International Journal for Parasitology 49 (2019) 769–777

Contents lists available at ScienceDirect

International Journal for Parasitology

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

Metagenomics for broad and improved parasite detection: a proof-of-concept study using swine faecal samples ☆

Claudia Wylezich [a,*], Ariane Belka [a], Dennis Hanke [a,1], Martin Beer [a], Sandra Blome [a], Dirk Höper [a,*]

[a] Institute of Diagnostic Virology, Friedrich-Loeffler-Institut (FLI), Südufer 10, 17493 Greifswald-Insel Riems, Germany

ARTICLE INFO

Article history:
Received 26 February 2019
Received in revised form 18 April 2019
Accepted 24 April 2019
Available online 27 July 2019

Keywords:
Shotgun metagenomics
Parasite detection
Subtyping
Taxonomic assignment
False-positives
Intestinal protists
RIEMS
Pig faeces

ABSTRACT

Efficient and reliable identification of emerging pathogens is crucial for the design and implementation of timely and proportionate control strategies. This is difficult if the pathogen is so far unknown or only distantly related with known pathogens. Diagnostic metagenomics – an undirected, broad and sensitive method for the efficient identification of pathogens – was frequently used for virus and bacteria detection, but seldom applied to parasite identification. Here, metagenomics datasets prepared from swine faeces using an unbiased sample processing approach with RNA serving as starting material were re-analysed with respect to parasite detection. The taxonomic identification tool RIEMS, used for initial detection, provided basic hints on potential pathogens contained in the datasets. The suspected parasites/intestinal protists (*Blastocystis, Entamoeba, Iodamoeba, Neobalantidium, Tetratrichomonas*) were verified using subsequently applied reference mapping analyses on the base of rRNA sequences. Nearly full-length gene sequences could be extracted from the RNA-derived datasets. In the case of *Blastocystis*, subtyping was possible with subtype (ST)15 discovered for the first known time in swine faeces. Using RIEMS, some of the suspected candidates turned out to be false-positives caused by the poor status of sequences in publicly available databases. Altogether, 11 different species/STs of parasites/intestinal protists were detected in 34 out of 41 datasets extracted from metagenomics data. The approach operates without any primer bias that typically hampers the analysis of amplicon-based approaches, and allows the detection and taxonomic classification including subtyping of protist and metazoan endobionts (parasites, commensals or mutualists) based on an abundant biomarker, the 18S rRNA. The generic nature of the approach also allows evaluation of interdependencies that induce mutualistic or pathogenic effects that are often not clear for many intestinal protists and perhaps other parasites. Thus, metagenomics has the potential for generic pathogen identification beyond the characterisation of viruses and bacteria when starting from RNA instead of DNA.

© 2019 The Author(s). Published by Elsevier Ltd on behalf of Australian Society for Parasitology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Evolution of pathogens, changes in host populations and behaviour as well as habitat changes favour the emergence of infectious diseases in humans and animals. The implementation of appropriate control measures and tools for swift identification of pathogens are of paramount importance. In this context, generic, unbiased approaches such as metagenomics are needed to include so far unknown or unrelated pathogens. Metagenomic workflows have been successfully implemented for viral and bacterial pathogens (e.g., Frank et al., 2011; Hoffmann et al., 2015, 2012; Hanke et al., 2017), however some important diseases may be caused by parasites. Parasitic infections are especially a problem in developing countries, but globalisation (e.g., extensive global trade and tourism) and climate change may largely shift such infections to developed countries where their diagnosis might be a challenge for inexperienced practitioners. Beside parasitic infections, many intestinal protists and helminths historically regarded as parasites are subjects of currently controversy and discussion as to whether they are parasites or commensals of the gut and to what extent their virulence depends on interactions between them and the bacterial community, possibly influencing the variability in disease expression (Chabé et al., 2017; Stensvold and van der Giezen, 2018). Recent review and opinion articles report on 'omics' studies investigating gut protozoa (Chabé et al., 2017; Marzano et al.,

☆ *Note:* The 18S rRNA sequences generated within this study are accessible under the GenBank accession numbers **MK801356–MK801508**.

* Corresponding authors.
E-mail addresses: Claudia.Wylezich@fli.de (C. Wylezich), Dirk.Hoeper@fli.de (D. Höper).
[1] Present address: Institute of Microbiology and Epizootics, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, 14163 Berlin, Germany.

https://doi.org/10.1016/j.ijpara.2019.04.007
0020-7519/© 2019 The Author(s). Published by Elsevier Ltd on behalf of Australian Society for Parasitology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

2017), and call for in-depth investigations of gut microorganisms, and not neglecting parasites but to clarify their role in health and disease and to differentiate between long-term colonisation and infection causing disease (Stensvold and van der Giezen, 2018). Also, for the purpose of renal transplantations, it might be crucial to accurately estimate the intestinal eukaryotic fauna of the donor stool before transplantation (Azami et al., 2010). A generic approach that covers all community members including pathogens and their interrelationships helps to detect any pathogens and can also illuminate the connections between them and other community members. However, the detection of parasitic infections via an 18S rDNA amplicon-based approach cannot be applied as a generic approach due to the huge diversity of eukaryotes resulting in non-availability of any generic primers.

Shotgun metagenomics sequencing is an undirected and therefore species-independent analysis tool for the identification of all taxa contained in a certain sample. This includes potential pathogens (viruses, bacteria, parasites) possibly comprised in the sample. Regarding protist and helminth parasites, however, metagenomics is of limited use since reference genomes are lacking for many of them (Ryan et al., 2017). There are only a few studies that have applied metagenomics to find parasite signatures in sequence datasets (e.g. Greninger et al., 2015). In a case of Malayan filariasis that remained undiscovered with standard diagnostic tests for helminth infections for 12 months, only metagenomic sequencing using DNA extracted from two different patient samples revealed the causative agent (*Brugia malayi*; 0.9% of the total dataset in the higher loaded eye discharge sample), and lead to optimal therapy curing the patient (Gao et al., 2016). In a study of foodborne outbreaks with unidentified pathogens, a myxozoan fish parasite (*Kudoa septumpunctata*), detected in metagenomic datasets from fish fillet (2.6% of the total dataset), was identified as the most likely causative agent (Kawai et al., 2012). In these examples, highly infected material was investigated which contained a large amount of nucleic acids from the pathogen, resulting in a reasonable quantity of reads associated with *B. malayi* or *K. septumpunctata*, respectively. In most cases, however, the abundance of parasites in tissue or faecal samples is inherently low or rather hidden among huge amounts of sequence reads originating from the host or the microbial background of the sample. Moreover, the parasite distribution is often patchy in the host tissue, hence different subsamples from the same infected tissue might have varying pathogen loads as in wild boar tissue infected with liver fluke resulting in low amounts of parasite-associated reads (<0.008%) (Wylezich et al., 2018).

As learned from the referenced examples, one criterion for successful metagenomic parasite detection is to investigate a sample with a sufficient load. Furthermore, the applied sample processing workflow should be capable of making available all nucleic acids enclosed in the sample for downstream processing including the ones encapsulated in robust cyst walls as for some bacteria and parasites, and not lose them during the whole sample processing workflow (Wylezich et al., 2018).

