



## Invited Review

# *Cryptosporidium* within-host genetic diversity: systematic bibliographical search and narrative overview

Alex Grinberg<sup>a,\*</sup>, Giovanni Widmer<sup>b</sup><sup>a</sup> Massey University, Institute of Veterinary, Animal and Biomedical Sciences, Infectious Diseases Group, Private Bag 11 222, Palmerston North, New Zealand<sup>b</sup> Tufts University, Cummings School of Veterinary Medicine, Department of Infectious Disease and Global Health, 200 Westboro Road, North Grafton, MA 01536, USA

## ARTICLE INFO

## Article history:

Received 19 January 2016

Received in revised form 29 February 2016

Accepted 6 March 2016

Available online 25 March 2016

## Keywords:

*Cryptosporidium*

Protozoa

Genetic diversity

Genotyping

Genomics

Population genetics

Recombination

## ABSTRACT

Knowledge of the within-host genetic diversity of a pathogen often has broad implications for disease management. *Cryptosporidium* protozoan parasites are among the most common causative agents of infectious diarrhoea. Current limitations of in vitro culture impose the use of uncultured isolates obtained directly from the hosts as operational units of *Cryptosporidium* genotyping. The validity of this practice is centred on the assumption of genetic homogeneity of the parasite within the host, and genetic studies often take little account of the within-host genetic diversity of *Cryptosporidium*. Yet, theory and experimental evidence contemplate genetic diversity of *Cryptosporidium* at the within-host scale, but this diversity is not easily identified by genotyping methods ill-suited for the resolution of DNA mixtures. We performed a systematic bibliographical search of the occurrence of within-host genetic diversity of *Cryptosporidium* parasites in epidemiological samples, between 2005 and 2015. Our results indicate that genetic diversity at the within-host scale, in the form of mixed species or intra-species diversity, has been identified in a large number ( $n = 55$ ) of epidemiological surveys of cryptosporidiosis in variable proportions, but has often been treated as a secondary finding and not analysed. As in malaria, there are indications that the scale of this diversity varies between geographical regions, perhaps depending on the prevailing transmission pathways. These results provide a significant knowledge base from which to draw alternative population genetic structure models, some of which are discussed in this paper.

© 2016 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

In many infectious diseases there is a presence of distinct microbial lineages within the host, and the within-host genetic diversity of the pathogen is often associated with its ability to adapt to selective pressures exerted during the infection. For instance, in hepatitis C and HIV infections the presence of viral sub-populations enhances evasion from the immune response and resistance to chemotherapy (Wolinsky et al., 1996; McMichael and Phillips, 1997; Farci et al., 2000; Briones et al., 2006). It has been reported that approximately 20% of people infected with *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa, harboured genetically heterogeneous infections, and the presence of microbial heterogeneity was associated with increased odds of treatment failure (Cohen et al., 2016). Examples in the protozoan world also abound. Polyclonal infections with *Eimeria tenella* (poultry) and *Theileria annulata* (cattle) seem to be common (Weir et al., 2011; Blake et al., 2015), and mixed infections with *Plasmodium*

*falciparum* clones were already observed in people during the 1990s (Viriyakosol et al., 1995). Furthermore, in some regions most malaria patients harbour multiple clones (Arnot, 1998; Apinjoh et al., 2015). Other studies have found associations between the within-host genetic diversity of *Plasmodium* spp. and the natural history of malaria, and it is currently recognised that this diversity may influence the evolution of virulence, hamper chemotherapeutic control and potentially promote the emergence of vaccine escape variants (de Roode et al., 2004; Kwiek et al., 2007; Juliano et al., 2010a,b; Tyagi et al., 2013). Thus, knowledge of the genetic diversity of a pathogen within the host may have broad implications for disease management.

