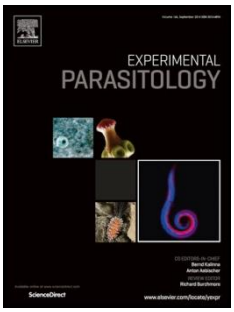
Paparini, A., Gofton, A., Yang, R., White, N., Bunce, M. and Ryan, U.M. (2015) Comparison of Sanger and next generation sequencing performance for genotyping Cryptosporidium isolates at the 18S rRNA and actin loci. Experimental Parasitology, 152. pp. 21-27.

Accepted Manuscript



Title: Comparison of sanger and next generation sequencing performance for genotyping *cryptosporidium* isolates at the 18s rRNA and actin loci.

Author: Andrea Paparini, Alexander Gofton, Rongchang Yang, Nicole White, Michael Bunce, Una M Ryan

PII: S0014-4894(15)00032-6
DOI: http://dx.doi.org/doi: 10.1016/j.exppara.2015.02.001
Reference: YEXPR 6996

To appear in: *Experimental Parasitology*

Received date: 18-11-2014
Revised date: 27-1-2015
Accepted date: 1-2-2015

Please cite this article as: Andrea Paparini, Alexander Gofton, Rongchang Yang, Nicole White, Michael Bunce, Una M Ryan, Comparison of sanger and next generation sequencing performance for genotyping *cryptosporidium* isolates at the 18s rRNA and actin loci., *Experimental Parasitology* (2015), http://dx.doi.org/doi: 10.1016/j.exppara.2015.02.001.

Comparison of Sanger and next generation sequencing performance for genotyping *Cryptosporidium* isolates at the 18S rRNA and actin loci.

Andrea Paparini[a], Alexander Gofton[a], Rongchang Yang[a], Nicole White[a,b], Michael Bunce[a,b], Una M Ryan[a*].

[a]*School of Veterinary and Life Sciences, Murdoch University, Murdoch WA, 6150, Australia.*

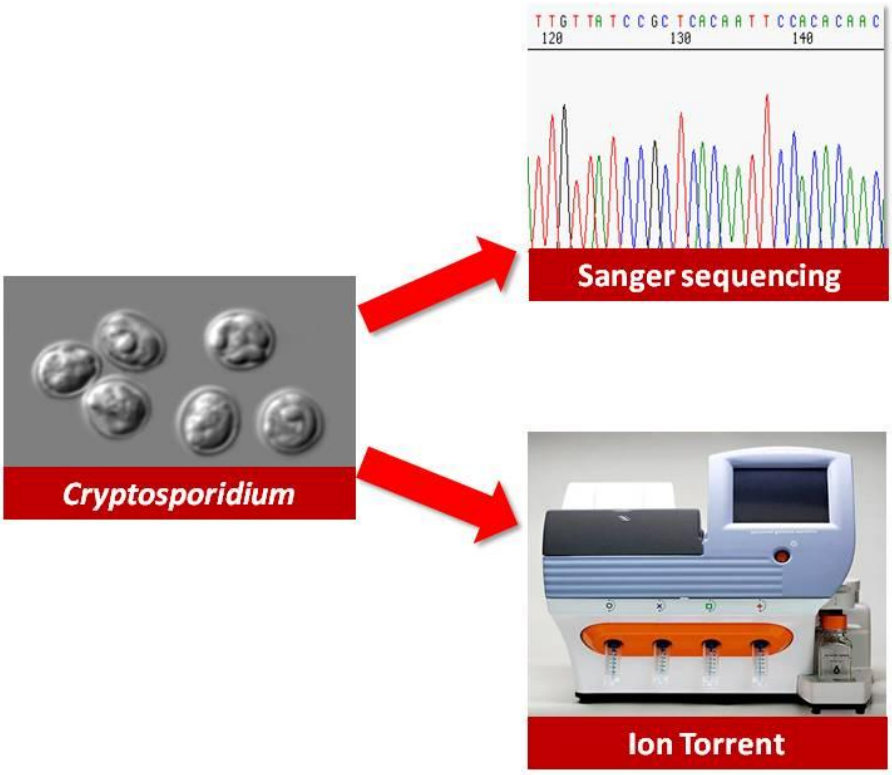
[b]*Present address: Trace and Environmental DNA (TrEnD) laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6102, Australia*

[*]Corresponding author. Mailing address: School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch, Western Australia, 6150, Australia. E–mail: Una.Ryan@murdoch.edu.au

Highlights

- Genotyping *Cryptosporidium* isolates at two loci
- Comparison of Sanger and Ion Torrent sequencing performance
- Ion Torrent superior at identifying mixed infections
- Loss of PCR efficiency on low-template samples for Ion Torrent
- Costs comparable on large numbers of samples

Graphical Abstract



Abstract

Cryptosporidium is an important enteric pathogen that infects a wide range of humans and animals. Rapid and reliable detection and characterisation methods are essential for understanding the transmission dynamics of the parasite. Sanger sequencing, and high-throughput sequencing (HTS) on an Ion Torrent platform, were compared to each other for their sensitivity and accuracy in detecting and characterizing 25 *Cryptosporidium*-positive human and animal faecal samples.