Another important but largely neglected point, possibly influencing the successful parasite detection with regard to the molecules initially used for metagenomics, is the use of DNA or RNA. In a study investigating stool samples via metagenomic sequencing, only one protist parasite (*Entamoeba*) was detected but none of the other pre-diagnosed protists (*Blastocystis, Cryptosporidium, Giardia*) (Schneeberger et al., 2016). A reason for missing parasitic reads in the mentioned study might be that DNA was used as the starting material. In cellular organisms, the advantage of RNA instead of DNA as the starting material is due to the high share of rRNA in total RNA that is much higher than the share of rRNA genes (rDNA) in total DNA. For bacteria, this might be approximately 80% (Filiatrault, 2011) or up to 90% of total RNA (Blazewicz et al., 2013, and references therein). rRNA can be used for taxonomic assignments of parasites and bacteria due to the huge amount of ribosomal reference sequences in contrast to missing fully sequenced and well annotated reference genomes for many parasite taxa (Marzano et al., 2017; Stensvold and van der Giezen, 2018). Therefore, rRNA sequences extracted from metagenomic datasets might be more reliable to assign to a certain taxon than many protein-coding genes, at least based on the current database status. Since easy to amplify ribosomal sequences have been used as phylogenetic markers for approximately 30 years, they have also been evaluated for the subtyping of parasites; for example being applicable for *Blastocystis* and *Entamoeba* (Stensvold et al., 2011; Alfellani et al., 2013), in contrast to *Dientamoeba* which shows a rather low level of genetic variation (Caccio et al., 2016).

In the present study, we analysed available metagenomic sequence datasets generated within a study of porcine epidemic diarrhoea virus (PEDV)-infected pigs. The datasets were already analysed for viruses and bacteria (Hanke et al., 2015, 2017) but not for sequences of parasites. For this purpose, mapping analyses with reference genomes and/or ribosomal reference sequences of the suspected candidates were performed and analysed in detail.

2. Materials and methods

2.1. Sampling and sequencing

Forty-one diarrheal samples (mostly faeces or swabs) of PEDV-infected pigs were collected and metagenomics sequenced starting from RNA as described in previous studies (Hanke et al., 2015, 2017).

2.2. Bioinformatic analysis of metagenomic datasets for parasite detection

In the present study, raw reads of the 41 metagenomic datasets from swine faeces were analysed using the software RIEMS version 4 (Scheuch et al., 2015) using the BLAST database version from August 17, 2016 to provide an overview of the taxonomic assignment of all reads. Dataset L00901, representing a negative control for parasite detection, was generated from a cell culture supernatant sample for cultivation of PEDV. For an in-depth analysis of sequences of intestinal protist parasites detected in RIEMS result protocols, the Genome Sequencer software suite (versions 2.6; Roche) was used to perform mapping analysis using 37 out of the 41 datasets. *Blastocystis hominis* subtype (ST) 7 (NCBI database **NW_013171815**), *Entamoeba histolytica* (**AAFB00000000**), and *Entamoeba invadens* (**NW_004443648**) were used as reference genomes. Additionally, the ST definition from Alfellani et al. (2013) was used for subtyping with 18S ribosomal sequences of *Blastocystis* STs with the following reference sequences: ST1 (**AB070989**), ST2 (**EU445487**), ST3 (**AB107963**), ST4 (**BSU51152**), ST5 (**AB107964**), ST6 (**EU445485**), ST7 (**AY244621**), ST8 (**AB107970**), ST9 (**KT438703**), ST10 (**KC148207**), ST11 (**GU256922**), ST12 (**EU427515**), ST13 (**KC148209**), ST14 (**KC148205**), ST15 (**KC148210**), ST16 (**EU427512**), ST17 (**KC148208**). Suspected relatives of the family Entamoebidae were identified with the 18S ribosomal sequences of *Endolimax nana* (**AF149916**), *E. histolytica* (**X56991**), *Entamoeba polecki* ST1 (**AF149913**; ST definition according to Stensvold et al., 2018), ST2 (**AF149912**), ST3 (**AJ566411**) and ST4 (**FR686392**), *Entamoeba suis* (**DQ286372**), *Iodamoeba* RL1 (**JN635742**) and RL2 (**JN635740**) as references. The mapping analyses with ribosomal sequences were refined using different identity thresholds (97%, 99%, 100%) and a minimum overlap length of reads of 95%. In addition, selected datasets were mapped against the 18S

ribosomal sequences of the pathogens *Cryptococcus neoformans* (**JQ794497**), *Cryptosporidium parvum* (**AF108865**), *Epispathidium amphoriforme* (**KT246080**), *Haemonchus contortus* (**EU086375**), *Hexamita inflata* (**L07836**), *Neobalantidium coli* (**KJ170368**), *Onchocerca cervicalis* (**DQ094174**), *Parastrongyloides trichosuri* (**AB923885**), *Protopolystoma xenopodis* (**AM051078**), *Spirometra erinaceieuropaei* (**KY552802**), *Syphacia muris* (**KY462829**) and different Trichomonadidae species, *Trichomonas vaginalis* (**JX943584**, **KM603341**), *Trichomonas gallinae* (**EU215373**), *Tritrichomonas suis* (**AY055800**) detected with RIEMS as references with 95% similarity threshold. According to the results, additional mappings using different identity thresholds (97%, 99%, 100%) and a minimum overlap length of reads of 95% were done for the two positive taxa, *N. coli* (**KJ170368**) and *Tetratrichomonas buttreyi* (**JX565058**).

The extracted 18S rRNA sequences are accessible under the GenBank accession numbers **MK801356**–**MK801508**.

2.3. Phylogenetic reconstruction using ribosomal sequences

After identifying the *Blastocystis*, Entamoebidae, *Neobalantidium, Tritrichomonas* and *Tetratrichomonas*-related sequence reads, the resulting contigs were aligned together and with sequences retrieved from GenBank using MAFFT version 7.388 (Katoh and Standley 2013) as implemented in Geneious version 10.2.3 (Biomatters, Auckland, New Zealand). Four different alignment datasets were prepared: for *Blastocystis*-related sequences, for sequences related to Entamoebidae including *Endolimax, Entamoeba, Iodamoeba*, for *Neobalantidium* and for *Tritrichomonas* and *Tetratrichomonas*-related sequences. Phylogenetic trees were constructed using PhyML version 3.0 (Guindon et al., 2010) using the GTR + GAMMA + I model with 1000 bootstrap replications and MrBayes version 3.2.6 (Ronquist and Huelsenbeck, 2003) using the GTR model with eight rate categories and a proportion of invariable sites within the Geneious software package. The Bayesian analysis was performed for 1,000,000 generations and sampled every 1000 generations for four simultaneous chains.

3. Results and discussion

3.1. Parasite-related sequences tracked with RIEMS

The RIEMS result protocols basically provide a useful overview of the taxonomic composition of the datasets and valuable hints on potential pathogens possibly contained in the sample that have to be followed up with in-depth analyses, as has already been shown for virus reads (e.g., Hoffmann et al., 2015; Pfaff et al., 2017). In our study focussing on parasite reads, a RIEMS analysis of the 41 sequence datasets (mean size 2.16 million reads) collected in relation to PEDV outbreaks in pigs (Hanke et al., 2017) resulted in many reads related to Entamoebidae (*Entamoeba* spp., *Iodamoeba* sp., *E. nana*; Amoebozoa) and *Blastocystis* (Stramenopiles) in most datasets, ranging from 0.00004 to 0.2% for Entamoebidae and from 0.002 to 0.5% for *Blastocystis* (Table 1). Thus, many 18S rRNA sequences can be extracted via reference mapping that are exploitable for follow-up analyses (see below). This is in contrast to the study of Forsell et al. (2017) who could only extract relatively smaller amounts of ribosomal sequences from metagenomic datasets generated from DNA (0.000008% – 0.002% for *Blastocystis* STs). The values for Balantiidae (Ciliophora) and Trichomonadidae (Excavata, Parabasalia)-related reads ranged from 0.0004% to 0.5% and from 0.0001% to 0.2%, respectively. The four mentioned taxonomic groups had the highest portion of eukaryotic reads beside reads assigned to mammalian taxa (originating from the host). Other eukaryotic reads frequently found with low to moderate read numbers belonged to nematodes, fungi, other ciliates, other excavates, and cercozoans. In addition, in some datasets exceptionally high numbers (0.0002% (L00919) to 3.6% (see Table 1)) were assigned to sequences coming from environmental clone library studies named "uncultured eukaryote" (TaxID 100272). These might increase the percentages for the individual eukaryote groups shown in Table 1 when correctly assigned. Nonetheless, the percentage of parasites/intestinal protists was relatively high in most datasets, considering the enormous host and bacterial background of faecal samples masking their presence, indicating the fidelity and suitability of the used metagenomics workflow.

