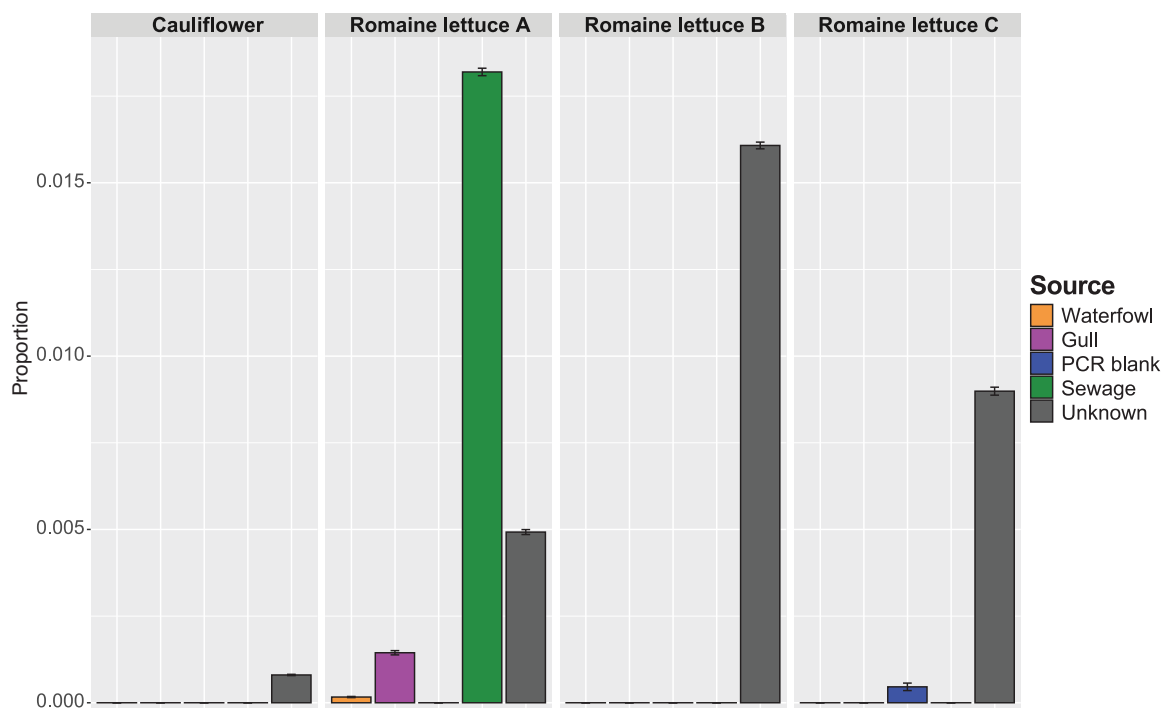

Fig. 1. The inferred source proportions of reads originating from faecal and unknown source environments of the three romaine lettuce samples and the cauliflower control sample, where standard deviations of the estimate are shown on top of the bars. The colour of the bars indicate the source of the signal. The sources calf, cow, dog, poultry, horse, swine and sheep were included in the microbial source tracking library but excluded from the figure as not represented in the samples.

3.2. Microbiological investigations

3.2.1. Analysis and typing of Cryptosporidium

Analyses of clinical samples by microscopy, performed at the local public health laboratory, confirmed seven positive cases. Five samples were analysed for species determination and subtyping. All five samples showed 100% identity to *C. parvum* 18S (GenBank accession number KM012040.1.) and gp60 subtype IIdA24G1 (GenBank accession number JQ028865).

Triplicate DNA preparations were performed from three romaine lettuce heads obtained from the aforementioned producer. The 18S nested PCR generated positive results on two lettuce heads, in two out of three triplicates, respectively. No sample was positive after IMS. The sequenced 18S amplicons showed 100% identity to *C. parvum*.

E. coli was detected in two of the three romaine lettuce samples in numbers corresponding to 270 and 10 CFU/ g respectively, sulphite-reducing anaerobes in two (10 CFU/ g) whereas intestinal enterococci were not detected (detection limit 100 CFU/ g). In the other vegetables, *E. coli* was detected (10 CFU/ g) in one cauliflower sample but apart from that no specific faecal indicator was found (Table 1).