*Cryptosporidium* is a genus of sexually reproducing protozoan parasites of amphibians, fish, reptiles, birds and mammals. Intestinal species are major contributors to the burden of infant diarrhoea in many developing countries (Kotloff et al., 2013). The species *Cryptosporidium parvum* and *Cryptosporidium hominis*, in particular, are among the most common and cosmopolitan causes of diarrhoea in people, and *C. parvum* is also a leading cause of diarrhoea in young farmed ruminants (Al Mawly et al., 2015). Despite the availability of the sequences of the *C. parvum* and *C. hominis*

\* Corresponding author. Tel.: +64 6 3569099; fax: +64 6 3505635.

E-mail address: [a.grinberg@massey.ac.nz](mailto:a.grinberg@massey.ac.nz) (A. Grinberg).

genomes (Abrahamsen et al., 2004; Xu et al., 2004), our understanding of cryptosporidiosis is still incomplete. Efficient anti-*Cryptosporidium* drugs and vaccines are lacking and there is poor knowledge on the role of genetic exchange in the evolution of the parasites' virulence. One of the main reasons for the poor progress in many fields is the limitations of in vitro culture. Despite recent progress (Morada et al., 2015), methods for *Cryptosporidium* culture are still inefficient and clonal lineages derived from individual sporozoites, which are the basic cells carrying the parasite's haploid genome, are not available. As a consequence, *Cryptosporidium* strains remain loosely defined (Cama et al., 2006a) and most research is currently performed using isolates composed of parasites obtained from single infected hosts, without previous axenic culture. The use of uncultured isolates as operational units of *Cryptosporidium* genotyping is also an established practice (Ryan et al., 2014), the working assumption being that a broadly phylogenetically homogeneous parasite infects the host. However, as in malaria, where individuals can harbour several genetically distinct parasites as a result of multiple mosquito bites, or single bites from mosquitoes bearing multiple clones (Talisuna et al., 2007), *Cryptosporidium* within-host diversity may originate from multiple exposures, or a single exposure to genetically diverse oocysts. In fact, studies have identified water as a major route of *Cryptosporidium* transmission in some areas (Xiao et al., 2004), and it has been demonstrated that water sources can contain multiple *Cryptosporidium* spp. originating from disparate locations or hosts (Ruecker et al., 2005, 2013; Nichols et al., 2006; Feng et al., 2009). Thus, one would expect to find heterogeneous parasites within hosts exposed to such sources, as well as in secondary infections.

There is long-standing evidence for the existence of within-host genetic diversity in *Cryptosporidium* (Widmer, 1998; Tanriverdi et al., 2003, 2008). However, most epidemiological studies of cryptosporidiosis used PCR-based genotyping methods able to resolve the dominant sequence but insensitive to minority variants and ill-suited for resolution of complex DNA mixtures (Suzuki and Giovannoni, 1996; Rochelle et al., 2000; Reed et al., 2002; Liu et al., 2008; Papparini et al., 2015). Furthermore, when identified in the field, the *Cryptosporidium* within-host diversity has been often treated as a marginal finding, and only a few groups have pursued further analyses at the within-isolate scale (Cama et al., 2006b; Jeníková et al., 2011; Shrestha et al., 2014). As a consequence, the extent of within-host *Cryptosporidium* diversity in nature remains unknown.

There are practicalities in working with the assumption of a within-host genetic homogeneity of *Cryptosporidium*, and genotyping of uncultured isolates has significantly aided our understanding of the epidemiology of cryptosporidiosis. Nevertheless, from the study of other pathogens we learn that genetic analyses at the within-host level could enhance our ability to answer many unresolved questions on cryptosporidiosis. Here, we present the results of a systematic bibliographical search of the epidemiological evidence for *Cryptosporidium* within-host genetic diversity. We also assemble current understandings of aspects of the parasites' life cycle pertaining to genetic exchange, to hypothesise alternative population genetic structures that challenge the assumption of within-host genetic homogeneity. Some gaps between actual and desired genotyping practices for the study of the genetic structure of *Cryptosporidium* populations are also addressed.