Table 1
Results from taxonomic assignment analyses using RIEMS. Samples re-investigated in this study (specified with the library number taken from the original study) with read numbers of the total datasets obtained in the study of Hanke et al. (2017). The percentage of read counts taken from the RIEMS result protocols assigned to uncultured eukaryotes, *Blastocystis*, Entamoebida-, Balantidae-, and Trichomonadidae-related taxa are given as a heatmap. In some datasets, no reads were found (n.d., none detected).

Lib-No	Total read number	Uncultured eukaryote	*Blastocystis*	Entamoebidae	Balantiidae	Trichomonadidae
Lib00719	2,424,038	1.773	0.034	0.043	0.039	0.009
Lib00721	2,224,040	0.034	0.092	0.014	0.022	0.005
Lib00798	2,349,124	2.493	0.042	0.012	0.001	0.085
Lib00799	2,670,508	0.074	0.089	0.016	0.007	0.039
Lib00855	1,828,934	0.044	0.075	0.003	0.029	0.012
Lib00857	1,721,454	1.890	0.007	0.012	0.018	0.002
Lib00862	2,175,422	0.003	0.005	0.003	0.023	0.000
Lib00901	3,669,342	0.192	n.d	n.d.	n.d.	0.000
Lib00906	2,627,912	0.033	0.013	0.013	0.004	0.003
Lib00907	1,731,532	1.542	0.006	0.067	0.019	0.026
Lib00908	2,307,230	0.056	0.003	0.044	0.011	0.017
Lib00918	485,818	1.854	0.076	0.152	0.024	0.052
Lib00919	565,084	0.000	n.d	n.d.	n.d.	0.000
Lib00926	2,177,223	0.026	0.014	0.011	0.018	0.011
Lib00927	4,375,983	0.014	0.002	0.001	0.003	0.001
Lib00928	1,093,621	0.004	n.d	n.d.	n.d.	0.000
Lib00929	1,225,146	0.054	0.019	0.005	0.041	0.004
Lib00931	5,008,665	0.407	n.d	0.000	n.d.	0.014
Lib00932	919,246	0.020	0.013	0.016	0.030	0.004
Lib00933	910,996	2,100	0.293	0.057	0.002	0.029
Lib00998	14,504,329	0.107	0.029	0.064	0.493	0.040
Lib00999	35,470	0.169	0.028	0.020	0.008	0.028
Lib01011	1,809,062	0.132	0.193	0.054	0.009	0.061
Lib01012	1,831,855	1.206	0.107	n.d.	0.053	0.157
Lib01013	1,783,903	0.226	0.262	0.013	0.482	0.002
Lib01014	1,487,492	1.022	0.006	0.003	0.000	0.002
Lib01015	1,832,406	2.043	0.003	0.017	0.060	0.002
Lib01017	1,577,308	1.429	0.089	0.108	0.112	0.070
Lib01018	2,177,160	1.471	0.081	0.057	0.012	0.090
Lib01019	1,734,154	0.139	0.037	0.018	0.015	0.138
Lib01020	1,696,220	1.347	0.094	0.043	0.137	0.075
Lib01059	1,724,050	0.102	0.005	0.005	0.053	0.004
Lib01060	2,020,926	0.015	0.007	0.022	0.045	0.012
Lib01061	1,907,022	3.605	0.038	0.013	0.138	0.010
Lib01062	1,444,788	0.158	0.247	0.042	0.058	0.065
Lib01063	1,766,924	2.830	0.012	0.029	0.018	0.131
Lib01064	1,427,918	0.091	0.046	0.014	0.017	0.014
Lib01065	1,630,164	0.068	0.015	0.008	0.041	0.007
Lib01329	1,062,588	2.671	0.062	0.128	0.032	0.038
Lib01330	1,243,164	0.277	0.499	0.203	0.423	0.090
Lib01420	1,282,824	0.056	0.139	0.019	0.008	0.006
Colour code		< 0.001%	0.001-0.01%	0.01-0.1%	0.1-1%	> 1%

3.2. Verification of suspected parasites using mapping analyses against ribosomal reference sequences

Thirty-seven datasets were analysed in detail for *Blastocystis* and Entamoebidae-related reads. The mapping analyses using complete genomes were less efficient and resulted in relatively short contigs for which ambiguous results were obtained using BLASTn. The ribosomal sequences resulting from this approach were also short and fragmentary. In contrast, the mappings with only ribosomal reference sequences resulted in contigs representing nearly full-length 18S rRNA gene sequences in most cases (fragment lengths are given in Table 2).

Table 2
Results from mapping analyses. Samples re-investigated in this study (specified with the library number taken from the original study) with sequences found for *Blastocystis* subtypes (STs) 1, 2, 3, 5, 15, *Entamoeba polecki* STs 1, 3, *Entamoeba suis*, *Iodamoeba* sp. RL2, *Neobalantidium coli* and *Tetratrichomonas buttreyi* based on different thresholds for the mapping analysis. The given numbers represent the fragment length of the obtained 18S ribosomal sequences found with minimum identities of 97% (grey numbers), 99% (black numbers, bold) and 100% (white bold numbers with dark grey background). Some datasets were not analysed (n.a.) due to the results shown in Table 1.

Lib-No	*Blasto-cystis* ST1	*Blasto-cystis* ST2	*Blasto-cystis* ST3	*Blasto-cystis* ST5	*Blasto-cystis* ST15	*E. polecki* ST1	*E. polecki* ST3	*E. suis*	*Iod-amoeba* RL2	*N. coli*	*Tetratri-chomo-nas*
L00719	-	-	-	**1749**	1877	-	-	-	2181	**1601**	-
L00721	**1713**	-	**1752**	**1737**	-	-	-	-	-	**1598**	-
L00798	**1731**	-	**1724**	**1744**	-	-	-	-	-	**1605**	-
L00799	**1731**	1654	**1724**	**1741**	-	**1809**	-	-	-	**1600**	-
L00855	-	1632	-	1738	1903	-	-	-	-	1617	-
L00857	-	-	-	1737	-	-	1801	-	-	**1600**	-
L00862	-	-	-	-	-	-	-	-	-	-	-
L00901	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
L00906	-	-	-	**1749**	1878	-	-	-	-	1606	1377
L00907	-	-	-	**1648**	-	1824	**1604**	-	2171	**1600**	-
L00908	-	-	-	1737	-	**1821**	1713	1937	1945	1608	1537
L00918	-	-	-	**1737**	-	-	1799	-	2168	1607	-
L00919	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
L00926	1631	-	-	**1738**	-	-	-	-	-	**1607**	1546
L00927	-	-	-	-	-	-	-	-	-	-	-
L00928	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.
L00929	-			**1729**	-	-	-	-	-	**1607**	-
L00931	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
L00932	1534	-	-	-	-	-	-	-	-	-	-
L00933	-	-	-	**1748**	1880	-	-	-	-	**1499**	-
L00998	-	-	-	**1757**	1865	**1832**	-	1941	2170	**1626**	**1547**
L00999	-	-	-	-	-	-	-	-	-	-	-
L01011	1709	-	-	**1751**	1892	-	-	-	2167	**1602**	**1544**
L01012	**1647**	-	1727	**1741**	-	**1800**	**1724**	-	-	**1635**	**1550**
L01013	-	-	-	**1753**	-	-	-	-	-	**1629**	-
L01014	-	-	-	**1737**	1873	-	-	-	-	**1430**	-
L01015	-	-	-	1736	1836	-	-	1941	-	**1636**	-
L01017	1675	-	-	**1751**	**1875**	**1828**	**1714**	1926	2176	**1629**	-
L01018	-	-	-	**1741**	-	**1798**	-	-	2173	**1607**	1544
L01019	1593	-	-	**1712**	-	-	-	-	2170	**1606**	1457
L01020	**1725**	-	**1727**	**1740**	1905	**1834**	1805	1768	-	**1601**	1540
L01059	-	-	-	**1733**	-	-	-	-	-	**1612**	-
L01060	-	-	-	1778	-	-	-	-	-	**1607**	-
L01061	**1630**	-	**1673**	**1775**	-	-	-	-	-	**1601**	-
L01062	**1722**	-	-	**1763**	1896	**1811**	**1727**	-	2171	**1612**	1530
L01063	-	-	-	**1746**	-	1816	-	-	-	**1619**	-
L01064	-	-	-	**1749**	1879	1829	-	-	-	**1607**	-
L01065	-	-	-	**1739**	-	-	-	-	-	**1606**	-
L01329	-	-	-	**1748**	-	1811	1738	-	2171	**1608**	-
L01330	-	-	-	**1776**	**1858**	**1815**	**1722**	**1784**	2180	**1601**	-
L01420	-	-	-	**1740**	1878	-	-	-	2169	1600	