Several attempts were made to get a positive *gp60* PCR but no sample was positive in any of the preparations.

3.2.2. Microbial source tracking

To infer the source origin of the contaminated romaine lettuce, collected samples (i.e. sink and background samples) were first sequenced by using amplicon sequencing of the 16S rRNA gene, V4 hypervariable region, to obtain the bacterial community composition or fingerprints (Fig. S1). To aid the MST library, both waterfowl faeces (i.e. mallard and goose) and milli-Q reference water (i.e. PCR blanks) were collected and sequenced. One cauliflower sample, collected at the same vegetable producer farm, was included together with the three romaine lettuce samples as sink samples in the MST analysis, in order to check for joint faecal signals. Other faecal environments in the MST library represented cow, calf, dog, horse, poultry, swine, sheep, and untreated sewage (Fig. S2). A sewage contamination was inferred in the romaine lettuce sample replicate A, with a source proportion of 1.8% of the total sequence (Fig. 1, Table 1). A weak gull bird faecal signal was detected in the same sample totalling 0.1% of the reads, with standard error interval did not include zero proportion. The source origin for some proportion of the data could not be resolved (to the pre-defined source environments in the faecal library) and were classified to the "Unknown" source in all three romaine lettuce samples and was most prominent in the A sample. The main component of the inferred source proportion was the background environment, ranging between 96 and 99.9% of the data, suggesting that the background variation was properly taken care of, which reduces the risk of false positive signals. Inadvertent contamination, via no- template PCR control, were detected in the assemblages. In total, 18 ASVs underpinned the inferred sewage signal in the romaine lettuce sample A, with taxonomic assignments to bacterial families Planococcaceae (Firmicutes), Coxiellaceae (Proteobacteria), Sphingobacteriaceae (Bacteroidetes) and Cytophagaceae (Bacteroidetes) which have been detected in faecal environments. Also, four ASVs assigned to bacterial family Moraxellaceae, which has previously been classified to sewer infrastructure-associated environments (Newton et al., 2013). The signal to the unknown environment consisted of $n = 11$, $n = 53$, and $n = 41$ ASVs for romaine lettuce samples A, B and C respectively. For sample B and C, the unknown signal contained families which was identified in the sewage signal for sample A, which suggests a similar source but less accurate in terms of which ASVs that were detected within families. The bacterial families that did overlap between unknown signals in samples B and C with the sewage signal in sample A was related to faecal environments including Moraxellaceae, Planococcaceae, Coxiellaceae and Sphingobacteriaceae. Interestingly, the family Paenibacillaceae (Firmicutes) was also present in the unknown signal which has been isolated in various faecal environments. The result of the MST analysis on the cauliflower sample did not show any faecal contamination and almost no sequences assigned to the unknown environment.
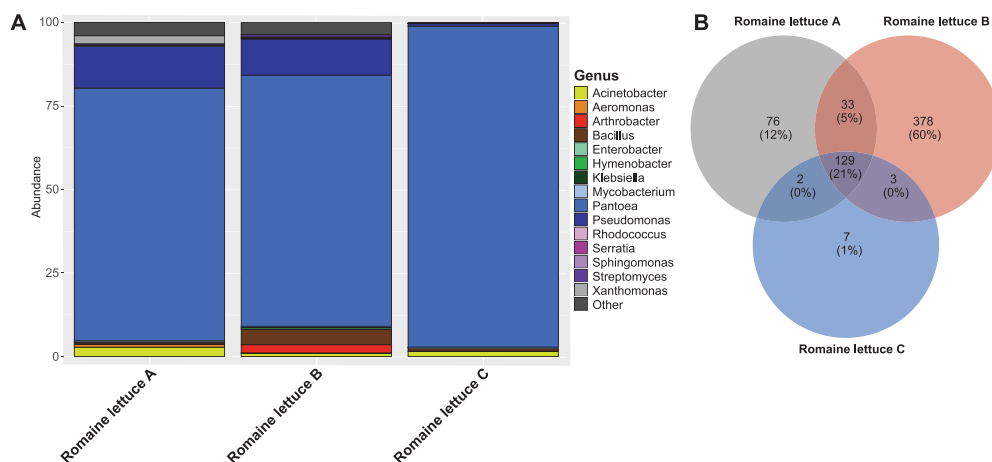

Fig. 2. A, Taxonomic composition at the genus level of the human faecal signal in the romaine lettuce samples where the top 20 abundant genus are shown. **B**, distribution of species in the samples associated with the human faeces database.