Ion Torrent reads (*n*=123,857) were obtained at both 18S rRNA and actin loci for 21 of the 25 samples. Of these, one isolate at the actin locus (Cattle 05) and three at the18S rRNA locus (HTS 10, HTS 11 and HTS 12), suffered PCR drop-out (i.e. PCR failures) when using fusion-tagged PCR. Sanger sequences were obtained for both loci for 23 of the 25 samples and showed good agreement with Ion Torrent-based genotyping. Two samples both from pythons (SK 02 and SK 05) produced mixed 18S and actin chromatograms by Sanger sequencing but were clearly identified by Ion Torrent sequencing as *C. muris*. One isolate (SK 03) was typed as *C. muris* by Sanger sequencing but was identified as a mixed *C. muris* and *C. tyzzeri* infection by HTS. 18S rRNA Type B sequences were identified in 4/6 *C. parvum* isolates when deep sequenced but were undetected in Sanger sequencing.

Sanger was cheaper than Ion Torrent when sequencing a small numbers of samples, but when larger numbers of samples are considered (*n*=60), the costs were comparative. Fusion-tagged amplicon based approaches are a powerful way of approaching mixtures, the only draw-back being the loss of PCR efficiency on low-template samples when using primers coupled to MID tags and adaptors. Taken together these data show that HTS has excellent potential for revealing the “true” composition of species/types in a *Cryptosporidium* infection, but that HTS workflows need to be carefully developed to ensure sensitivity, accuracy and contamination are controlled.

58 *Keywords: Cryptosporidium*; Ion Torrent; Sanger sequencing; 18S rRNA; actin

59

1.0 Introduction

Cryptosporidium is one of the most prevalent waterborne parasites worldwide and represents the major public health concern of water utilities in developed nations, as the oocyst stage is highly resistant to environmental stresses and disinfection treatments (including chlorine treatment of community water supplies) (Karanis et al., 2007; Baldursson and Karanis, 2011). Between 2004 and 2010, *Cryptosporidium* was responsible for 60.3% of worldwide reported waterborne protozoan parasitic outbreaks (Baldursson and Karanis, 2011). The parasite is responsible for enteric disease in a wide range of hosts and may be chronic and life-threatening in immunocompromised individuals (Hunter et al., 2007). In humans, cryptosporidiosis mainly manifests itself as watery diarrhoea with varying severity, abdominal cramps, loss of appetite, nausea, vomiting and low-grade fever (Xiao, 2010). Currently, nitazoxanide (NTZ) is approved for the treatment of cryptosporidiosis in children and immunocompetent adults in the U.S.A., however treatment failures are common and NTZ is ineffective in immunocompromised individuals (Amadi et al., 2009).

As most species of *Cryptosporidium* are morphologically indistinguishable, molecular tools are required for robust identification. Molecular data indicates that at least twenty *Cryptosporidium* species/genotypes have been identified in humans including *C. hominis, C. parvum, C. meleagridis, C. felis, C. canis, C. ubiquitum* and *C. cuniculus* (Xiao, 2010; Ryan et al., 2014), with *C. hominis* and *C. parvum* being by far the most prevalent species in humans populations worldwide (Xiao, 2010).

Current molecular characterisation methods for *Cryptosporidium* generally rely on PCR followed by Sanger sequencing, with the 18S ribosomal RNA gene (rRNA), actin and glycoprotein 60 (gp60) being the most commonly characterised loci (Xiao, 2010). Most of the commonly adopted molecular protocols involve nested-PCR amplifications, to generate

sufficient material for subsequent sanger sequencing (Mayer and Palmer, 1996). A limitation of this approach, however, is that rare species and/or mixed infections are generally not detected due to low relative abundance (Grinberg, et al., 2013). This leads to a possible underestimation of the prevalence of mixed infections and/or zoonotic *Cryptosporidium* species. Nested PCR approaches also have an inherent risk of contamination and can exhibit strong PCR bias and/or stochastic variation (Park and Crowley, 2010). An alternative approach is to employ deep sequencing amplicon-based technologies.

The advent of second generation, high-throughput sequencing (HTS) platforms have enabled the rapid sequencing of genes, genomes and metagenomes (Metzker, 2010). Their application to the analysis of mixed *Cryptosporidium* infections in human faecal samples is likely to audit the species composition more comprehensively and in a more cost-effective manner than currently practiced. The aim of the present study was to compare Sanger and a HTS method (Ion Torrent), for molecular detection and characterization of *Cryptosporidium*-positive faecal samples from humans and animals at the 18S rRNA and actin loci and to compare the cost of both methods.