For *Blastocystis*, five out of 17 tested STs were confirmed in the investigated datasets. These are ST1 (in 12 datasets), ST2 (in 2 datasets), ST3 (in 6 datasets), ST5 (in 33 datasets) and ST15 (in 14 datasets). The identified sequences were detected mostly based on high identity to the used reference. ST15 and especially ST2 sequences from pig faeces were more divergent from the used reference sequences. For Entamoebidae, no contigs were detected for *E. nana*, *E. histolytica* and *Iodamoeba* RL1. In contrast, we found *E. suis* (in 6 datasets), and *Iodamoeba* RL2 (in 13 datasets). For *E. polecki*, four subtypes were tested, but only ST 1 (in 13 datasets) and ST3 (in 10 datasets) were detected. The reference strain NIH:1293:1 (**AF149913**) was isolated from a pig source (ST1, Stensvold et al., 2011) and resulted in the most sequences found, largely based on high identity to the reference. With respect to other taxa found in the RIEMS result protocols that might have pathogenic potential, only two of the additionally tested taxa could be confirmed with the mapping analysis; *N. coli* (in 33 datasets) and *T. buttreyi* (in 10 datasets) (see Table 2) and *T. suis* in only one dataset (L00998).

3.3. Discrepancy between methods with respect to false-positive assignments

Several mappings using ribosomal sequences of suspected taxa (results from RIEMS) revealed mis-assignments rendering them false-positives. For example, assignments to *C. parvum* and *E. amphoriforme* turned out to be reads of *N. coli* whereas assignments to *T. vaginalis* turned out to be reads of *T. buttreyi*. Then, contigs gained from mappings with *Cryptococcus neoformans* (Fungi) and *Protopolystoma xenopodis* (Platyhelminthes) turned out to belong to uncultured fungi rather than being *Cryptococcus* or *Protopolystoma,* whereas contigs gained from mappings with the nematode taxa *H. contortus, O. cervicalis, P. trichosuri, S. muris* and the platyhelminth *S. erinaceieuropaei* were identified as mis-assignments of mammalian sequences. This type of mis-assignment with parasites is very common in metagenomic datasets. For instance, the genome of the apicomplexan parasite *Hammondia hammondi* was found to contain human as well as microbial sequence contaminations (Kirstahler et al., 2018), possibly delivering false-positive *Hammondia* hits for datasets containing human sequences as in the mentioned study. In the case of Entamoebidae, *E. histolytica* relationships were detected with RIEMS (Table 1) although they could not be verified using reference mapping. This is due to the lack of a reference genome for *Iodamoeba* in public databases. Hence, the found protein-coding genes of *Iodamoeba* show the highest similarity to such genes of genomes available for *E. histolytica.* With the specific reference mapping analysis, *Iodamoeba* could be verified.

3.4. Relevance of detected parasite signatures

In 34 out of 41 datasets, at least one of the 11 thoroughly tested different taxa/STs was found. *Neobalantidium coli* was the most

prevalent species (80%), with sequences nearly identical to each other and to the reference sequence branching within a cluster of closely related sequences (Fig. 1). Pig and wild boar are the reservoir of *N. coli*, and humans are rarely infected with this zoonotic pathogen (Bellanger et al., 2013). Balantidiosis is a cosmopolitan but often neglected parasitosis that is considered an emerging protozoan pathogen for humans transmitted by ingesting contaminated water or food causing dysentery or, in rare cases, pneumonia-like disease (Schuster and Ramirez-Avila, 2008). In pigs, the species is asymptomatic and high prevalences are reported, increasing with age from 0% for 1-week old piglets up to 100% for ≥4-week old pigs (Hindsbo et al., 2000; Damriyasa and Bauer, 2006). The prevalence of 80% as found in our datasets from pigs of different ages is in agreement with the data from the literature.

Blastocystis hominis is one of the eukaryotes most frequent found in mammalian intestines with 30–50% presence in the human population in Europe, or a higher percentage in developing countries. It is mostly apathogenic but may be an opportunistic agent in cases of long-lasting diarrhoea or in immunosuppressed patients, and is seen as an emerging pathogen depending on parasite localisation (Tan, 2008; Stensvold and Clark, 2016). To date, 17 *Blastocystis* STs are known (Alfellani et al., 2013), with partly unknown correlations between ST and pathogenicity (Chabé et al., 2017). *Blastocystis* ST5 was reported with high prevalence in pigs with infrequent occurrence of ST1 and ST3, which are more commonly reported in humans (Wang et al., 2014). ST2 is typically found in primates (Stensvold et al., 2009). In the present study, five STs were confirmed in pig faeces with ST5 being the most prevalent ST (80%). For *Blastocystis*-related reads, ST3 and ST5 sequences were mostly detected when a high similarity threshold was applied, and show a close relationship to each other within the respective ST cluster (Fig. 2). In contrast, for ST1, ST2 and ST15, the sequences extracted from the metagenomic datasets show a higher heterogeneity among the sequences within the ST clusters, indicating deviating STs. In addition, ST15 was detected for the first known time in swine faeces (with a prevalence of 34%) exhibiting a slightly lower identity to the reference sequence (approximately 97%) and a higher diversity among the determined sequences. ST15 was defined when it was first detected in camel and gibbon faecal material in a study screening zoo animals and livestock for *Blastocystis* (Alfellani et al., 2013). The vicinity of the ST to reptilian *Blastocystis* rather than to mammalian *Blastocystis* was critically discussed with respect to possible contamination during sampling. This possibility was rejected since the ST15 isolate was cultivated at 37 °C, which would not have been possible with reptilian *Blastocystis* (Alfellani et al., 2013). The frequent detection of ST15-related sequences in swine faeces from different farms in the present study supports the mammalian origin of the *Blastocystis* ST15 and reveals a so far unnoticed ST for pigs.