## 2. Within-host *Cryptosporidium* genetic diversity: epidemiological evidence

We aimed to estimate the extent of *Cryptosporidium* within-host genetic diversity in nature based on information published

in the scientific literature. We noted that in many papers the genetic heterogeneity revealed in the isolates was not the main focus of the study, and when identified, it was often defined as 'mixed infection'. Therefore, we interrogated the PubMed database (<http://www.ncbi.nlm.nih.gov/PubMed>) using the descriptor '*Cryptosporidium*' (all fields); and 'mixed' (all fields). Filters applied were 'species' (values: 'humans'; and 'other animals'); 'text availability' (value: 'abstract'); and 'languages' (value: 'English'). The search was limited to the decade of 2005–2015. This strategy returned all the papers in English, containing the terms '*Cryptosporidium*' and 'mixed', referring to infections in humans and animals for which an abstract was available in PubMed, published between 2005 and 2015. The database was accessed repeatedly from 31 August 2014. Relevant papers retrieved by us elsewhere were also used.

The articles were downloaded and assessed for the type of intra-isolate diversity reported (mixed-species/intra-species diversity). Each article was scored by its evidentiary value as follows: Evidentiary Score 1 (ES1), studies in which spurious intra-isolate genetic diversity was unlikely by virtue of the type of genotyping method used, or in which the possibility of spurious diversity was controlled; ES2, studies in which spurious intra-isolate genetic diversity was possible, and this was acknowledged or discussed by the authors; ES3, studies in which spurious intra-isolate genetic diversity was possible, but it was not discussed.

Of 144 articles initially retrieved, 89 reported mixed infections with *Cryptosporidium* and other organisms, or analysed *Cryptosporidium* mixtures in water, or involved in vitro experiments and were discarded. A total of 55 articles reported intra-isolate genetic diversity, worldwide. Twelve articles reported intra-isolate genetic diversity in humans and 41 in animal species. Two articles reported it in both humans and animals. Forty articles reported mixed species and 13 intra-species diversity, whilst two reported both occurrences. In many articles the diversity was observed in a significant proportion of isolates, and in one it was reported in 42% of the isolates. A total of 26/55 (47%) articles received an ES1 score, five received ES2 and 22 received ES3. Two articles could not be scored due to missing information. The details of the surveyed articles are presented in [Supplementary Table S1](#).

## 3. Origins of *Cryptosporidium* within-host genetic diversity

The bibliographic search indicated a substantial number of epidemiological studies reporting *Cryptosporidium* intra-isolate genetic diversity. Hence, to establish the origins of this diversity it is necessary to review aspects of the parasites' life cycle pertinent to genetic exchange. We note as a limitation, that much of what we assume today about this life cycle is based on inferences from electron microscopy studies performed before the species could be differentiated by genotyping (Tzipori, 1986; Tzipori and Griffiths, 1998; Tzipori and Ward, 2002).

Infections with *C. parvum* (or as a generalisation, with any intestinal *Cryptosporidium* sp.) are acquired through the ingestion of oocysts, and the life cycle culminates with excretion in the faeces of millions of highly resistant oocysts containing four haploid sporozoites. With some exceptions (Karanis et al., 2008; Koh et al., 2013), it is generally assumed that *Cryptosporidium* only replicates inside the host. In the gut, sporozoites excyst from the oocyst, invade the host cells and undergo subsequent rounds of conservative cell divisions in an intracellular but extracytoplasmic position within the enterocytes, followed by differentiation into micro- or macrogametocytes and mating of gametes, resulting in the formation of a transient diploid zygote. It is not known whether gamete dimorphism is determined by epigenetic regula-

