Parasitology Research (2019) 118:3195–3204
https://doi.org/10.1007/s00436-019-06537-x

GENETICS, EVOLUTION, AND PHYLOGENY - REVIEW

Comparative genomics: how has it advanced our knowledge of cryptosporidiosis epidemiology?

Yingying Fan[1] · Yaoyu Feng[1] · Lihua Xiao[1]

Received: 12 September 2019 / Accepted: 30 October 2019 / Published online: 14 November 2019

Abstract

Whole genomic sequencing (WGS) and comparative genomics are increasingly used in the characterization of *Cryptosporidium* spp. They are facilitated by the establishment of procedures for WGS analysis of clinical specimens without laboratory propagation of pathogens. Results of recent comparative genomics analysis suggest that gene duplication might be associated with broad host ranges of some zoonotic *Cryptosporidium* species and subtypes, while genetic recombination could be involved in the emergence of virulent subtypes. The availability of WGS data has further facilitated the development of advanced molecular typing tools. The use of these tools together with comparative genomics analyses has begun to improve the investigations of outbreaks in industrialized nations. More WGS data, however, are needed from both industrialized nations and developing countries before we can have in-depth understanding of the population genetics and evolution of *Cryptosporidium* spp. and genetic determinants of various phenotypic traits in human-pathogenic subtypes.

Keywords *Cryptosporidium* · Whole genome sequencing · Comparative genomics · Epidemiology

Introduction

Protozoan parasites *Cryptosporidium* spp. are leading causes of diarrhea in humans and various animals (Kotloff 2017). Currently, nearly 40 *Cryptosporidium* species and about the same number of genotypes are recognized, most of which have some host specificity (Feng et al. 2018). A few species, such as *C. parvum*, *C. ubiquitum*, and *C. muris*, have broad host ranges, causing zoonotic infections together with several other *Cryptosporidium* spp. (Ryan et al. 2016). The genetic basis for host adaptation in *Cryptosporidium* spp. is unclear, partially because of the lack of whole genome sequence (WGS) data despite the sequencing of the *C. parvum* and *C. hominis* genomes 15 years ago (Abrahamsen et al. 2004; Xu et al. 2004). With the advent of next-generation sequencing (NGS) technologies, WGS analysis is increasingly used in the characterization of *Cryptosporidium* spp. This review describes recent advances in research on genomics of *Cryptosporidium* spp. and discusses their impacts on our understanding of the epidemiology of cryptosporidiosis.

Historical perspective on genomics of *Cryptosporidium* spp.

A major breakthrough in research on genomics of *Cryptosporidium* spp. is the WGS analysis of *C. parvum* Iowa isolate and *C. hominis* TU502 using the random shotgun Sanger sequencing approach (Abrahamsen et al. 2004; Xu et al. 2004). Prior to this, the chromosome 6 of the *C. parvum* IOWA isolate was sequenced (Bankier et al. 2003). The genomes of *C. parvum* and *C. hominis* both appear to have 8 chromosomes. However, only the *C. parvum* genome was fully assembled at the chromosome level using a HAPPY map to scaffold and order the

Handling Editor: Una Ryan

Yaoyu Feng
yyfeng@scau.edu.cn

Lihua Xiao
lxiao@scau.edu.cn

[1] College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, Guangdong, China

sequence contigs. The *C. hominis* genome, in contrast, was represented by 1422 contigs (Jex and Gasser 2014). Both genomes nevertheless have similar genome size (~ 9.1 Mb), GC contents (~ 30%), intron distribution (5–20% of their genes), and nucleotide sequences (95–97% identity). Altogether, 3994 and 3952 (3805 protein-encoding) genes were predicted in the *C. hominis* and *C. parvum* genomes, respectively (Abrahamsen et al. 2004; Bankier et al. 2003; Xu et al. 2004).