The parasite *E. histolytica* is the causative agent of amoebic dysentery in humans and mammals. *Entamoeba* and *Iodamoeba* are genera mostly apathogenic in pigs, and moderate to high prevalence is reported, increasing with age (Damriyasa and Bauer, 2006). *Entamoeba suis* and *E. polecki* are known as typical inhabitants of the gut microbiome in pigs but may also have pathogenic effects (Matsubayashi et al., 2014, 2015, 2016). *Entamoeba polecki* ST3 was linked to the occurrence of lesions in the intestinal lamina propria of pigs with abnormal faeces (Hirashima et al., 2017). *Entamoeba polecki* ST1 and ST3 have the potential to infect both humans and pigs. Based on similar morphologies, the differentiation between *Entamoeba* spp./STs using microscopical methods is difficult and ribosomal sequences are typically used for species and ST determination. A recently designed diagnostic PCR

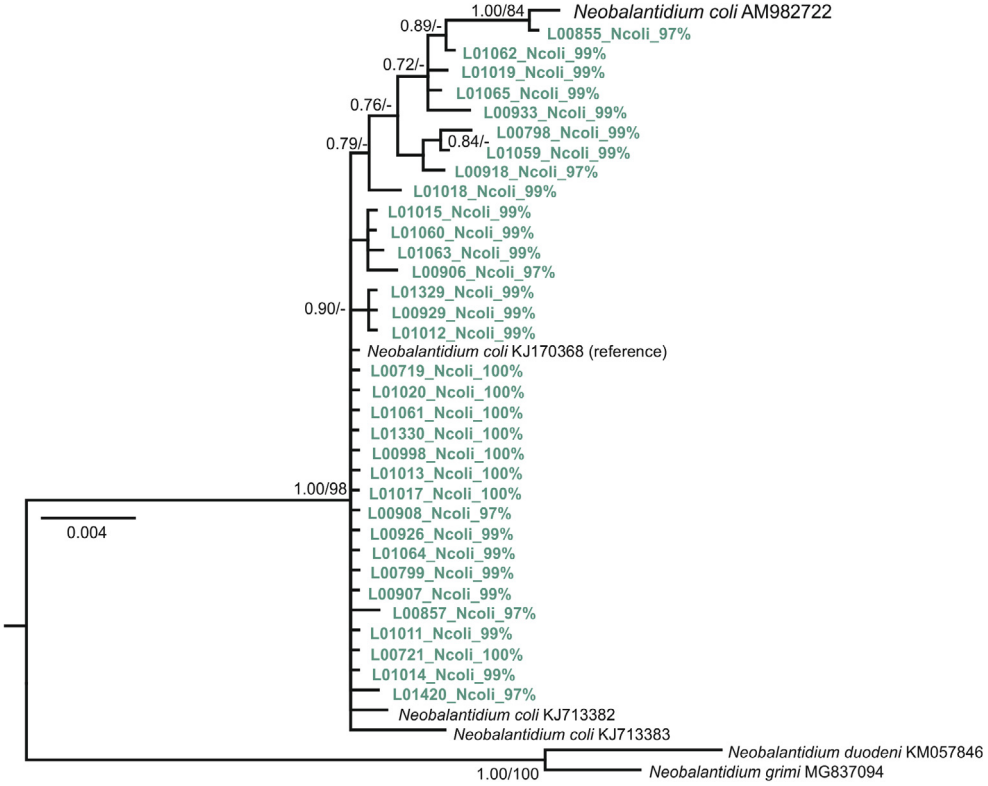

Fig. 1. Phylogenetic relationships of *Neobalantidium*-related sequences. Relationship of the *Neobalantidium coli* sequences extracted from the metagenomics datasets (in colour). The reference sequence used for mapping analysis is included and labelled. Support values (MrBayes/Maximum Likelihood) above 70% are given. The tree was rooted with the outgroup taxon *Epispathidium* sp. **KT246081** (NCBI database) (not shown).

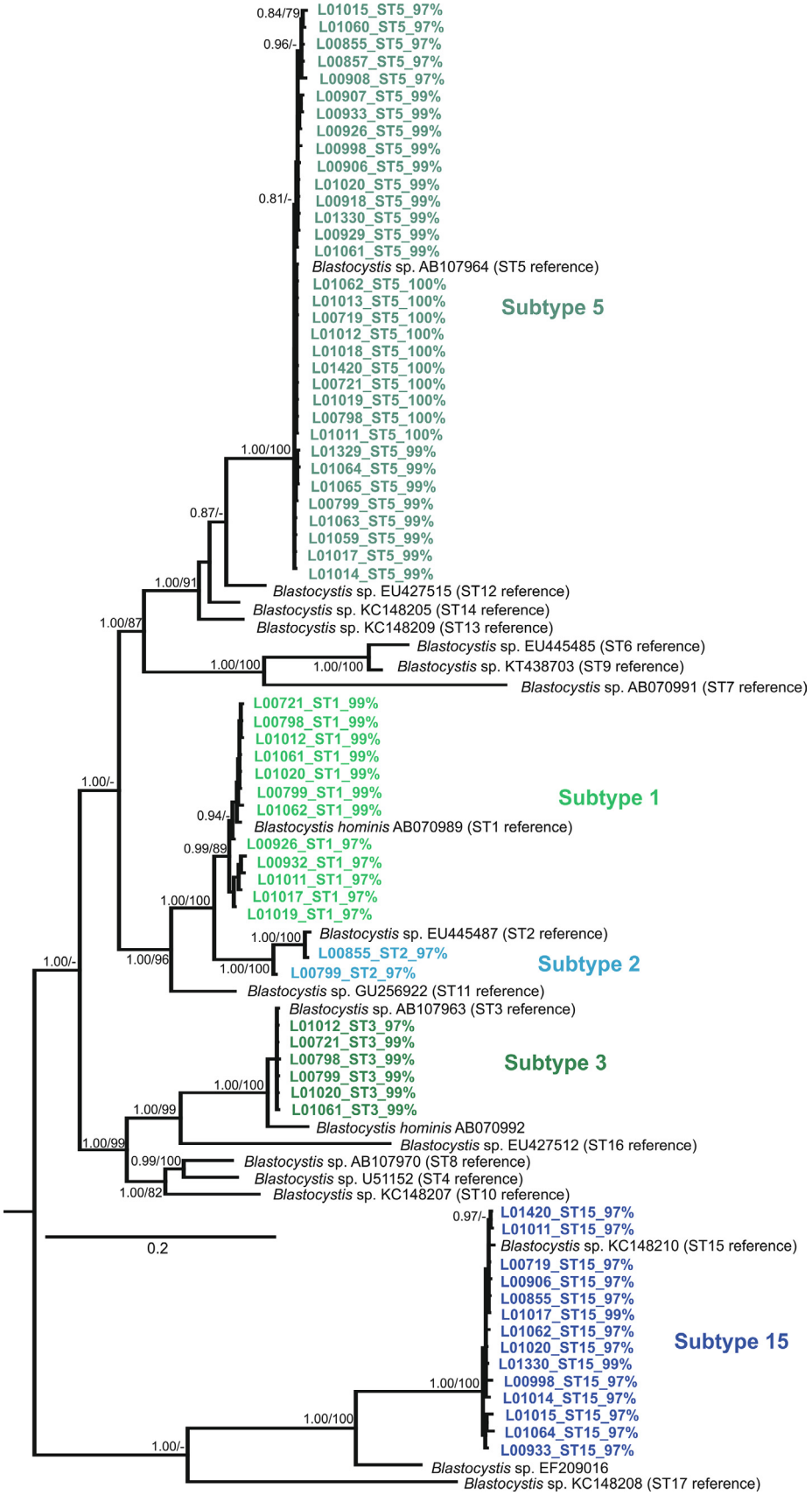

Fig. 2. Phylogenetic relationships of *Blastocystis*-related sequences. The relationships of the 17 *Blastocystis* subtypes (STs) are shown including new sequences extracted from the metagenomics datasets in coloured clusters representing ST1, ST2, ST3, ST5 and ST15. The reference sequences used for mapping analysis are included and labelled. Support values (MrBayes/Maximum Likelihood) above 70% are given. The tree was rooted with the outgroup taxon *Blastocystis lapemi* **AY266471** (NCBI database) (not shown).

for *Entamoeba* in pigs allows the distinction between *E. suis*, *E. polecki* ST1 and ST3, but does not include other species or STs of this genus (Hirashima et al., 2017). In the present study, *E. polecki* ST1 (typical for pigs)-related sequences could be extracted from metagenomic datasets based on a 99% threshold, showing a close relationship to each other (Fig. 3). Sequences of *E. suis* are closely related to each other, although they were found mostly based on 97% sequence identity, indicating a slightly differing sequence type than used as the reference. In the case of *Iodamoeba* RL2 (found in 32% of the datasets), the sequences extracted from the metagenomic datasets using 97% sequence identity show a higher heterogeneity.