tion or is genetically pre-determined. This is a key distinction: a sporozoite bearing the genome of one side of the sexual duality would be unable to independently complete the life cycle or continuously grow in vitro. Thereafter, the zygote undergoes a reduction division (meiosis) with the generation of two haploid sporozoites, and it is widely accepted that chromosomal reassortment and crossing-over may take place during this division. This view is consistent with observations suggesting that recombinant parasite progeny can be generated in co-infections with heterogeneous isolates in laboratory mice (Tanriverdi et al., 2007). Then, each cell undergoes a single conservative division generating two pairs of haploid sporozoites, while the oocyst wall forms. In the case of heterogamy (i.e., mating of genetically different gametes), the resulting oocyst will contain two pairs of genetically different sporozoites, each of which could differ from the parental genotypes due to recombination. Thin and thick-walled oocysts have been observed using electron microscopy, and it is believed that the thin-walled ones rupture in the gut and provide further rounds of life cycle (autoinfection), increasing recombinatorial diversification. However, it is not known whether thin-walled oocysts are produced in every infection. This life cycle predicts that the isolates, which may differ genetically between the hosts, can also display sporozoite diversity at the within-host level as follows:

- (i) The host excretes a population of parasites (the isolate) for which genetic diversity stems from the initial diversity of the infecting oocysts, the mutations and recombinations accumulating in the gut through billions of cell divisions, and the loss of diversity occurring during the infection, representable as:

$$D_{(\text{tot})ith} = D_{(\text{ing})ith} + D_{(\text{mut})ith} - D_{(\text{lost})ith}$$

where  $D_{(\text{tot})ith}$  is the number of alleles at the  $ith$  locus present in the isolate;  $D_{(\text{ing})ith}$  is the number of alleles ingested with the infection dose;  $D_{(\text{mut})ith}$  is the number of new alleles generated through mutations during the infection; and  $D_{(\text{lost})ith}$  is the number of alleles lost during the infection due to natural selection and genetic drift.

- (ii) According to the principle of reproductive isolation, which predicts mating incompatibility between sexually reproducing species (Mayr, 1942), co-infecting species are not expected to cross-mate and each species will generate distinct oocysts.
- (iii) Conversely, the same principle predicts a lack of, or marginal, mating incompatibility between variants within the species. Hence, two gametes from the same species encountering in an intestinal cell will be able to mate and produce an oocyst. In the case of homogamy (synonym: selfing, i.e. mating of identical gametes), the oocyst will contain four virtually identical sporozoites, which will be identical to the parental gametes. Conversely, in heterogamy it will contain two distinct pairs of sporozoites that could also differ from the parental cells due to recombinatorial changes (see Fig. 1, host C). Sexually reproducing populations characterised by unconstrained random mating (homogamous and/or heterogamous) are said to be ‘panmictic’. However, some authors consider sexually reproducing populations dominated by homogamy as ‘clonal’ (Tibayrenc and Ayala, 2014).
- (iv) As a derivation, *Cryptosporidium* within-host genetic diversity may occur both within and between the oocysts (Fig. 1, hosts B and C).

The available evidence supports a population structure characterised by genetic variation within and between the oocysts. In

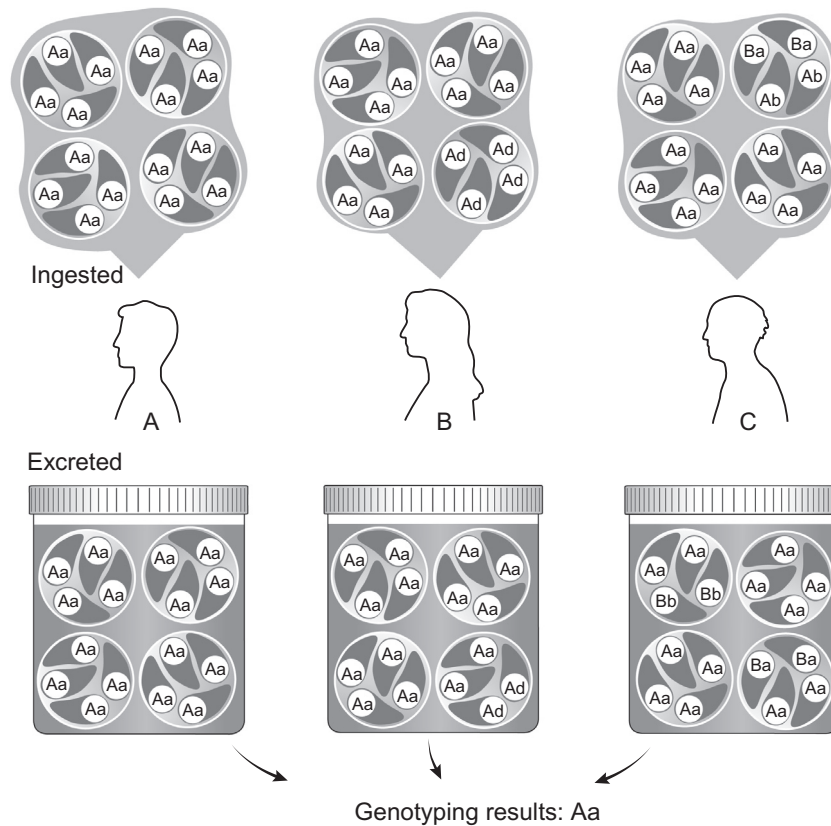
particular, *Cryptosporidium* intra-species mating incompatibility has not been documented, and studies in mice have suggested that recombinant parasite progeny can be generated in infections with heterogeneous isolates (Tanriverdi et al., 2007). However, reproductive isolation of the species has not been studied and many reproductive aspects remain theoretical and unverified.