2.0 Materials and methods

2.1 Sample selection and molecular analyses

Faecal DNA from animals (cattle - *Bos taurus*, n=5; South-west Carpet Python - *Morelia spilota imbricata*, n=4) and humans (n=16) was extracted using the MOBIO soil kit (MOBIO, USA; animal samples) or the QIAamp DNA Stool Mini Kit (Qiagen, USA; human samples). Samples were confirmed as *Cryptosporidium*-positive and pre-selected for the

study based on quantitative PCR (qPCR)-based screening at the 18S rRNA locus using a qPCR assay previously reported (Yang et al., 2014).

For screening, DNA was amplified using single round PCR at the 18S rRNA and actin loci, using respectively the *Cryptosporidium*-specific primer pairs 18S iF/iR (Morgan, et al., 1997) and actin primers CrActin 8F (5'-CTGTDGGWAGYGARAGATTYAG-3')/CrActin 6R (5'-GGDGCAACRACYTTRATCTTC-3') (this study). Samples identified as low template amounts (by cycle threshold-CT values) and/or that presented with severe inhibition (via a qPCR dilution series) were not processed further.

Once Sanger sequencing and Ion Torrent were completed for all pre-selected samples, the final selection of samples were made. In line with the (comparative) objective of the present analysis, all samples providing *Cryptosporidium*-specific sequences by either Sanger sequencing or Ion Torrent (at least one locus), were included in the final analysis.

2.2. Sanger and Ion Torrent sequencing

For Sanger sequencing, after electrophoresis, gel-purified single-round PCR amplicons were sequenced using the ABI Prism Terminator Cycle Sequencing kit (Applied Biosystems, USA), on an Applied Biosystem 3730 DNA Analyzer. The same primer pairs were also used for HTS, using an Ion Torrent system (ion semiconductor sequencing) (Life Technologies, USA). Sequencing was performed according to the manual published by Life Technologies. Fusion primers (IDT, USA) were based on the 18S iF/iR primers (Morgan et al., 1997) and novel CrActin 8F/6R primers, and included unique sample-specific barcodes (MID tags) and P1 and A adaptors. All PCR amplicons were double purified using the Agencourt AMPure XP Bead PCR purification protocol (Beckman Coulter Genomics, USA), pooled in roughly equimolar ratios and sequenced.

2.3. Data deconvolution and bioinformatics

Amplicon sequence reads (hereafter referred to as sequences) were filtered using Geneious 7.1.7 (Biomatters Ltd, available from http://www.geneious.com/) and sorted into sample batches based on the unique DNA tags, primer sequences and adaptors. Only sequences exhibiting exact matches to the flanking regions were processed further. To further reduce noise associated with sequencing error and low abundance, unique sequences occurring in single copy within the set of reads obtained from a given sample (singletons), were removed from the dataset and unique sets of identical reads were obtained (for simplicity these sets of identical reads, excluding singletons, will be referred to as "reads" hereafter). Each set comprised a variable number of identical reads greater than two. Sequences were then BLAST-searched against an "in-house" library of *Cryptosporidium* 18S and actin sequences. An in-silico probe (5'-TATTACTATATT-3') was used to identify Type B 18S rRNA sequences in *C. parvum* isolates.

3.0. Results

3.1. HTS sequencing of 18S and actin loci on the Ion torrent

The total number of reads obtained from 21 samples was 9,014 (18S rRNA) and 114,843 (actin). There was a positive linear correlation ($R^2 > 0.79$) between the total number of reads and the total unique reads (i.e., excl. redundant and singletons). For both loci, intra-sample genetic variation was observed in both human and animal isolates, even after filtering, when redundant reads and singletons were removed. One sample at the actin locus (Cattle 05)

and three samples at the 18S rRNA locus (HTS 10, HTS 11 and HTS 12) could not be typed by HTS due to lack of amplification of these isolates, presumably due to the fusion primers (included MID tags and adaptors) that impact on PCR efficiency. The total number of sets of unique reads/sample was higher for actin than for 18S, and showed a broad range (0-514 or 0-154, for actin and 18S, respectively), with medians of 116 (actin) and 10 (18S rDNA) and large standard deviations (129 and 37, for actin and 18S, respectively) (Table 1).

By Ion Torrent, all 18S reads obtained were *Cryptosporidium*-specific. For actin, however, 9 samples (Cattle 04, HTS 05, HTS 16, HTS 17, HTS 18, HTS 21, SK 02, SK 03 and SK 05) showed variable numbers of non-specific unique reads (no singletons), ranging from 1 to 36 (HTS 21 and Cattle 4, respectively), but generally represented a negligible fraction (~4%) of the total number of reads obtained from each sample. This result is likely due to differential quality/amount of template DNA, but may also reveal differences in the amplification performance of the two primer pairs for the two loci.

For the 21 samples for which *Cryptosporidium*-specific Ion Torrent reads were obtained at both loci (i.e., excluding Cattle 5, HTS 10, HTS 11 and HTS 12), complete agreement between 18S- and actin-based identifications were observed (Table 1).