Tetratrichomonas buttreyi is commonly found in pig faeces and is known to be apathogenic but may generally cause disease in cattle (Castella et al., 1997). We found a moderate prevalence of 24% of the infected pig samples, mostly based on 97% sequence identity, showing a higher heterogeneity among the sequences, indicating a slightly differing sequence type than used as the reference (Fig. 4), agreeing well with data from the literature (e.g., Li et al., 2015). *Tritrichomonas suis* was found in only one dataset (L00998) but with 100% identity to the used reference sequence.

3.5. Scientific considerations in relation to the parasites in the investigated cases

Many parasites/intestinal protists are typically found in pig faeces, with *Neobalantidium (Balantidium) coli* possibly exhibiting the highest prevalence (e.g., Pakandl, 1994). The prevalence in piglets

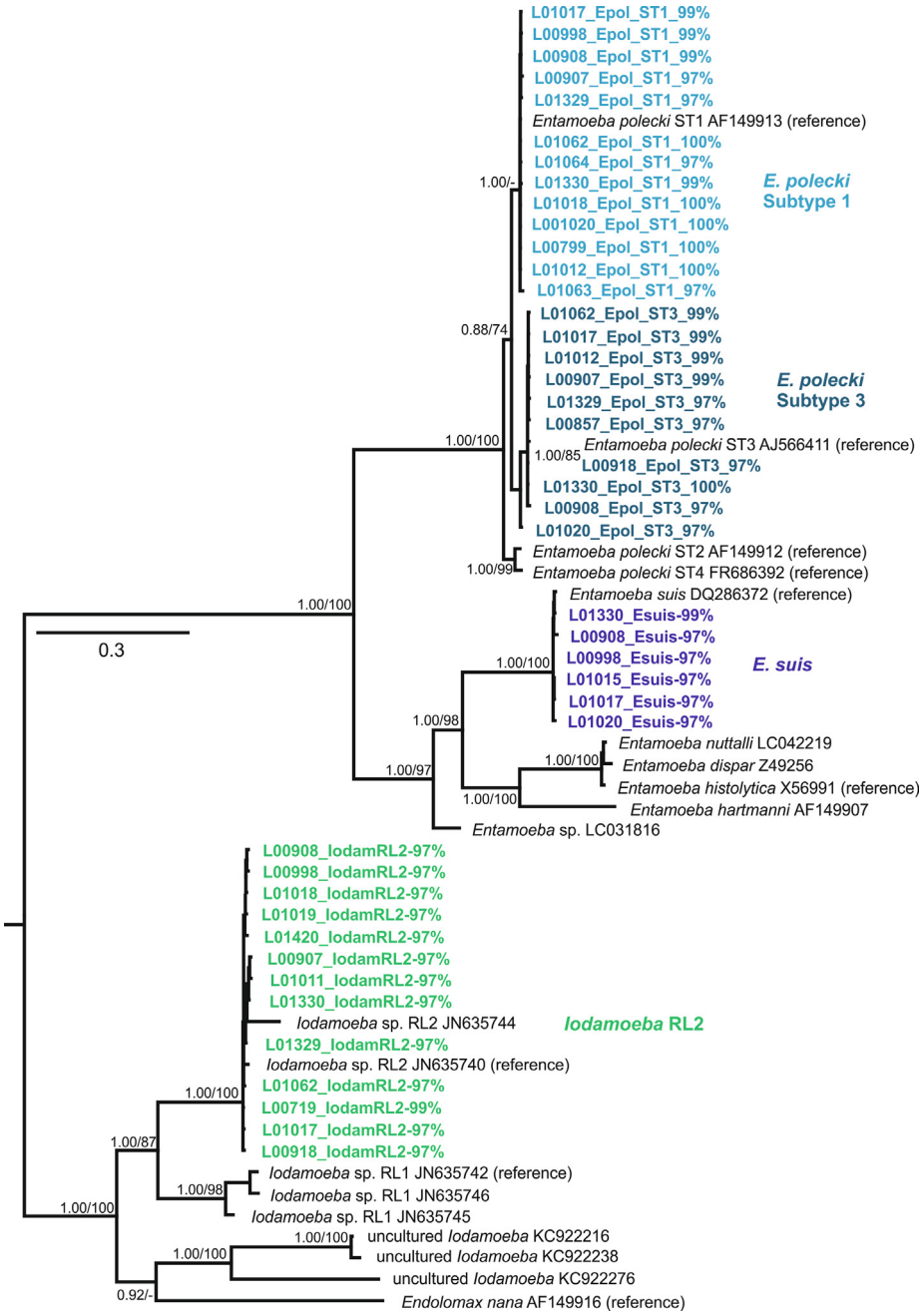

Fig. 3. Phylogenetic relationships of Entamoebidae-related sequences. The relationships of *Entamoeba* and *Iodamoeba* sequences are shown including new sequences extracted from the metagenomics datasets in coloured clusters representing *Entamoeba polecki* subtypes (STs) 1 and 3, *Entamoeba suis* and *Iodamoeba* sp. RL2. The reference sequences used for mapping analysis are included and labelled. Support values (MrBayes/Maximum Likelihood) above 70% are given. The tree was rooted with the outgroup taxon *Mastigamoeba simplex* **AF421218** (NCBI database) (not shown).

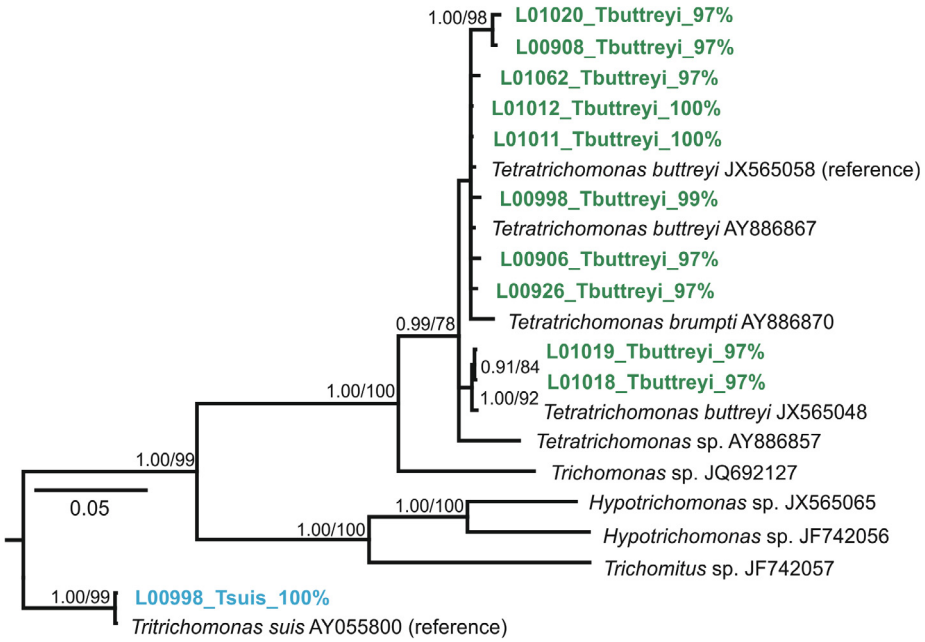

Fig. 4. Phylogenetic relationships of Trichomonadidae-related sequences. The relationships of the *Tetratrichomonas buttreyi*-like sequences and one *Tritrichomonas suis* sequence extracted from the metagenomics datasets (in colour). The reference sequence used for mapping analysis is included and labelled. Support values (MrBayes/Maximum Likelihood) above 70% are given. The tree was rooted with the outgroup taxon *Histomonas meleagridis* **AF293056** (NCBI database) (not shown).

was found to be very low for intestinal parasites in general (Pakandl, 1994; Damriyasa and Bauer, 2006). However, studies on the diversity, but especially prevalence and epidemiology, are unreliable as they often use microscopic investigations that are time-consuming and depend on the type of preparation/method and the taxonomic expertise of the researcher. Here we report on the protist parasite composition of 41 datasets from faecal samples derived from sows, fattening pigs and piglets (Hanke et al., 2017) as extracted from metagenomic datasets.