To further decipher the faecal contamination of the romaine lettuce samples, shotgun sequencing was carried out and the resulting species' assignments were compared to a human gut metagenome database GMrepo. The composition of human gut related bacteria differed in the samples with sample B having a relatively large proportion of the genus *Bacillus*, which was not as abundant in samples A and C (Fig. 2A). Interestingly, a large number of species did overlap with about 20–30% of the total detected species in the samples (Fig. 2B). A particularly large number of species overlapped in sample B, with 60% of all human gut-associated species in the lettuce samples were unique to this assemblage. In the set of overlapping species, there were some which are highly associated with human and animal faecal environments: *Escherichia coli, Enterococcus faecium, Enterococcus faecalis, Clostridium* spp. (10 species including *Clostridium perfringens*). Also, reads taxonomically classified to potential pathogenic species were found such as *Legionella pneumophila* (sample A and B), *Mycobacterium tuberculosis* (in all samples), *Campylobacter jejuni* (in all samples), and *Coxiella burnetii* (sample A).

Several *Cryptosporidium* species were detected in the taxonomic classification analysis, with *C. parvum* being the most abundant, albeit at a low total frequency: 4219, 307 and 249 unique reads for samples A, B and C, respectively. The other species detected were *Cryptosporidium muris* and *C. hominis*, present in all samples. We mapped all reads to the reference genome of *C. parvum* (the Iowa II strain), which resulted in a mapping frequency of 1.8%, 1.0% and 0.7% of the trimmed reads and an average coverage of all bases (i.e. average number of bases mapped at a position) in the reference of 15.5, 9.5 and 8.4 in samples A, B and C, respectively. Although no complete genomes were recovered, these results suggests that *C. parvum* were present in the collected romaine lettuce samples.

Taken together, there are indications of a human faecal contamination in the samples, both derived via metagenomics analyses from 16S rRNA amplicon and shotgun sequencing. A sewage signal was inferred in the 16S rRNA data for sample A, while sample B and C seemed to have similar set of taxa suggesting a similar origin of contamination. A large proportion of the total species found with shotgun metagenomics analysis was overlapping with a human gut microbiome database, with a particularly diverse set in sample B.

4. Discussion

By a combination of traditional outbreak investigating methods and DNA sequencing methods we pinpoint the origin of the contaminated vegetables involved in a major *Cryptosporidium* outbreak in Sweden, and in addition indicate the likely source of the contamination of the vegetables. The evidence pointing towards romaine lettuce from a specific farm was convincing, since it could further explain cases from other cities for which the batch of lettuce was the only (known) common denominator. Since there was DNA and eluate left from samples from the same crops and farm (batch) we had the opportunity to establish the link between food and patient and, identify possible faecal sources of the romaine lettuce causing the described outbreak.

We successfully determined species to *C. parvum* on romaine lettuce from the same batch as associated with the outbreak, as well as from the infected humans. *C. parvum* is the most common species to cause cryptosporidiosis in humans in Sweden (Insulander et al., 2013; Lebbad et al., 2021). In a recent study 84% (211 of 250) of all domestically acquired *Cryptosporidium* infections were caused by *C. parvum* (Lebbad et al., 2021). Humans and livestock, particularly pre-weaned calves, are considered to be main reservoir hosts of *C. parvum* in the transmission of human cryptosporidiosis (Burnet et al., 2013). Wildlife, including wild ruminants, wild boar and birds may also be carriers of *C. parvum* with potential to contaminate the environment, farmed vegetables and water (García-Presedo et al., 2013; Oliveira et al., 2017; Wells et al., 2015). Further molecular typing of the human samples identified *gp60* subtype IIdA24G1. Several Swedish foodborne outbreaks have been associated with IIdA24G1 (Gherasim et al., 2012; National Veterinary Institute, 2020). The *gp60* subtype IIdA24G1 has never been observed in Swedish animals, however, IIdA24G1c, which is highly homologous to IIdA24G1, has previously been identified in four calves from two herds in Sweden (Silverlås and Blanco-Penedo, 2013). In an outbreak among veterinary students, IIdA24G1 was one of two identified *gp60* subtypes and although the original source to those cases was never identified, the epidemiological investigation suggested those patients were infected on a cattle farm (Kinross et al., 2015).