#### 4. Is genetic homogeneity of *Cryptosporidium* isolates a self-fulfilling prophecy?

Species are arguably the most recognisable taxonomic units, and there is a consensus that *Cryptosporidium* comprises multiple species and ‘genotypes’ (Ryan et al., 2014) (the latter are taxa for which knowledge is incomplete, and here we treat them as species for simplicity). However, there is no clarity on what constitutes a *Cryptosporidium* sp. and the criteria for species delineation are still a matter of debate (Slapeta, 2006). Hence, various descriptors such as ‘valid’, ‘accepted’ or ‘named species’ have been used by different groups (Xiao and Feng, 2008; Ryan et al., 2014; Thompson and Ash, 2015). The main reasons for this lack of clarity are the absence of morphological traits enabling the differentiation of multiple species (Xiao et al., 2004) and, perhaps more importantly, the limitations of culture and lack of clonal lineages allowing experimental assessment of biological species traits, in particular reproductive isolation. The species conundrum is not unique to *Cryptosporidium*, but rather common in asexually reproductive protists (Schlegel and Meisterfeld, 2003), and in some microorganisms the species are still delineated based on convenience criteria.

With the lack of defining morphological and biological traits, PCR-based genotyping of uncultured isolates has long been the sole method used for identification of *Cryptosporidium* spp. in the field, and the main method for new species characterisation. Several taxonomically informative loci conserved within each species, and which are polymorphic between species, have been used for species identification, usually using PCR of uncloned amplicons followed by Sanger sequencing or restriction fragment length polymorphism analysis (RFLP). The inclusion of the 18S rDNA in *Cryptosporidium* spp. identification schemes is common practice, and the locus is often among the first to be characterised when new variants are found. Thus, public repositories contain a large number of 18S rDNA sequences, underpinning comparative analyses. Other coding or non-coding regions are also used for species assignment (Carraway et al., 1996, 1997; Peng et al., 1997; Spano et al., 1998; Caccio et al., 1999; Tanriverdi et al., 2003), but often these sequences are known for a limited number of taxa. PCR protocols which use a single pair of primers, or nested PCR protocols are commonly used, but it has been repeatedly shown that these assays tend to resolve the dominant allele and are ill-suited for the identification of minority variants (Reed et al., 2002; Grinberg et al., 2013; Papparini et al., 2015) (in unpublished experiments we found that a sequence representing less than 25% of the DNA remained undetected by RFLP). In contrast to PCR with a conserved pair of primers, Tanriverdi et al. (2003) used two species-specific primers to show that minority variants representing as little as 0.01% of the population could be detected.