The sequencing of *C. parvum* and *C. hominis* genomes was followed by WGS analysis of the *C. muris* RN66 isolate through a hybrid sequencing approach (~ 4.5-fold coverage using the traditional Sanger technology and ~ 10-fold coverage using the Roche 454 technology). The outcome of this WGS effort has never been formally published, although the WGS data have been available on CryptoDB since 2008 (www.cryptodb.org). The WGS data of *C. muris* are valuable in understanding biological differences among *Cryptosporidium* species. Unlike the genetically related intestinal species *C. hominis* and *C. parvum*, *C. muris* is a divergent gastric species of rodents. The *C. muris* WGS data consist of 45 contigs of 9.2 Mb, which is slightly larger than genomes of *C. parvum* and *C. hominis*. It is predicted to encode ~ 3900 protein-coding genes. Unlike *C. parvum* and *C. hominis*, which rely on anaerobic metabolism for energy production, *C. muris* apparently has all enzymes associated with the TCA cycle and a conventional respiratory chain system (Mogi and Kita 2010). Thus, the sequencing of the *C. muris* genome has provided much needed data for comparative genomic studies.

As host adaptation exists within *C. parvum*, the genome of the *C. parvum* isolate TU114 of the anthroponotic subtype IIcA5G3b propagated in immunosuppressed mice was sequenced using the 50-bp paired-end Illumina technique (Widmer et al. 2012). The WGS data were compared with existing data from *C. parvum* IOWA (IIaA15G2R1 subtype) and *C. hominis* TU502 (IaA25R3 subtype). Over 12,000 single-nucleotide polymorphisms (SNPs) were identified between the two *C. parvum* isolates. At some polymorphic genetic loci, there was more sequence similarity between the anthroponotic *C. parvum* subtype and *C. hominis* than between anthroponotic and zoonotic *C. parvum*. It was speculated that the anthroponotic phenotype of both *C. parvum* TU114 and *C. hominis* was probably under the influence of the same positive selection. This unexpected observation suggests that some divergent genetic loci might be associated with the host preference (Widmer et al. 2012). The same isolate was sequenced again recently used the newer 250-bp paired-end Illumina technique, with the data being deposited on the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov).

Development of procedures for sequencing clinical isolates directly from fecal samples and single cell sequencing

An initial obstacle in genomics research of *Cryptosporidium* spp. was the lack of access to sufficient numbers of highly purified oocysts for NGS analysis due to the absence of an *in vitro* culture system for the propagation of parasites. Previous WGS analyses of *Cryptosporidium* spp. were done on oocysts purified from infected laboratory animals (Abrahamsen et al. 2004; Widmer et al. 2012; Xu et al. 2004). To overcome this technical difficulty, procedures have been developed for the isolation and enrichment of *Cryptosporidium* genomic DNA directly from stool specimens (Guo et al. 2015b; Li et al. 2014). This was achieved by oocyst purification using a combination of sucrose flotation and immunomagnetic separation [IMS], removal of residual contaminants by hypochloride treatment of purified oocysts, DNA extraction and whole genome amplification (WGA) of the extracted DNA using commercial kits, and assessment of *Cryptosporidium* DNA quantity using qPCR analysis of WGA products (Guo et al. 2015b). Both Roche 454 and Illumina paired-end technologies were effectively used in WGS analysis of over 20 field isolates from six *Cryptosporidium* species and genotypes (Feng et al. 2017; Guo et al. 2015a; Guo et al. 2015b; Guo et al. 2015c; Li et al. 2014). A similar approach has been used by other researchers in WGS analysis of *Cryptosporidium* spp. in clinical specimens (Andersson et al. 2015; Hadfield et al. 2015; Troell et al. 2016). Recently, Troell et al. (2016) developed a robust method to obtain WGS data from flow cytometry-sorted single *Cryptosporidium* oocyst (Troell et al. 2016). This technology could be an effective tool of WGS analysis of other uncultivable unicellular organisms. With the rapid falling cost of NGS analysis, we are expected to see an explosion of WGS data on *Cryptosporidium* spp. of both humans and animals.

Available WGS data on *Cryptosporidium* spp.