The farm on which the romaine lettuce originated used only chemical fertilizers. Hence, no contamination is likely to come from manure. Irrigation water is taken from a deep well on site and no surface water has been used. According to Livsmedelsverket (2019), the deep well water was not contaminated at the time, but may have been later in the distribution system at the farm. No domestic animals are kept on the grounds, but wild ruminants may enter the fields and graze, and wild birds may also visit the area. Although a common carrier of *C. parvum*, faeces from cattle was likely not the source of contamination, whereas wild animals cannot be excluded.

Based on the source tracking analysis, the most likely source of the faecal pollution was human, either directly or from wastewater contamination of the irrigation water. First, we identified bacterial abundance patterns in the lettuce assemblages which matched with abundance pattern in sewage assemblages, with present bacterial families typical for human faecal environments. Second, the shotgun generated data allowed us to identify bacteria at the species level, which partly overlapped with the species content in a human gut microbiome database. Potential contamination sources of sewage at the farm, such as on-site wastewater contaminating the irrigation water source or the water used for washing, or infected people handling the lettuce, are all possible explanations of such event. However, there were uncertainty in the 16S based source tracking analysis: the proportion of reads assigned to the unknown environment were present in all three lettuce assemblages. This result suggests either that the contaminating environments defined in the source tracking library were not representative for the contamination source at the site, or that the contaminating source was not present in the library at all. Spatiotemporal effects might have altered the composition of the sources (O'Dea et al., 2019), as most library samples were collected in different regions in Sweden three years earlier (Hägglund et al., 2018) and thus might not capture the bacterial diversity of the faecal contamination. By analysing spiked mesocosms containing different source contaminations, Brown et al. (2019) found that faecal source library composition substantially influenced the MST ability to infer source contributions correctly, with unknown sources (not accounted for) and large within source variability in composition as most detrimental for prediction accuracy. We cannot rule out the possibility of missing faecal sources in the library, as both deer and wild boar potentially could have contaminated the crops (Ballesté et al., 2020; McKee et al., 2021). Another source of uncertainty is the presence of

cosmopolite taxa which overlaps with multiple source environments, and could result in erroneous predictions, in particular with the background environment of the crops at the farm and faecal source library environments (Brown et al., 2019; Hägglund et al., 2018). In the present study, however, we included multiple samples from different crops at the same site as background representative communities which should capture the variability in community composition of the crops and the contamination could thus be separated.

We utilized different data sources in the metagenome and source tracking analyses to improve the source prediction. The amplicon 16S rRNA sequencing provides us with a broad bacterial taxonomic coverage and profiles of a large set of samples in multiple environments as this technique nowadays are cheap and easily accessible for diagnostic labs. Shotgun metagenomics, on the other hand, access entire genomes of all organisms present in the sample, not only bacteria but also viruses and eukaryotes. Furthermore, it's less susceptible to the biases that are inherent in targeted gene amplification of the 16S rRNA, by avoiding the PCR step(s) and primer limitations in capturing the diversity of the sample (see e.g. Berry et al. (2011); Cai et al. (2013); Tremblay et al. (2015) for sources and quantification of bias in 16S rRNA gene amplification). Even though this sequencing technique becomes cheaper and more accessible, it is still difficult to sequence a large set of samples, both with respect to financial and data resources, required in microbial source tracking investigations as the variability in community composition of library and sink environments must be correctly assessed by the investigators, as in any supervised learning problem. Thus, a reasonable way forward might be to utilize both types of sequencing techniques: 16S rRNA to access the variability in composition in a large set of samples to perform signature based source tracking (Hägglund et al., 2018; Henry et al., 2016; Roguet et al., 2018) and shotgun metagenomics to investigate a few sink communities much more in detail (i.e. broader taxonomic coverage and finer phylogenetic resolution) to detect specific organisms of interest such as *Cryposporidium* spp. (Lokmer et al., 2019; Wylezich et al., 2019) and faecal and gut associated species (Almeida et al., 2019; Walker et al., 2014; Yatsunenko et al., 2012).