Altogether, the composition of parasites/intestinal protists detected in the present study is typical for most taxa with *Blastocystis* ST15 discovered for the first known time in pig faeces. A high mortality in piglets was reported for datasets L00906-908, L00998, L01060 and L01420 for which the PEDV sequences clustered in two distinct clusters (Hanke et al., 2017). Interestingly, in the mentioned samples the pig-typical *Blastocystis* ST1 and ST3, but also ST2, are absent, pointing to the possibility that pigs from other samples (with lower mortality) might have a certain *Blastocystis*-induced protection against a disadvantageous bacterial composition (Chabé et al., 2017; Forsell et al., 2017; Stensvold and van der Giezen, 2018). In two of the six mentioned datasets (L00907, L00908), *E. polecki* ST3, known to cause lesions in the intestinal lamina propria, was detected which might have contributed to the high mortality rates. Altogether, the occurrence of the detected taxa is plausible and consistent with the literature but drawing any conclusions is greatly hampered by missing metadata for some of the datasets.

3.6. Concluding remarks

Our study underlines that metagenomic sequencing is the preferred tool for parasite detection in contrast to PCR-based approaches since no PCR-caused biases influence the outcome and unexpected taxa not targeted by used primers cannot be overlooked. Moreover, it is a holistic approach that can help study viruses, eukaryotes and prokaryotes in parallel, and find interdependencies inducing mutualistic or pathogenic effects (Andersen et al., 2015). The present study clearly demonstrates that metagenomic sequencing is applicable not only for viruses and bacteria but also for the detection of protist and metazoan endobionts (parasites, commensals or mutualists).

We recommend analysis of metagenomic datasets with a combination of taxonomic assignment tools such as RIEMS, and subsequent reference mapping analyses for successful parasite detection, offering an easy detection tool that allows clear species assignment and subtyping as shown for *Blastocystis* and *E. polecki*. Prerequisites for successful detection of pathogen signatures are the avoidance of pre-treatments such as filtration or centrifugation that will always exclude bacteria or parasites, and the inclusion of an efficient sample disintegration step prior to the nucleic acid extraction (Wylezich et al., 2018). Another prerequisite for successful pathogen detection is to start from RNA instead of DNA, as shown in the present study. This approach exploits the presence of 18S rRNA sequences, a biologically abundant biomarker, for the detection of pathogenic eukaryote (or prokaryote) taxa.

With respect to quantitative analysis, the abundance of sequence reads is occasionally used to discuss the organism abundance. Although quantification is hampered by a lack of clarity regarding the number of ribosomes within a cell (when using RNA as starting material) depending on its trophic state, quantification using sequences coming from amplification-free metagenomic datasets is basically possible rather than with sequences coming from highly biased amplicon sequencing studies. Due to the mis-assignments as discussed above, however, the quantification based on RIEMS counts has to be assessed with caution since reference databases contain many ribosomal sequences without a species designation.

Current challenges in metagenomic parasite detection are the lack of wellcurated reference genomes (e.g., Kirstahler et al., 2018) often delivering false-positive hits and the general lack of genomes for representatives of many taxonomic groups as well as different functional groups such as pathogen, symbiont and host taxa. The extraction of 18S rRNA sequences from metagenomics can circumvent the current shortcoming of missing genomes. Metagenomics datasets can be re-analysed once well annotated reference genomes have been made available to further assess

complete profiles of genes and transcripts, providing a proper perspective of parasite biology. The approach developed and applied in the present proof-of-concept study was tested only for intestinal protists/parasites in faecal samples and organ material (Wylezich et al., 2018), but the detection of parasitic pathogens in an outbreak scenario is feasible.

Acknowledgements

This work was financially supported by European Union Horizon 2020 Research and Innovation Program COMPARE (Grant Number 643476).

References

Alfellani, M.A., Taner-Mulla, D., Jacob, A.S., Imeede, C.A., Yoshikawa, H., Stensvold, C. R., Clark, C.G., 2013. Genetic diversity of *Blastocystis* in livestock and zoo animals. Protist 164, 497–509.

Andersen, L.O., Bonde, I., Nielsen, H.B., Stensvold, C.R., 2015. A retrospective metagenomics approach to studying *Blastocystis*. FEMS Microbiol. Ecol. 91, fiv072.

Azami, M., Sharifi, M., Hejazi, S.H., Tazhibi, M., 2010. Intestinal parasitic infections in renal transplant recipients. Braz. J. Infect. Dis. 14, 29–32.

Bellanger, A.P., Scherer, E., Cazorla, A., Grenouillet, F., 2013. Dysenteric syndrome due to *Balantidium coli*: a case report. New Microbiol. 36, 203–205.

Blazewicz, S.J., Barnard, R.L., Daly, R.A., Firestone, M.K., 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. 7, 2061–2068.

Caccio, S.M., Sannella, A.R., Bruno, A., Stensvold, C.R., David, E.B., Guimaraes, S., Manuali, E., Magistrali, C., Mahdad, K., Beaman, M., Maserati, R., Tosini, F., Pozio, E., 2016. Multilocus sequence typing of *Dientmoeba fragilis* identified a major clone with widespread geographical distribution. Int. J. Parasitol. 46, 793–798.

Castella, J., Munoz, E., Ferrer, D., Gutierrez, J.F., 1997. Isolation of the trichomonad *Tetratrichomonas buttreyi* (Hibler et al., 1960) Honigberg, 1963 in bovine diarrhoeic faeces. Vet. Parasitol. 70, 41–45.

Chabé, M., Lokmer, A., Segurel, L., 2017. Gut protozoa: Friends or foes of the human gut microbiota? Trends Parasitol. 33, 925–934.

Damriyasa, I.M., Bauer, C., 2006. Prevalence and age-dependent occurrence of intestinal protozoan infections in suckling piglets. Berl. Münch. Tierärztl. Wochenschr. 119, 287–290.

Filiatrault, M.J., 2011. Progress in prokaryotic transcriptomics. Curr. Opin. Microbiol. 14, 579–586.

Forsell, J., Bengtsson-Palme, J., Angelin, M., Johansson, A., Evengård, B., Granlund, M., 2017. The relation between *Blastocystis* and the intestinal microbiota in Swedish travellers. BMC Microbiol 17, 231.

Frank, C., Werber, D., Cramer, J.P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., Wadl, M., Zoufaly, A., Jordan, S., Kemper, M.J., Follin, P., Müller, L., King, L.A., Rosner, B., Buchholz, U., Stark, K., Krause, G., 2011. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. N. Engl. J. Med. 365, 1771.

Gao, D., Yu, Q., Wang, G., Wang, G., Xiong, F., 2016. Diagnosis of a malayan filariasis case using a shotgun diagnostic metagenomics assay. Parasites Vectors 9, 86.