Recently, an array of new WGS data become available. A collection of genomic data are currently available in CryptoDB and NCBI databases (Heiges et al. 2006), including updated versions of WGS data from *C. hominis* TU502_2012 and *C. parvum* IOWA and new WGS data from additional isolates (Feng et al. 2017; Gilchrist et al. 2018; Guo et al. 2015c; Hadfield et al. 2015; Ifeonu et al. 2016; Isaza et al. 2015; Liu et al. 2016; Nader et al. 2019; Nash et al. 2018; Sikora et al. 2017) (Table 1; Fig. 1). These data provide valuable genomic information on *Cryptosporidium* spp. of public health and veterinary

Table 1 *Cryptosporidium* WGS data available in CryptoDB and NCBI

Species	Isolate	Subtype	Host	Geographic sources	Sequencing techniques	BioProject ID	Accession number
C. andersoni	37034	N/A	Cattle	Egypt: Kafr El Sheikh	Illumina GAIIx	PRJNA308889	LRBU01000000
C. andersoni	30847	N/A	Cattle	Canada: Alberta	Illumina GAIIx	PRJNA308889	LRBS01000000
C. andersoni	31729	N/A	Cattle	China: Zhengzhou	Illumina GAIIx	PRJNA308889	LRBT01000000
C. baileyi	TAMU-09Q1	N/A	*Colinus virginianus*	USA; Texas	Illumina MiSeq	PRJNA222835	JIBL01000000
C. cuniculus	UKCU2	VbA28	*Homo sapiens*	UK	Illumina HiSeq	PRJNA315496	PVQC01000000
C. hominis	UdeA01	IeA11G3T3	*Homo sapiens*	Colombia	454	PRJEB10000	CXWB01000000
C. hominis	UKH1	IbA10G2	*Homo sapiens*	UK: Wales	Illumina MiSeq	PRJNA222837	JIBN01000000
C. hominis	UKH3	IbA10G2	*Homo sapiens*	UK: England	Illumina MiSeq	PRJNA253834	LJRW01000000
C. hominis	UKH4	IaA14R3	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA253838	LKHI01000000
C. hominis	UKH5	IbA10G2	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA253839	LKHJ01000000
C. hominis	TU502	IaA25R3	*Homo sapiens*	Uganda: Kampala	Sanger sequencing	PRJNA13200	AAEL01000000
C. hominis	TU502_2012	IbA10G2	*Homo sapiens*	Uganda	Illumina MiSeq	PRJNA222836	JIBM01000000
C. hominis	SWEH1	IbA9G3	*Homo sapiens*	Ethiopia	IonTorrent	PRJNA307563	LQXE01000000
C. hominis	30974	IbA10G2	*Homo sapiens*	USA: Columbia, SC	454	PRJNA252787	JTAH01000000
C. hominis	30976	IaA28R4	*Homo sapiens*	USA: St. Louis	Illumina GAIIx	PRJNA252787	JTAI01000000
C. hominis	33537	IaA28R4	*Homo sapiens*	USA: Walsenburg, CO	454	PRJNA252787	JTAJ01000000
C. hominis	37999	IbA10G2	*Homo sapiens*	USA: Twin Fall, Idaho	Illumina GAIIx	PRJNA252787	JRXJ01000000
C. meleagridis	UKMEL1	IIIbA22G1R1c	*Homo sapiens*	India	Illumina MiSeq	PRJNA222838	JIBK01000000
C. meleagridis	UKMEL3	IIIgA23G3	*Homo sapiens*	UK	Illumina HiSeq	PRJNA315502	PVQF01000000
C. meleagridis	UKMEL4	IIIhA7	*Homo sapiens*	UK	Illumina HiSeq	PRJNA315503	PUXV01000000
C. meleagridis	TU1867	IIIeA21G2R1	*Homo sapiens*	Uganda	Illumina MiSeq	PRJNA192428	SRX253214
C. muris	RN66	N/A	Wild rat	Japan: Osaka	Hybrid sequencing	PRJNA19553	AAZY02000000
C. parvum	TU114	IIcA5G3b	*Homo sapiens*	Uganda	Illumina MiSeq	PRJNA437480	PYBU01000000
C. parvum	IOWA	IIaA15G2R1	Cattle	USA, Iowa	Sanger sequencing	PRJNA144	AAEE01000000
C. parvum	35090	IIaA15G1R1	Cattle	Egypt: El Beheira	Illumina GAIIx	PRJNA320419	LXLE01000000
C. parvum	34902	IIdA20G1	*Bubalus bubalis*	Egypt: Kafr El Sheikh	Illumina GAIIx	PRJNA320419	LXLD01000000
C. parvum	31727	IIdA19G1	Cattle	China: Henan	Illumina GAIIx	PRJNA320419	LXLC01000000
C. parvum	UKP1	IIaA17G1R1	*Homo sapiens*	UK	Illumina HiSeq	PRJNA439211	PYCJ01000000
C. parvum	UKP2	IIaA19G1R2	*Homo sapiens*	UK: England	Illumina MiSeq	PRJNA253836	LKHK02000000
C. parvum	UKP3	IIaA18G2R1	*Homo sapiens*	UK: Wales	Illumina HiSeq	PRJNA253840	LKHL01000000
C. parvum	UKP4	IIaA15G2R1	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA253843	LKHM01000000
C. parvum	UKP5	IIaA15G2R1	*Homo sapiens*	UK	Illumina HiSeq	PRJNA253845	LKHN01000000
C. parvum	UKP6	IIaA15G2R1	*Homo sapiens*	UK	Illumina HiSeq	PRJNA253846	LKCK01000000
C. parvum	UKP7	IIaA17G1R1	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA253847	LKCL01000000
C. parvum	UKP8	IIdA22G1	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA253848	LKCJ01000000
C. parvum	UKP12	IIcA5G3p	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA315504	PUXS01000000
C. parvum	UKP13	IIcA5G3a	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA315505	PVQD00000000
C. parvum	UKP14	IIcA5G3a	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA315506	PUXT01000000
C. parvum	UKP15	IIcA5G3a	*Homo sapiens*	UK: Wales	Illumina HiSeq	PRJNA315507	PVQE01000000
C. parvum	UKP16	IIcA5G3j	*Homo sapiens*	UK: Wales	Illumina HiSeq	PRJNA315508	PUXU01000000
C. ubiquitum	UKUB1	XIIb	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA315509	QZWX01000000
C. ubiquitum	UKUB2	XIIb	*Homo sapiens*	UK: Wales	Illumina HiSeq	PRJNA315510	QZWY01000000