Subtyping, or whole genome sequencing for bacteria, has improved outbreak investigations and is a valuable tool in order to detect an outbreak as well as to connect it to a specific food (Didelot et al., 2012). However, we do believe that microbial source tracking can be an important complement to typing, epidemiological and environmental investigations in a root cause analysis (Kase et al., 2017); especially in order to come as close to the upstream source as possible in order to suggest interventions or risk mitigation options.

Data availability

Supplementary data related to this article can be accessed at the Short Read Archive (https://www.ncbi.nlm.nih.gov/sra), project reference PRJNA768652. The obtained gp60 sequence has accession number OK317684.

Declaration of Competing Interest

The authors report no conflict of interest.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fawpar.2021.e00142.

References

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L.Y., Gorelenkov, V., Graeff, A., Hlavina, W., Holmes, B., Johnson, M., Kattman, B., Khotomlianski, V., Kimchi, A., Kimelman, M., Kimura, M., Kitts, P., Klimke, W., Kotliarov, A., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lathrop, S., Lee, J.M., Leubsdorf, C., Lu, Z., Madden, T.L., Marchler-Bauer, A., Malheiro, A., Meric, P., Karsch-Mizrachi, I., Mnev, A., Murphy, T., Orris, R., Ostell, J., O'Sullivan, C., Palanigobu, V., Panchenko, A.R., Phan, L., Pierov, B., Pruitt, K.D., Rodarmer, K., Sayers, E.W., Schneider, V., Schoch, C.L., Schuler, G.D., Sherry, S.T., Siyan, K., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Thibaud-Nissen, F., Todorov, K., Trawick, B.W., Vakatov, D., Ward, M., Yaschenko, E., Zasypkin, A., Zbicz, K., 2018. Database resources of the National Center for biotechnology information. Nucleic Acids Res. 46, D8–D13. https://doi.org/10.1093/nar/gkx1095.

Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., Finn, R.D., 2019. A new genomic blueprint of the human gut microbiota. Nature 568, 499–504. https://doi.org/10.1038/s41586-019-0965-1.

Alves, M., Xiao, L., Sulaiman, I., Lal, A.A., Matos, O., Antunes, F., 2003. Subgenotype analysis of Cryptosporidium isolates from humans, cattle, and zoo ruminants in Portugal. J. Clin. Microbiol. 41, 2744–2747. https://doi.org/10.1128/JCM.41.6.2744-2747.2003.

Ballesté, E., Demeter, K., Masterson, B., Timoneda, N., Sala-Comorera, L., Meijer, W.G., 2020. Implementation and integration of microbial source tracking in a river watershed monitoring plan. Sci. Total Environ. 736 https://doi.org/10.1016/j.scitotenv.2020.139573.

Berry, D., Mahfoudh, K. Ben, Wagner, M., Loy, A., 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl. Environ. Microbiol. 77, 7846–7849. https://doi.org/10.1128/AEM.05220-11.

Björkman, C., von Brömssen, C., Troell, K., Svensson, C., 2018. Disinfection with hydrated lime may help manage cryptosporidiosis in calves. Vet. Parasitol. 264, 58–63. https://doi.org/10.1016/j.vetpar.2018.11.004.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods 10, 57–59. https://doi.org/10.1038/nmeth.2276.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, C.M., Mathai, P.P., Loesekann, T., Staley, C., Sadowsky, M.J., 2019. Influence of library composition on sourcetracker predictions for community-based microbial source tracking. Environ. Sci. Technol. 53 https://doi.org/10.1021/acs.est.8b04707.

Burnet, J.B., Ogorzaly, L., Tissier, A., Penny, C., Cauchie, H.M., 2013. Novel quantitative TaqMan real-time PCR assays for detection of Cryptosporidium at the genus level and genotyping of major human and cattle-infecting species. J. Appl. Microbiol. 114, 1211–1222. https://doi.org/10.1111/jam.12103.