Greninger, A.L., Messacar, K., Dunnebacke, T., Naccache, S.N., Federman, S., Bouquet, J., Mirsky, D., Nomura, Y., Yagi, S., Glaser, C., Vollmer, M., Press, C.A., Klenschmidt-DeMasters, B.K., Dominguez, S.R., Chiu, C.Y., 2015. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med. 7, 113.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hanke, D., Jenckel, M., Petrov, A., Ritzmann, M., Stadler, J., Akimkin, V., Blome, S., Pohlmann, A., Schirrmeier, H., Beer, M., Höper, D., 2015. Comparison of porcine epidemic diarrhea viruses from Germany and the United States, 2014. Emerg. Infect. Dis. 21, 493–496.

Hanke, D., Pohlmann, A., Sauter-Louis, C., Höper, D., Stadler, J., Ritzmann, M., Steinrigl, A., Schwarz, B.A., Akimkin, V., Blome, S., Beer, M., 2017. Porcine epidemic diarrhea in Europe: in-detail analyses of disease dynamics and virus evolution. Viruses 9, 177.

Hindsbo, O., Nielsen, C.V., Andressen, J., Willingham, A.L., Bendixen, M., Nielsen, M., Nielsen, N.O., 2000. Age-dependent occurrence of the intestinal ciliate *Balantidium coli* in pigs at a Danish research farm. Acta Vet. Scand. 41, 79–83.

Hirashima, Y., Manchanayake, T., Yano, T., Kitahara, S., Koreeda, T., Kamimura, S., Sasai, K., Matsubayashi, M., Shibahara, T., 2017. Development of molecular diagnostic protocols for detecting three types of *Entamoeba* from diarrheal and asymptomatic pigs and environmental moist soils. Parasitol. Res. 116, 2001–2007.

Hoffmann, B., Scheuch, B., Höper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., Eschbaumer, M., Goller, K.V., Wernike, K., Fischer, M., Breithaupt, A., Mettenleiter, T.C., Beer, M., 2012. Novel orthobunyavirus in cattle, Europe, 2011. Emerg. Infect. Dis. 18, 469.

Hoffmann, B., Tappe, D., Höper, D., Herden, C., Boldt, A., Mawrin, C., Niederstraßer, O., Müller, T., Jenckel, M., van der Grinten, E., Lutter, C., Abendroth, B., Teifke, J.P., Cadar, D., Schmidt-Chanasit, J., Ulrich, R.G., Beer, M., 2015. A variegated squirrel bornavirus associated with fatal human encephalitis. N. Engl. J. Med. 373, 154.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kawai, T., Sekizuka, T., Yahata, Y., Kuroda, M., Kumeda, Y., Iijima, Y., Kamata, Y., Sugita-Konishi, Y., Ohnishi, T., 2012. Identification of *Kudoa septempunctata* as the causative agent of novel food poisoning outbreaks in Japan by consumption of *Paralichthys olivaceus* in raw fish. Clin. Infect. Dis. 54, 1046–1052.

Kirstahler, P., Bjerrum, S.S., Friis-Møller, A., la Cour, M., Aarestrup, F.M., Westh, H., Pamp, S.J., 2018. Genomics-based identification of microorganisms in human ocular body fluid. Sci. Rep. 8, 4126.

Li, W., Li, W., Gong, P., Zhang, C., Yang, J., Zhang, X., Li, J., 2015. The prevalence of intestinal trichomonads in Chinese pigs. Vet. Parasitol. 211, 12–15.

Marzano, V., Mancinelli, L., Bracaglia, G., Del Chierico, F., Vernocchi, P., Di Girolamo, F., Garrone, S., Kuekou, H.T., D'Argenio, P., Dallapiccola, B., Urbani, A., Putignani, L., 2017. Omic" investigations of protozoa and worms for a deeper understanding of the human gut "parasitome. PLoS Negl. Trop. Dis. 11, e0005916.

Matsubayashi, M., Suzuta, F., Terayama, Y., Shimojo, K., Yui, T., Haritani, M., Shibahara, T., 2014. Ultrastructural characteristics and molecular identification of *Entamoeba suis* isolated from pigs with hemorrhagic colitis: implications for pathogenicity. Parasitol. Res. 113, 3023–3028.

Matsubayashi, M., Murakoshi, N., Komatsu, T., Tokoro, M., Haritani, M., Shibahara, T., 2015. Genetic identification of *Entamoeba polecki* subtype 3 from pigs in Japan and characterisation of its pathogenic role in ulcerative colitis. Infect. Genet. Evol. 36, 8–14.

Matsubayashi, M., Sasagawa, Y., Aita, T., Tokoro, M., Haritani, M., Shibahara, T., 2016. First report of mixed *Entamoeba polecki* (ST 1) and *E. suis* infection in piglets shedding abnormal feces by histopathological and molecular surveys. Acta Parasitol. 61, 665–670.

Pakandl, M., 1994. The prevalence of intestinal protozoa in wild and domestic pigs. Vet. Med. (Praha) 39, 377–380.

Pfaff, F., Schulze, C., König, P., Franzke, K., Bock, S., Hlinak, A., Kämmerling, J., Ochs, A., Schüle, A., Mettenleiter, T.C., Höper, D., Beer, M., 2017. A novel alphaherpesvirus associated with fatal diseases in banded Penguins. J. Gen. Virol. 98, 89–95.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Ryan, U., Paparini, A., Oskam, C., 2017. New technologies for detection of enteric parasites. Trends Parasitol. 33, 532–546.

Scheuch, M., Höper, D., Beer, M., 2015. RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. BMC Bioinf. 16, 69.

Schneeberger, P.H.H., Becker, S.L., Pothier, J.F., Duffy, B., N'Goran, E.K., Beuret, C., Frey, J.E., Utzinger, J., 2016. Metagenomic diagnostics for the simultaneous detection of multiple pathogens in human stool specimens from Côte d'Ivoire: a proof-of-concept study. Infect. Genet. Evol. 40, 389–397.

Schuster, F.L., Ramirez-Avila, L., 2008. Current world status of *Balantidium coli*. Clin. Microbiol. Rev. 21, 626–638.

Stensvold, C.R., Alfellani, M., Nørskov-Lauritsen, S., Prip, K., Victory, E.L., Maddox, C., Nielsen, H.V., Clark, C.G., 2009. Subtype distribution of *Blastocystis* isolates from synanthropic and zoo animals and identification of a new subtype. Int. J. Parasitol. 39, 473–479.

Stensvold, C.R., Clark, C.G., 2016. Current status of *Blastocystis*: A personal view. Parasitol. Int. 65, 763–771.

Stensvold, C.R., Lebbad, M., Victory, E.L., Verweij, J.J., Tannich, E., Alfellani, M., Legarraga, P., Clark, C.G., 2011. Increased sampling reveals novel lineages of *Entamoeba*: consequences of genetic diversity and host specificity for taxonomy and molecular detection. Protist 162, 525–541.

Stensvold, C.R., van der Giezen, M., 2018. Associations between gut microbiota and common luminal intestinal parasites. Trends Parasitol. 34, 369–377.

Stensvold, C.R., Winiecka-Krusnell, J., Lier, T., Lebbad, M., 2018. Evaluation of a PCR method for detection of *Entamoeba polecki*, with an overview of its molecular epidemiology. J. Clin. Microbiol. 56, e00154–18.

Tan, K.S.W., 2008. *Blastocystis* in humans and animals: new insights using modern methodologies. Vet. Parasitol. 126, 121–144.

Wang, W., Bielefeldt-Ohmann, H., Traub, R.J., Cuttell, L., Owen, H., 2014. Location and pathogenic potential of *Blastocystis* in the porcine intestine. PLoS ONE 9, e103962.

Wylezich, C., Papa, A., Beer, M., Höper, D., 2018. A versatile sample processing workflow for diagnostic metagenomics. Sci. Rep. 8, 13108.