Table 1 (continued)

Species	Isolate	Subtype	Host	Geographic sources	Sequencing techniques	BioProject ID	Accession number
C. ubiquitum	39725	XIIc	*Homo sapiens*	USA: Wisconsin	Illumina GAIIx	PRJNA246478	LRBR01000000
C. ubiquitum	39726	XIIb	*Homo sapiens*	USA: Wisconsin	Illumina GAIIx	PRJNA534291	LRBP01000000
C. ubiquitum	39668	XIIb	*Homo sapiens*	USA: Wisconsin	Illumina GAIIx	PRJNA246478	LRBQ00000000
C. tyzzeri	UGA55	IXdA6	*Mus musculus domesticus*	USA: Athens, Georgia	Illumina MiSeq	PRJNA388495	SRS2284677
C. viaorum	UKVIA1	XVaA3f	*Homo sapiens*	UK: England	Illumina HiSeq	PRJNA492837	QZWW01000000
Chipmunk genotype I	LX-2015	XIVaA14G2T2	*Homo sapiens*	USA: Vermont	Illumina GAIIx	PRJNA272389	JXRN01000000
Chipmunk genotype I	37763	XIVaA14G2T2	*Homo sapiens*	USA: Vermont	Illumina HiSeq	PRJNA511361	SCFC00000000

N/A not available

Fig. 1 Molecular phylogeny (based on the partial SSU rRNA gene) of *Cryptosporidium* spp.. *Cryptosporidium* species or genotypes for which whole genomic data are available in public databases are highlighted in red (adapted from Xiao and Feng 2008)

interests. Comparative genomics analyses of these data have significantly improved our understanding of the genetics and evolution of *Cryptosporidium* spp. and genetic determinants for some phenotypic traits.