At the actin locus, minor (<1%) genetic variants of *C. hominis* and *C. parvum* sequences were identified in 20 samples, excluding Cattle 5. These genetic variants consisted of 1-2 bp insertions or deletions (INDELs) or single nucleotide polymorphisms (SNPs), and, being > 99% identical to known *Cryptosporidium* spp., did not prevent confident identification of the respective isolates. Overall, Ion Torrent sequencing-based identifications were divided as follows: *C. hominis* (*n*=14), *C. parvum* (*n*=7), *C. muris* (*n*=3) and a *C. muris* and *C. tyzzeri* mixed infection for sample SK 03 (Table 1).

3.2. In silico probe identification of Type B 18S rRNA for C. parvum.

Type B 18S rRNA sequences were detected in five of the seven *C. parvum* isolates with no Type B sequence detected in the human sample HTS 05 or the cattle isolate Cattle 03 (Table 1 and Fig. 1). The number of sets of unique reads of *C. parvum* Type B represented varying percentages of the total unique sets, ranging from 3.1% (HTS 03) to 50% (Cattle 02) (Fig. 1).

3.3. Sanger sequencing of the 18S and actin loci

Sanger sequencing of single-round PCR products showed good agreement with Ion Torrent-based genotyping at both loci (Table 1). Identifications based on Sanger sequencing also showed complete agreement between the two loci, for the 23 samples typed (human and animal).

Clean chromatograms were obtained for 23 of the 25 isolates typed at both loci. Sanger sequencing-based identifications were divided as follows: *C. hominis* (*n*=14), *C. parvum* (*n*=7), and *C. muris* (*n*=2). *Cryptosporidium* spp., identifications were not possible for two snake samples (SK 02 and SK 05), as mixed 18S and actin chromatograms were obtained due to non-specific amplification using these two primer sets. Sample SK 03, which was identified as a *C. muris* and *C. tyzzeri* mixed infection by Ion Torrent, was identified as *C. muris* only by Sanger, presumably due to a larger amount of *C. muris* template in the mixture dominating the PCR reaction. All cattle samples harboured *C. parvum*, which was also found in two human samples (HTS 03 and HTS 05). Of the sixteen human *Cryptosporidium*-positive faecal samples screened, *C. parvum* was detected in two samples (12.5%). Overall Sanger sequencing-based identifications were divided as follows: *C. hominis* (*n*=14), *C. parvum* (*n*=6) and *C. muris* (*n*=2) (Table 1). Unlike the result obtained

from Ion Torrent sequencing, Type B 18S rRNA sequences were not identified using a Sanger sequencing approach.

3.4. Comparative Costs of Sanger and HTS approaches when genotyping

Sanger was cheaper than HTS when sequencing small numbers of samples is required (Table 2). The cost of Sanger sequencing varies greatly depending on the format chosen (i.e., tubes, PCR strip tubes, or 96-well plates) and the gel-purification method; the cost per sample can range from AUD$10-$20 or AUD$3.00 to $10.00 if sequencing 60 samples in a 96 well plate format. For HTS, the cost varies depending on the level of coverage needed platform and library build method. Excluding labour, the cost for an Ion Torrent run, with the 400 bp chemistry and including 60 samples, varies from AUD$570 (314 chip) to AUD$820 (316 chip). The expected average number of reads per sample, however, is approximately 8,600 (314 chip) and 500,000 (316 chip) (Table 2). For labour (i.e., hands-on time), Ion Torrent is comparable to Sanger sequencing when a number of samples (*n*=60) is processed simultaneously (Table 2).

4.0 Discussion

In the present study, 25 *Cryptosporidium* positive faecal samples were assessed by molecular interrogation using a HTS Ion Torrent platform and a Sanger sequencing approach for single-round PCR's two loci, 18S rRNA and actin. Our results show good agreement between the two sequencing technologies (Table 1). The DNA sequences obtained from the Ion Torrent platform were largely congruent with the *Cryptosporidium* species identifications made from Sanger sequencing. One sample at the actin locus (Cattle 05) and three samples at

the 18S rRNA locus (HTS 10, HTS 11 and HTS 12) could not be typed by HTS due to lack of amplification of these isolates, presumably due to the fusion primers (included MID tags and adaptors) that impacted on PCR efficiency. However, Ion Torrent was able to type two isolates (SK 02 and SK 05), which could not be typed by Sanger sequencing but were identified by Ion Torrent (Table 1). One isolate (SK 03) was typed as *C. muris* by Sanger sequencing but was typed as a *C. muris/C. tyzzeri* mixture by Ion Torrent. Although computational deciphering of mixed Sanger chromatograms is practiced, these situations are usually resolved by cloning amplicons: a time-consuming and expensive procedure (Carr et al., 2009; Fantin et al., 2013; Paparini et al., 2013. The results from this study, supports HTS as a better means for the identification of mixed infections.