Several groups have been able to resolve more than one species in single isolates using ad hoc genotyping approaches. For example, using the dihydrofolate reductase and the *Cryptosporidium* oocyst wall protein (COWP) genes in a second round of genotyping, Cama and co-workers (2006b) identified mixed infections in 7/21 human patients initially diagnosed with *Cryptosporidium canis* or *Cryptosporidium felis* mono-infections. Elwin and Chalmers (2008) reported *Cryptosporidium* isolates in sheep showing 18S rDNA RFLP banding patterns of *Cryptosporidium bovis* and COWP patterns of *Cryptosporidium* ‘cervine genotype’. Most recently, the combined



**Fig. 1.** The cartoon represents three hypothetical infections with a panmictic *Cryptosporidium* sp. acquired from a single source containing genetically heterogeneous parasites. Hosts A, B and C ingest infective doses represented by four oocysts, each containing two pairs of genetically identical haploid sporozoites. Within the sporozoites, letters represent allelic variants at two unlinked loci. The alleles of locus 1 are represented in uppercase letters and of locus 2 in lowercase. As expected from any sampling process, the predominant haplotype (Aa) is ingested by the three hosts, but there are differences in minority variant composition between the infective doses of the hosts. Host A ingests a genetically homogeneous infective dose composed of sporozoites 'Aa' (allele 'A' at locus 1 and 'a' at locus 2). In this host, all gamete matings are homogamous and the excreted parasite is genetically homogeneous. Host B ingests a heterogeneous infective dose composed of two types of oocysts: 3/4 of the oocysts carry the predominant haplotype Aa and 1/4 a minority haplotype Ad. As a result of heterogamy, 1/4 of the oocysts obtained from this host display within-oocyst heterogeneity (oocysts containing haplotypes Aa and Ad). Host C ingests a heterogeneous dose composed of the predominant haplotype, Aa, and the minority haplotypes Ba and Ab. In the gut of host C, recombination between variants generated a new genome ('Bb'), which was not present at the source. *Note:* PCR-based genotyping identifies the predominant haplotype (Aa) for the three isolates, leading to the conclusion of a monoclonal, rather than panmictic, population.

use of PCR sequencing of the 18S rDNA, heat shock protein 70 and actin genes facilitated the detection of the three common endemic *Cryptosporidium* spp. of cattle (*C. parvum*, *C. bovis* and *Cryptosporidium andersoni*) in a bovine faecal sample (Shrestha et al., 2014). These cases are commonly assumed to represent 'mixed species' infections, where each species is reproductively isolated and generates distinct oocysts in the gut (although this assumption remains unverified). To better characterise these mixed-species infections, it could be possible to genotype individual oocysts or sporozoites, but such work is technically challenging and has rarely been reported (Tanriverdi et al., 2002; Ikarashi et al., 2013). An alternative is cloning of PCR amplicons and sequencing of multiple plasmids. This application has enabled to identify intra-isolate sequence heterogeneity at the rRNA internal transcribed spacer locus of the microsporidian *Enterocytozoon bieneusi* (Widmer et al., 2013). Although the procedure is technically straightforward, it requires careful primer design in order to cover a wide range of species. Furthermore, significant time and labour requirements hinder the application of amplicon cloning to routine genotyping of large numbers of field isolates, so our ability to detect and characterise mixed species infection remains limited.

If discrimination of variants within a species is required, genotyping at loci displaying intra-species polymorphism (i.e., sub-genotyping) can also be used. Length polymorphism analysis of micro- and minisatellite regions spanning variable numbers of