Comparative genomics of *Cryptosporidium* spp.

Genome similarity between *C. parvum* and *C. hominis*

WGS analysis of *C. parvum* IOWA strain and *C. hominis* TU502 has enabled researchers to compare their genetics. Initial comparative genomics analyses of WGS data from the two isolates indicated that chromosomes in the two species were completely syntenic in gene organization and had only 3–5% differences in sequences, with no evidence of major insertions, deletions, or rearrangements (Fig. 2). Their gene set appeared to be almost identical, with the < 50 additional genes in *C. hominis* largely present around the known gaps in the genome assembly of the species (Mazurie et al. 2013). These observations indicate that both of them have the same or similar genomic characteristics, with the differences in phenotypic traits, such as host range (Xiao et al. 2004), virulence, and the biological fitness (Bouzid et al. 2013), possibly being caused by subtle sequence variation in coding regions or relating to expression levels of positively selected genes (Mazurie et al. 2013). The synteny in gene organization and other similarities in genomic features between *C. parvum* and *C. hominis* have been confirmed in recent comparative genomics analyses (Guo et al. 2015c; Isaza et al. 2015; Liu et al. 2016; Xu et al. 2019). These shared features are listed in Table 2.

Copy number variations in subtelomeric genes between *C. hominis* and *C. parvum*

More recent comparative genomic analysis of newly acquired WGS data from several *C. hominis* isolates (of the IbA10G2 and IaA28R4 subtypes) and the fully assembled reference *C. parvum* IOWA genome has identified differences in copy numbers of several multigene families, which were neglected in previous studies probably due to the fragmented nature of the *C. hominis* TU502 genome (Guo et al. 2015c). These gene insertions and deletions mostly occur in the subtelomeric regions of chromosomes 5 and 6, and involve mainly two *Cryptosporidium*-specific secretory protein families: MEDLE proteins and insulinase-like proteases. Of the six genes of the MEDLE family in *C. parvum* (cgd5_4580, cgd5_4590, cgd5_4600, cgd5_4610, cgd6_5480, and cgd6_ 5490), only one, the ortholog of cgd5_4600, exists in *C. hominis*. Similarly, two subtelomeric genes (cgd6_5510 and cgd6_5520) that encode insulinase-like proteases are absent in *C. hominis*. It was suggested that these proteins could be responsible for host expansion in *C. parvum* (Guo et al. 2015c). This has been recently supported by biological