Cacciò, S.M., Chalmers, R.M., 2016. Human cryptosporidiosis in Europe. Clin. Microbiol. Infect. 22, 471–480. https://doi.org/10.1016/j.cmi.2016.04.021.

Cai, L., Ye, L., Tong, A.H.Y., Lok, S., Zhang, T., 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One 8. https://doi.org/10.1371/journal.pone.0053649.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8.

Chalmers, R.M., Davies, A.P., 2010. Minireview: clinical cryptosporidiosis. Exp. Parasitol. 124, 138–146. https://doi.org/10.1016/j.exppara.2009.02.003.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E.A., Crook, D.W., 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. 13, 601–612. https://doi.org/10.1038/nrg3226.

EFSA, 2014. The European Union summary report on trends and sources of Zoonoses, zoonotic agents and food-borne outbreaks in 2012. EFSA J. 12 https://doi.org/10.2903/j.efsa.2014.3547.

EFSA, 2021. The European Union one health 2020 Zoonoses report. EFSA J. 19, 1–324.

European Commission, 2017. Commission notice on guidance document on addressing microbiological risks in fresh fruits and vegetables at primary production through good hygiene. Off. J. Eur. Union C 163, 1–40.

Fu, L.L., Li, J.R., 2014. Microbial source tracking: a tool for identifying sources of microbial contamination in the food chain. Crit. Rev. Food Sci. Nutr. 54, 699–707. https://doi.org/10.1080/10408398.2011.605231.

García-Presedo, I., Pedraza-Díaz, S., González-Warleta, M., Mezo, M., Gómez-Bautista, M., Ortega-Mora, L.M., Castro-Hermida, J.A., 2013. Presence of Cryptosporidium scrofarum, C. suis and C. parvum subtypes IIaA16G2R1 and IIaA13G1R1 in Eurasian wild boars (Sus scrofa). Vet. Parasitol. 196, 497–502. https://doi.org/10.1016/j.vetpar.2013.04.017.

Gharpure, R., Perez, A., Miller, A.D., Wikswo, M.E., Silver, R., Hlavsa, M.C., 2009. Morbidity and mortality weekly report cryptosporidiosis outbreaks-United States, 2009-2017. Morb. Moral. Wkly. Rep. 68, 568–572.

Gherasim, A., Lebbad, M., Insulander, M., Decraene, V., Kling, A., Hjertqvist, M., Wallensten, A., 2012. Two geographically separated food-borne outbreaks in Sweden linked by an unusual cryptosporidium parvum subtype, october 2010. Eurosurveillance 17, 1–8. https://doi.org/10.2807/ese.17.46.20318-en.

Global GAP, 2016. Integrated farm assurance. In: All Farm Base - Crops Base - Fruit and Vegetables.

Hägglund, M., Bäckman, S., Macellaro, A., Lindgren, P., Borgmästars, E., Jacobsson, K., Dryselius, R., Stenberg, P., Sjödin, A., Forsman, M., Ahlinder, J., 2018. Accounting for bacterial overlap between raw water communities and contaminating sources improves the accuracy of signature-based microbial source tracking. Front. Microbiol. 9, 1–15. https://doi.org/10.3389/fmicb.2018.02364.

Harvala, H., Ögren, J., Boman, P., Riedel, H.M., Nilsson, P., Winiecka-Krusnell, J., Beser, J., 2016. Cryptosporidium infections in Sweden—understanding the regional differences in reported incidence. Clin. Microbiol. Infect. 22, 1012–1013. https://doi.org/10.1016/j.cmi.2016.09.012.

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'Brien, P., Deletic, A., McCarthy, D., 2016. Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. Water Res. 93, 242–253. https://doi.org/10.1016/j.watres.2016.02.029.

Insulander, M., de Jong, B., Svenungsson, B., 2008. A food-borne outbreak of cryptosporidiosis among guests and staff at a hotel restaurant in Stockholm county, Sweden, September 2008. Euro Surveill. 13, 12–13. https://doi.org/10.2807/ese.13.51.19071-en.