repeat units found throughout the genome was already applied for sub-genotyping of isolates in the late 1990s (Feng et al., 2000). During the last decade, PCR sequencing of the polymorphic repeat region of the 60 kDa zoite glycoprotein gene (the 'gp60' gene) has been widely used and many authors now refer to *C. parvum* and *C. hominis* gp60 variants as 'subtypes', effectively extending the *Cryptosporidium* taxonomy below the species level (Xiao et al., 2004; Ryan et al., 2014). Gp60 sub-genotyping relies on PCR and suffers from the same limitation in detecting intra-isolate diversity as do the methods for species identification. Cryptic intra-species gp60 heterogeneity has been revealed in *C. parvum* and *C. hominis* isolates by amplicon cloning (Waldron and Power, 2011; Couto et al., 2013; Grinberg et al., 2013; Ramo et al., 2014) and next generation sequencing (Grinberg et al., 2013; Guo et al., 2015). One study using next generation sequencing has identified extensive within-host gp60 diversity in *C. parvum*, indicating the isolates may represent highly heterogeneous populations, rather than discrete subtypes (Grinberg et al., 2013). Moreover, the definition of gp60 subtypes does not take into account the effect of genetic recombination within the species and that unlinked loci, including gp60, might occur in different combinations (Widmer and Lee, 2010; Fig. 1, hosts B and C). Intra-species diversity of *C. parvum*, *C. hominis* and *Cryptosporidium meleagridis* isolates was also observed by next generation sequencing of an imperfect simple sequence repeat marker (Widmer et al., 2015).

Several multilocus micro/minisatellite sub-genotyping schemes have also been used. In particular, multilocus sub-genotyping has been extensively used to generate hypotheses about *C. parvum* and *C. hominis* being ‘clonal’ or ‘panmictic’ via linkage disequilibrium analysis (LD), with conflicting results (Mallon et al., 2003; Morrison et al., 2008; Tanriverdi et al., 2008; De Waele et al., 2013; Caccio et al., 2015; Ramo et al., 2016). However, *Cryptosporidium* haplotypes obtained by PCR may not be suitable for LD analysis. In fact, the aim of LD analysis is to ascertain to what extent cross-mating breaks up associations between loci, but PCR is not suited to the determination of the mating patterns in the gut, as it often misses minority variants (Fig. 1).

Multi-locus schemes provide greater discriminatory power compared with single locus schemes, but increasing the number of markers also increases the chance of genotyping error due to PCR artifacts or allelic dropouts (Gagneux et al., 1997; Anderson et al., 2000; Greenhouse et al., 2007). The use of multi-locus sub-genotyping schemes underpinned detection of intra-species within-host genetic diversity in several studies (Mallon et al., 2003; Morrison et al., 2008; Tanriverdi et al., 2008; Caccio et al., 2015; Ramo et al., 2016). In a multinational study using multilocus sub-genotyping, intra-species diversity was observed in 18.2% of *C. hominis* and 11% *C. parvum* isolates from Uganda, but only in 2% of *C. hominis* from the UK (Tanriverdi et al., 2008). This pattern is reminiscent of the increased complexity of infection observed in malaria in regions of high transmission, attributable to multiple exposures (Greenhouse et al., 2007). Unlike the case of mixed species infections, it is assumed that intra-species variants may recombine in the gut, with an accumulation of diversity both between and within the oocysts (Fig. 1, hosts B and C).

Despite the long-standing evidence for occurrence of mixed species and intra-species within-host diversity in nature, most commonly used end-point genotyping tools resolve the dominant sequence present, self-fulfilling the expectation of within-host homogeneity of *Cryptosporidium*. Hence, it is likely that the parasites’ within-host genetic diversity is currently largely overlooked.

## 5. Discussion

For the last two decades, uncultured isolates composed of oocysts obtained from the infected hosts have been used as operational units of *Cryptosporidium* genotyping. This pragmatic practice, spanning from the limitations of culture, has played an important role and enhanced *Cryptosporidium* source tracking on several fronts. However, the practice is hard to reconcile with the fact that different species or variants can infect a host. Evidence for *Cryptosporidium* within-host genetic diversity has been available since the 1990s but it has been often reported in the field as a secondary finding. Therefore, we undertook an ad hoc bibliographical search to ascertain to what extent this diversity has been reported in epidemiological surveys.