Sensitivity is an important issue when typing infectious agents from faecal or environmental samples, which may be present in low quantities. However, it is also important to note that performance of amplicon sequencing critically depends on PCR efficiency. For Sanger sequencing, a nested PCR approach with unmodified primers were used but for HTS, amplification required the original primers to be modified by the addition of MID-tags and adaptors. The resultant loss of PCR sensitivity and efficiency will need to be carefully monitored if HTS is to become more widespread in *Cryptosporidium* genotyping. Other library build methods, including ligation of adaptors may be required when input template levels become limiting.

At the actin locus, minor genetic variants of *C. hominis* and *C. parvum* sequences were found in virtually all samples (except only Cattle 5) by HTS. However, as the reported "genetic" error rate of the Ion Torrent chemistry is estimated to be 1.7% (Quail et al., 2012), genetic differences ≤ 2% might be the results of artefacts, and should be treated with caution. All Sanger sequencing chromatograms were characterised by narrow and high peaks, which consistently allowed reliable, automatic assignment of each base (i.e., software-aided). A

fraction of samples, however, also displayed a low level of background noise, which, did not affect the interpretation, but may be indicative of sequence variants.

Evidence shows that sequencing genomes or genomic regions with biased base composition, can still pose significant challenges with currently available HTS platforms (Oyola et al., 2012). Like other pathogens such as *Plasmodium falciparum* and *Mycobacterium tuberculosis*, *Cryptosporidium* displays a biased base composition (AT-rich) (Abrahamsen et al., 2004), which, during standard library preparation procedures, may cause problems and require optimized protocols (Oyola et al., 2012). Currently the 18S rRNA gene is the only locus that will amplify all species and genotypes of *Cryptosporidium* and is therefore the most widely used locus (Xiao, 2010). However, the hypervariable region of the 18S rRNA gene used to distinguish different species and genotypes contains homopolymer runs. In *C. hominis*, (along with a few other SNPs), there is a stretch of 7 to 11 thymine residues in the poly-T region within the amplicon sequenced in this study, that is important for differentiating it from the closely related species *C. parvum* (Power et al., 2011), whose genome is 95-97% identical to *C. hominis* (Abrahamsen et al., 2004; Xu et al., 2004). Ion Torrent, like 454 pyrosequencing, has a weakness in resolving homopolymer sequences and false INDELs are common (Meldrum et al., 2011; Quail et al., 2012). In this regard the single base incorporation afforded by Illumina sequencing-by-synthesis may be preferable. Despite this limitation identification at the 18S locus was still possible using Ion Torrent.

There are two structurally distinct types of rRNA unit in the haploid *Cryptosporidium* genome; Type A (4 copies) and Type B (1 copy), with marked differences in the internal transcribed spacer regions and differences in the 18S region, particularly for *C. parvum* (Le Blancq et al., 1997; Morgan et al., 1999; Xiao et al., 1999). With Sanger sequencing, the Type B sequence rarely appears (Ryan, unpublished observations), however with HTS, both Type A and Type B sequences will be detected. In the present study, Type B sequences were

detected in four of the five *C. parvum* cattle isolates and one of the two human *C. parvum* isolates using HTS (Table 1) but not by Sanger sequencing. This however necessitated using an in silico probe to identify Type B sequences (Type B sequences were not searched for in *C. hominis* isolates, as the Type A and B sequences are very similar) (Morgan et al., 1999). Why Type B sequences weren't detected in all *C. parvum* isolates is unclear but may have been due to the low number of overall reads obtained for most isolates at the 18S rRNA locus. It may also be that Type B sequences are not present in all isolates as previous Type B specific amplification suggests this may be the case (Morgan et al., 1999). It is important that Type B sequences be correctly identified when using HTS as otherwise they could be mis-identified as novel genotypes. This is particularly problematic as although Type B sequences are known for *C. hominis*, *C. parvum*, *C. fayeri* and *C. felis* (Le Blancq et al., 1997; Morgan et al., 1999; Xiao et al., 1999), little is known about Type B sequences in other species and genotypes. However further HTS analysis of a range of *Cryptosporidium* species will likely solve this issue.

A recent study, which analysed the single-copy 70 kDa heat shock protein (hsp70) and the 60 kDa surface glycoprotein (gp60) genes in two *C. parvum* isolates on a Illumina HiSeq instrument (San Diego, CA, USA), identified two HSP70 and 10 gp60 genotypes in these isolates (Grinberg et al., 2013). This study suggests a much higher rate of mixed infections in *Cryptosporidium* faecal samples than previously thought. In the present study, a mixed *C. muris/C. tyzzeri* infection was only identified in one sample (SK 03), however it is difficult to make comparisons when different platforms, samples and loci are employed. Future studies should include analysis of a wide range of *Cryptosporidium* isolates at the gp60 locus using the Illumina platform to confirm the data by Grinberg et al. (2013).