Insulander, M., Silverlås, C., Lebbad, M., Karlsson, L., Mattsson, J.G., Svenungsson, B., 2013. Molecular epidemiology and clinical manifestations of human cryptosporidiosis in Sweden. Epidemiol. Infect. 141, 1009–1020. https://doi.org/10.1017/S0950268812001665.

Kase, J.A., Zhang, G., Chen, Y., 2017. Recent foodborne outbreaks in the United States linked to atypical vehicles — lessons learned. Curr. Opin. Food Sci. 18, 56–63. https://doi.org/10.1016/j.cofs.2017.10.014.

Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L., 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26, 1721–1729. https://doi.org/10.1101/gr.210641.116.

Kinross, P., Beser, J., Troell, K., Silverlas, C., Björkman, C., Lebbad, M., Winiecka-Krusnell, J., Lindh, J., Löfdahl, M., 2015. Cryptosporidium parvum infections in a cohort of veterinary students in Sweden. Epidemiol. Infect. 143, 2748–2756. https://doi.org/10.1017/S0950268814000318.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent microbial source tracking. Nat. Methods 8, 761–763. https://doi.org/10.1038/nmeth.1650.

Kvalitetssystem, Sigill, 2016. Standard för kvalitetssäkrad produktion av bär, frukt, potatis, frilands- och växthusgrönsaker med tillval för klimatcertifiering.

Lebbad, M., Winiecka-Krusnell, J., Stensvold, C.R., Beser, J., 2021. High diversity of cryptosporidium species and subtypes identified in cryptosporidiosis acquired in Sweden and abroad. Pathogens 10. https://doi.org/10.3390/pathogens10050523.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Livsmedelsverket, 2019. Bevattningsvatten - kunskapsunderlag. Livsmedelsverkets samarbetsrapport S 2019 nr 02.

Lokmer, A., Cian, A., Froment, A., Gantois, N., Viscogliosi, E., Chabé, M., Ségurel, L., 2019. Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. PLoS One 14, 1–20. https://doi.org/10.1371/journal.pone.0211139.

Lynch, M.F., Tauxe, R.V., Hedberg, C.W., 2009. The growing burden of foodborne outbreaks due to contaminated fresh produce: risks and opportunities. Epidemiol. Infect. 137, 307–315. https://doi.org/10.1017/S0950268808001969.

McKee, A.M., Bradley, P.M., Shelley, D., McCarthy, S., Molina, M., 2021. Feral swine as sources of fecal contamination in recreational waters. Sci. Rep. 11, 4212. https://doi.org/10.1038/s41598-021-83798-6.

McKerr, C., O'Brien, S.J., Chalmers, R.M., Vivancos, R., Christley, R.M., 2018. Exposures associated with infection with Cryptosporidium in industrialised countries: a systematic review protocol. Syst. Rev. 7 https://doi.org/10.1186/s13643-018-0731-8.

McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217. https://doi.org/10.1371/journal.pone.0061217.

National Veterinary Institute, 2020. Surveillance of Infectious Diseases in Animals and Humans in Sweden 2019.

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L., McLellan, S.L., 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb. Ecol. 65, 1011–1023. https://doi.org/10.1007/s00248-013-0200-9.

Enterococcus. In: NMKL (Ed.), 2011. Determination in Foods and Feeds, 5:th ed. NMKL 68, Oslo, Norway.

NMKL, 2015. Sulphite-reducing clostridia. In: NMKL (Ed.), Determination in Foods, 56. Oslo, Norway.

O'Dea, C., Zhang, Q., Staley, C., Masters, N., Kuballa, A., Fisher, P., Veal, C., Stratton, H., Sadowsky, M.J., Ahmed, W., Katouli, M., 2019. Compositional and temporal stability of fecal taxon libraries for use with SourceTracker in sub-tropical catchments. Water Res. 165, 114967 https://doi.org/10.1016/j.watres.2019.114967.

Oliveira, B.C.M., Ferrari, E.D., da Cruz Panegossi, M.F., Nakamura, A.A., Corbucci, F.S., Nagata, W.B., dos Santos, B.M., Gomes, J.F., Meireles, M.V., Widmer, G., Bresciani, K.D.S., 2017. First description of Cryptosporidium parvum in carrier pigeons (Columba livia). Vet. Parasitol. 243, 148–150. https://doi.org/10.1016/j.vetpar.2017.06.023.