Our search indicates that within-host genetic diversity has been identified in at least 55 studies in humans and animals between 2005 and 2015. This is a substantial number of studies if we consider that the commonly used end-point genotyping methods have been shown to miss minority variants in many cases. Most studies were published between 2010 and 2015 (data not shown), evidence of an increased awareness of this occurrence and a widening access to better genotyping platforms. This pattern is similar to the increasing number of reports of multiplicity of infection in malaria during the last decade. One of the cautions that must be considered is the presence of genotyping artifacts that could have led to over-estimation of the diversity. Approximately 40% of the studies in which genotyping error remained uncontrolled received an ES3 score. Yet, we determined that in half of the studies the

within-host genetic diversity could not be discounted on the basis of error, and these studies received an ES1 score. In these studies, techniques applied to reduce or control genotyping error included repetition of the assays, use of species-specific primers, verification of the results by Sanger sequencing or DNA cloning, iterative genotyping at multiple loci or other strategies. A thorough discussion of genotyping error is beyond the scope of this paper, but perhaps the most worrisome aspect of amplicon-based analyses of *Cryptosporidium* is the phenomenon of polymerase slippage during the amplification of short repeats. Slippage might generate multiple types of amplicon, which are readily detectable by fragment analysis or cloning. Our scoring system was subjective and we provide the PubMed identification numbers of the papers in [Supplementary Table S1](#), so readers can assess those using other criteria.

The presence of cryptic within-host genetic heterogeneity of a pathogen may bias the results of studies, in particular genomic analyses and antimicrobial drug assessments (Gagneux et al., 1997; Kwiek et al., 2007; Juliano et al., 2010a; Cohen et al., 2016; Mideo et al., 2016). Failure to recognise within-host genetic diversity might also impact *Cryptosporidium* research. As previously indicated, overlooking this diversity may bias the conclusions of LD analysis (Anderson et al., 2000; Pompanon et al., 2005). The use of LD analysis to define clonality or panmixia in sexually reproducing microparasites is, in general, controversial (Ramírez and Llewellyn, 2014). One angle of this controversy is simplified in Fig. 1: a hypothetical *Cryptosporidium* parasite diversifying by unconstrained recombination (hence, panmictic) appears monoclinal by PCR-based haplotype analysis.

Failure to identify minority variants could also impact anti-*Cryptosporidium* drug discovery studies. In malaria, for instance, the occurrence of cryptic resistant variants in infections consisting of predominantly drug-sensitive parasites is a recognised reason for unexplained chemotherapy failure (Djimé et al., 2001; Kwiek et al., 2007; Tyagi et al., 2013). Similarly, cryptic within-host diversity could affect vaccine research. *Cryptosporidium* vaccine candidate antigens have been recently reviewed, with emphasis on proteins involved in attachment to, and invasion of, host cells (Ludington and Ward, 2015). Some candidate antigens might be polymorphic, and immunisation with predominant antigens recognised in the population through between-host comparisons could prompt immune-escape by cryptic variants. In such a case, a vaccine showing high efficacy in a region dominated by homogeneity might fail to protect vulnerable populations exposed to highly heterogeneous parasites in other regions.

A highly heterogeneous within-host genetic structure could also shape the natural history of cryptosporidiosis. It has been hypothesised that super-infections or exposure to heterogeneous doses may increase the genetic diversity of *Cryptosporidium* in the gut (Tanriverdi et al., 2008), potentially promoting immune-escape by variants, and perhaps chronicity. Finally, the use of methods for the identification of within-host minority variants could underpin a better understanding of mixed species, genotype and sub-genotype infections, which may lead to a re-evaluation of the host range for certain taxa. For example, the numerous identifications of *C. hominis* in farmed ruminants during the last decade, either as the primary species (Ryan et al., 2005; Smith et al., 2005; Giles et al., 2009; Abeywardena et al., 2012; Connelly et al., 2013; Ghaffari et al., 2014), or as a minority variant (Tanriverdi et al., 2003), raises the possibility that *C. hominis* is not limited to humans as commonly assumed, but is also widespread as a minority variant in ruminants.

To conclude, the *Cryptosporidium* within-host genetic diversity is congenial to our current understanding of the parasites’ life cycle and is supported by long-standing experimental evidence. The results of our bibliographical search indicate *Cryptosporidium* mixed species and/or intra-species within-host genetic diversity