Comparing the costs of the sequencing is complicated by the ever-falling price of HTS. At the time of these experiments, the cost of Ion Torrent sequencing, while more

expensive for sequencing small numbers of samples, was comparable with Sanger for sequencing larger numbers (n=60) (Table 1). For Sanger sequencing, several variables affect the cost per sample (e.g., possibility of automation, choice of consumables, adoption of kits etc.), but hands-on time (labour) and possibility of adopting high-throughput protocols for the simultaneous processing of multiple samples are probably the most important factors to consider. For Ion Torrent, the cost is mainly determined by the platform and depth of coverage needed for a particular application. Based on Table 2, the cost per sample for Ion Torrent can vary from AUD$ 9.50 to $ 13.70, or ~ AUD$ 0.41 -1.43 for 1,000 reads, depending on the chip used.

The bioinformatics analysis of the Ion Torrent data generated was somewhat problematic, as existing bioinformatics pipelines did not yield useful data, mainly due to the lack of suitable curated databases. For example, simply querying sequences against the NCBI GenBank nucleotide data-base using BLASTn (Benson et al., 2006) and then importing the BLAST results obtained into MEtaGenome Analyzer v4 (MEGAN), where they were mapped and visualised against the NCBI taxonomic framework (Huson et al., 2007), only grouped the *Cryptosporidium* sequences to genus level. Another unsuccessful approach involved the use of the QIIME software package (Caporaso et al., 2010) for the analysis of 18S data, in combination with the Silva 104 and 108 releases (Quast et al., 2013). This was largely due to the abundance of *Cryptosporidium* sequences in GenBank that are not clearly identified to species or genotype level, and/or to the completeness of the databases used. As a result of this, an "in house" database of curated sequences was developed and Ion Torrent sequences were queried against this data.

The present study focused on the comparison of alternative techniques, to amplify and sequence the same loci (18S rRNA and actin), using the same single-round PCR primer sets (with the addition of MID tags and adaptors for Ion Torrent) on a range of *Cryptosporidium*-

positive samples. Our comparison proves the cost-effectiveness of Ion Torrent in comparison to Sanger sequencing, particularly for a sample set greater than ~60 samples. Although four of the samples analysed showed problematic amplification (most likely due to the loss of efficiency by the addition of MID tags and adaptors for Ion Torrent), the superior depth of coverage, offered by Ion Torrent, still represents a clear advantage over more traditional approaches, especially when mixed infections are present in a clinical sample. For example, isolate SK 03, was typed as *C. muris* only by Sanger sequencing but was identified as a *C. muris* and *C. tyzzeri* mixture by Ion Torrent sequencing. The higher sensitivity of Ion Torrent (and other types of HTS), to rare DNA species, requires the adoption of experimental strategies aimed at minimising cross-study DNA contamination. Other HTS technologies such as Illumina (Bentley et al., 2008) may be more suitable for genotyping *Cryptosporidium* isolates and future studies should compare *Cryptosporidium* isolate characterisation on a range of HTS platforms and at multiple loci including 18S, actin and the gp60 subtyping locus.

Acknowledgements

This study was financially supported by an Australian Research Council Linkage Grant number LP130100035. MB was supported by Australian Research Council Future Fellowship (FT0991741). Authors are grateful to Tim Hyndman for provision of snake samples and to Frances Brigg and Dave Berryman (State Agricultural Biotechnology Centre, Murdoch University, Australia) for support and discussions.

357 **References**

358

359 Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A.,
360 Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear,
361 P.H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., Kapur,
362 V., 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*.
363 Sci. 304, 441-5.
364 Amadi, B., Mwiya, M., Sianongo, S., Payne, L., Watuka, A., Katubulushi, M., Kelly, P.,
365 2009. High dose prolonged treatment with nitazoxanide is not effective for
366 cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial.
367 BMC Infect. Dis. 9, 195.
368 Baldursson, S., Karanis, P., 2011. Waterborne transmission of protozoan parasites: Review of
369 worldwide outbreaks - An update 2004-2010. Wat. Res. 45, 6603-6614.
370 Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G.,
371 Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. 2008. Accurate whole
372 human genome sequencing using reversible terminator chemistry. Nature. 456, 53-59.
373 Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. 2006. GenBank.
374 Nucleic Acids Res. 34 (Database issue), D16-20.
375 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K.,
376 Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T.,
377 Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D.,
378 Pirrung, M., Reeder, J., Sevinsky, J.R., Tumbaugh, P.J., Walters, W.A., Widmann, J.,
379 Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-
380 throughput community sequencing data. Nat. Methods 7, 335-336.

Carr, I.M., Robinson, J.I., Dimitriou, R., Markham, A.F., Morgan, A.W., Bonthron, D.T., 2009. Inferring relative proportions of DNA variants from sequencing electropherograms. Bioinformatics 25, 3244-3250.