Pullerits, K., Chan, S., Ahlinder, J., Keucken, A., Rådström, P., Paul, C.J., 2020. Impact of coagulation–ultrafiltration on long-term pipe biofilm dynamics in a full-scale chloraminated drinking water distribution system. Environ. Sci. Water Res. Technol. 6, 3044–3056. https://doi.org/10.1039/d0ew00622j.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., Kelso, J., 2015. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics 31, 770–772. https://doi.org/10.1093/bioinformatics/btu719.

Robertson, L.J., Chalmers, R.M., 2013. Foodborne cryptosporidiosis: is there really more in Nordic countries? Trends Parasitol. 29, 3–9. https://doi.org/10.1016/j.pt.2012.10.003.

Robertson, L.J., van der Giessen, J.W.B., Batz, M.B., Kojima, M., Cahill, S., 2013. Have foodborne parasites finally become a global concern? Trends Parasitol. https://doi.org/10.1016/j.pt.2012.12.004.

Roguet, A., Eren, A.M., Newton, R.J., McLellan, S.L., 2018. Fecal Source Identification Using Random Forest. Microbiome, pp. 1–15. https://doi.org/10.1186/s40168-018-0568-3 in Press.

Ryan, U., Hijjawi, N., Xiao, L., 2018. Foodborne cryptosporidiosis. Int. J. Parasitol. 48, 1–12. https://doi.org/10.1016/j.ijpara.2017.09.004.

Silverlås, C., Blanco-Penedo, I., 2013. Cryptosporidium spp. in calves and cows from organic and conventional dairy herds. Epidemiol. Infect. 141, 529–539. https://doi.org/10.1017/S0950268812000830.

Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J.L., Tringe, S.G., 2015. Primer and platform effects on 16S rRNA tag sequencing. Front. Microbiol. 6, 771. https://doi.org/10.3389/fmicb.2015.00771.

Walker, A.W., Duncan, S.H., Louis, P., Flint, H.J., 2014. Metagenomics of the human gut microbiota. Trends Microbiol. 22, 267–274.

Wells, B., Shaw, H., Hotchkiss, E., Gilray, J., Ayton, R., Green, J., Katzer, F., Wells, A., Innes, E., 2015. Prevalence, species identification and genotyping Cryptosporidium from livestock and deer in a catchment in the Cairngorms with a history of a contaminated public water supply. Parasit. Vectors 8, 1–13. https://doi.org/10.1186/s13071-015-0684-x.

Wu, S., Sun, C., Li, Y., Wang, T., Jia, L., Lai, S., Yang, Y., Luo, P., Dai, D., Yang, Y.-Q., Luo, Q., Gao, N.L., Ning, K., He, L., Zhao, X.-M., Chen, W.-H., 2020. GMrepo: a database of curated and consistently annotated human gut metagenomes. Nucleic Acids Res. 48, D545–D553. https://doi.org/10.1093/nar/gkz764.

Wylezich, C., Belka, A., Hanke, D., Beer, M., Blome, S., Höper, D., 2019. Metagenomics for broad and improved parasite detection: a proof-of-concept study using swine faecal samples. Int. J. Parasitol. 49, 769–777. https://doi.org/10.1016/j.ijpara.2019.04.007.

Xiao, L., Escalante, L., Yang, C., Sulaiman, I., Escalante, A.A., Montali, R.J., Fayer, R., Lal, A.A., 1999. Phylogenetic analysis of Cryptosporidium parasites based on the small- subunit rRNA gene locus. Appl. Environ. Microbiol. 65, 1578–1583. https://doi.org/10.1128/aem.65.4.1578-1583.1999.

Xiao, L., Bern, C., Limor, J., Sulaiman, I., Roberts, J., Checkley, W., Cabrera, L., Gilman, R.H., Lal, A.A., 2001. Identification of 5 types of Cryptosporidium parasites in children in Lima. Peru. J. Infect. Dis. 183, 492–497. https://doi.org/10.1086/318090.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R., Gordon, J.I., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227. https://doi.org/10.1038/nature11053.