Fantin, Y.S., Neverov, A.D., Favorov, A.V., Alvarez-Figueroa, M.V., Braslavskaya, S.I., Gordukova, M.A., Karandashova, I.V., Kuleshov, K.V., Myznikova, A.I., Polishchuk, M.S., Reshetov, D.A., Voiciehovskaya, Y.A., Mironov, A.A., Chulanov, V.P., 2013. Base-Calling Algorithm with Vocabulary (BCV) Method for Analyzing Population Sequencing Chromatograms. PLoS ONE 8, e54835.

Grinberg, A., Biggs, P.J., Dukkipati, V.S.R., George, T.T., 2013. Extensive intra-host genetic diversity uncovered in *Cryptosporidium parvum* using Next Generation Sequencing. Infect. Genet. Evol. 15, 18-24.

Le Blancq, S.M., Khramtsov, N.V., Zamani, F., Upton, S.J., Wu, T.W., 1997. Ribosomal RNA gene organization in *Cryptosporidium parvum*. Mol. Biochem. Parasitol. 90, 463-478.

Hunter, P.R., Hadfield, S.J., Wilkinson, D., Lake, I.R., Harrison, F.C.D., Chalmers, R.M., 2007. Subtypes of *Cryptosporidium parvum* in humans and disease risk. Emerg. Infect. Dis. 13**,** 82-88.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17(3), 377-386.

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

Mayer, C.L., Palmer, C.J., 1996. Evaluation of PCR, nested PCR, and fluorescent antibodies for detection of *Giardia* and *Cryptosporidium* species in wastewater. Appl. Environ. Microbiol. 62, 2081-2085.

405 Metzker, M., 2010. Applications of next-generation sequencing. Sequencing technologies -
406 the next generation. Nat. Rev. Genet. 11, 31 - 46.
407 Meldrum, C., Doyle, D.A., Tothill, R.W. 2011. Next-Generation Sequencing for Cancer
408 Diagnostics: a Practical Perspective. Clin. Biochem. Rev. 32(4), 177-195.
409 Morgan, U.M., Constantine, C.C., Forbes, D.A., Thompson, R.C.A., 1997. Differentiation
410 between human and animal isolates of *Cryptosporidium parvum* using rDNA
411 sequencing and direct PCR analysis. J. Parasitol. 83, 825-830.
412 Morgan, U.M., Deplazes, P., Forbes, D.A., Spano, F., Hertzberg, H., Sargent, K.D., Elliot,
413 A., Thompson, R.C., 1999. Sequence and PCR-RFLP analysis of the internal
414 transcribed spacers of the rDNA repeat unit in isolates of *Cryptosporidium* from
415 different hosts. Parasitol. 118, 49-58.
416 Oyola, S.O., Otto, T.D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D.J.,
417 MacInnis, B., Kwiatkowski, D.P., Swerdlow, H.P., Quail, M.A., 2012. Optimizing
418 illumina next-generation sequencing library preparation for extremely at-biased
419 genomes. BMC Genomics 13, 1.
420 Paparini, A., Jackson, B., Ward, S., Young, S., Ryan, U.M., 2012. Multiple *Cryptosporidium*
421 genotypes detected in wild black rats (*Rattus rattus*) from northern Australia. Exp.
422 Parasitol. 131, 404-412.
423 Park, J.W., Crowley, D.E., 2010. Nested PCR bias: a case study of *Pseudomonas* spp. in soil
424 microcosms. J. Environ. Monit. 12(4), 985-8.
425 Power, M.L., Holley, M., Ryan, U.M., Worden, P., Gillings, M.R., 2011. Identification and
426 differentiation of *Cryptosporidium* species by capillary electrophoresis single-strand
427 conformation polymorphism. FEMS Microbiol. Lett. 314, 34-41.
428 Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A.,
429 Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms:

comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 13, 341.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene data-base project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590-D596.

Ryan, U.M., Fayer R., Xiao, L., 2014. *Cryptosporidium* species in humans and animals: current understanding and research needs. Parasitol. 141, 1667-1685.

Xiao, L.H., 2010. Molecular epidemiology of cryptosporidiosis: an update. Exp. Parasitol. 124, 80-89.

Xiao, L., Limor, J.R., Li, L., Morgan, U., Thompson, R.C., Lal, A.A., 1999. Presence of heterogeneous copies of the small subunit rRNA gene in *Cryptosporidium parvum* human and marsupial genotypes and *Cryptosporidium felis*. J. Eukaryot. Microbiol. 46, 44S-45S.

Yang, R., Paparini, A., Monis, P., Ryan, U., 2014. Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of *Cryptosporidium* oocysts in faecal samples. Int. J. Parasitol. In press

Xu, P., Widmer, G., Wang, Y., Ozaki, LS., Alves, J.M., Serrano, M.G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A.J., Pearson, W.R., Dear, P.H., Bankier, A.T., Peterson, D.L., Abrahamsen, M.S., Kapur, V., Tzipori, S., Buck, G.A., 2004. The genome of *Cryptosporidium hominis*. Nature. 431, 1107-1112.

Fig. 1. Percentage of 18S rRNA Type B sequences in *C. parvum* isolates analysed by Ion Torrent. n=the number of total reads used in the calculations.

Table 1. Comparison of Ion Torrent and Sanger sequencing of 25 animal and human samples analysed at the 18S rRNA (type A and/or type B) and actin loci. The number of sets of unique and total reads (in parenthesis), obtained by Ion Torrent from each sample is shown. Single-copy reads (singletons) were discarded and therefore unique reads from any given sample are two or more reads. For example, for HTS 05, there were 107 total reads (including singletons); when singletons were removed, 51 reads forming 10 unique sets (*C. parvum* A) were obtained.

		Ion Torrent							
Locus	**Sample**	***C. hominis***	***C. parvum* A (18S only)**	***C. parvum* B (18S only)**	**Other *Cryptosporidium* spp.**	**Non-specific**	**Total unique (excl. redundant and singletons)**	**Total (incl. redundant and singletons)**	**Sanger sequencing**
18S	HTS 03		95 (772)	3 (7)			98	1,252	*C. parvum*
	HTS 04	35					35	359	*C. hominis*
	HTS 05		10 (51)				10	107	*C. parvum*
	HTS 06	24					24	205	*C. hominis*
	HTS 07	2					2	39	*C. hominis*
	HTS 08	56					56	660	*C. hominis*
	HTS 09	2					2	33	*C. hominis*
	HTS 10						0	0	*C. hominis*
	HTS 11						0	0	*C. hominis*
	HTS 12						0	0	*C. hominis*
	HTS 15	8					8	51	*C. hominis*
	HTS 16	21					21	135	*C. hominis*
	HTS 17	1					1	24	*C. hominis*
	HTS 18	5					5	59	*C. hominis*
	HTS 19	50					50	491	*C. hominis*
	HTS 21	154					154	1,793	*C. hominis*
	Cattle 01		6 (20)	5 (14)			11	92	*C. parvum*
	Cattle 02		2 (4)	2 (4)			4	36	*C. parvum*
	Cattle 03		2 (5)				2	41	*C. parvum*
	Cattle 04		8 (33)	3 (10)			11	101	*C. parvum*
	Cattle 05		16 (86)	7 (20)			23	281	*C. parvum*
	SK 02				28 (*C. muris*)		28	918	*Mixture*
	SK 03				2 (*C. muris* + *C.*		2	7	*C. muris*

				tyzzeri)				
	SK 04			84 (*C. muris*)		84	2,306	*C. muris*
	SK 05			2 (*C. muris*)		2	24	*Mixture*
	HTS 03		85			85	2,231	*C. parvum*
	HTS 04	116				116	3,738	*C. hominis*
	HTS 05		57		2	59	1,109	*C. parvum*
	HTS 06	151				151	5,490	*C. hominis*
	HTS 07	138				138	4,779	*C. hominis*
	HTS 08	409				409	18,741	*C. hominis*
	HTS 09	261				261	10,310	*C. hominis*
	HTS 10	161				161	5,888	*C. hominis*
	HTS 11	313				313	12,716	*C. hominis*
	HTS 12	92				92	2,634	*C. hominis*
	HTS 15	99				99	3,885	*C. hominis*
	HTS 16	132			1	132	5,460	*C. hominis*
	HTS 17	50			3	54	2,141	*C. hominis*
	HTS 18	197			4	201	9,353	*C. hominis*
	HTS 19	116				116	4,999	*C. hominis*
Actin	HTS 21	35		n.a.	1	36	903	*C. hominis*
	Cattle 01		235			235	4,739	*C. parvum*
	Cattle 02		122			122	2,304	*C. parvum*
	Cattle 03		72			72	1,286	*C. parvum*
	Cattle 04		514		36	550	11,287	C. parvum
	Cattle 05					0	0	*C. parvum*
	SK 02			1 (*C. muris*)	6	7	175	*Mixture*
	SK 03			3 (1x *C. muris; 2x C. tyzzeri*)	5	8	89	*C. muris*
	SK 04			19 (*C. muris*)		19	532	*C. muris*
	SK 05			1 (*C. muris*)	4	5	54	*Mixture*

Table 2. Comparison of the costs of Sanger sequencing versus Ion Torrent sequencing. All costs are in AUD dollars.

Method	Number of samples	Total turn-around time (h)	Approx. cost (excl. DNA extraction)	Labour time (excl. DNA extraction)	Turn-around time for data analysis (including deconvolution and/or phylogenetic reconstruction)	Notes
Sanger sequencing	1+ (tubes)	15	$10.00-$20.00 per sample	2.5	1	Identification based on BLAST-searches and phylogenetic reconstruction
Sanger sequencing	60 (Plate)	15	$3.00-$10.00 per sample	6	3	
Ion Torrent (314 chip v2)	60	25	$570.00 per run	5.5	4.0	Number of reads/chip: 400-550 thousand

	Ion Torrent (316 chip v2)	60	25	$820.00 per run	5.5	6.0	Number of reads/chip: 2-3 million