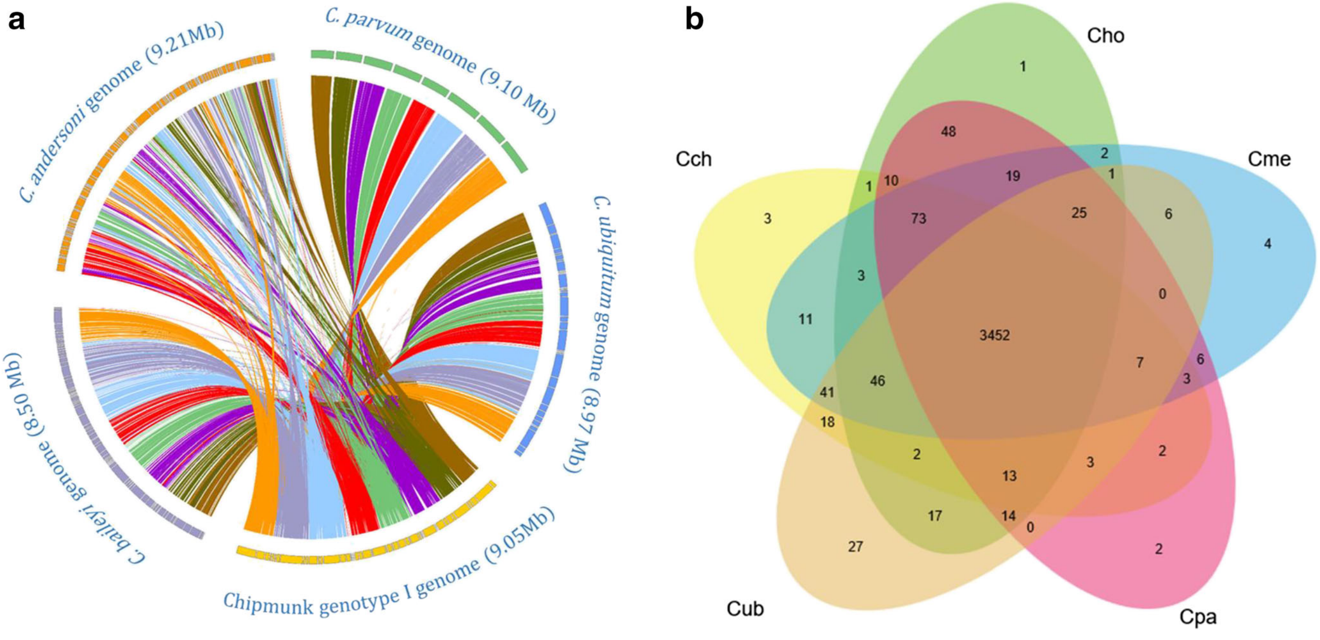

Fig. 2 Syntenic relationship and shared orthologous genes among *Cryptosporidium* spp.. **a** Syntenic relationship in gene organization among genomes of *Cryptosporidium* chipmunk genotype I, *Cryptosporidium parvum*, *C. hominis*, *C. ubiquitum*, *C. baileyi*, and *C. andersoni*. Syntenic blocks (regions with orthologous genes) are connected with lines, with the colors representing 8 chromosomes of *C. parvum*. **b** Venn diagram of shared orthologs and species-specific genes among five *Cryptosporidium* species. Abbreviations of taxa: *Cryptosporidium parvum* IOWA (Cpa); *C. hominis* Ude (Cho); *C. meleagridis* (Cme); *Cryptosporidium* chipmunk genotype I (Cch); *C. ubiquitum* (Cub) (Source: Xu et al. 2019)

characterizations of MEDLE proteins, which appear to have developmentally regulated expression and are potentially involved in the invasion of host cells (Fei et al. 2018; Li et al. 2017; Xu et al. 2019). In addition, a recent study has shown that cgd6_5510 and cgd6_5520 in *C. parvum* are actually fragments of one gene (cgd6_5520-5510) encoding a full insulinase-like protease (INS20-19), and provided some preliminary data on the potential involvement of the protein in the invasion or early developmental process of *C. parvum* (Zhang et al. 2019). A few other gene gains and losses are also present between *C. parvum* and *C. hominis*, including the presence of one *C. hominis*-specific gene, Chro.50011 (Guo et al. 2015c). Some of the differences in gene content between *C. parvum* and *C. hominis* have been observed in comparative genomics analysis of a *C. hominis* isolate from Colombia (Isaza et al. 2015).

Table 2 Comparison of genomic features among some *Cryptosporidium* species*

Feature	Intestinal						Gastric	
	C. parvum isolate IIdA20G1	*C. hominis* isolate 30976	*C. meleagridis* isolate UKMEL1	*C. ubiquitum* isolate 39726	*C. baileyi* isolate TAMU-09Q1	Chipmunk genotype I	*C. muris* isolate RN66	*C. andersoni* isolate 30847
Total length (Mb)	9.1	9.06	8.97	8.97	8.94	9.05	9.21	9.09
No. of super contigs	8	97	57	27	153	50	45	135
GC content (%)	30.3	30.1	31.0	30.8	24.3	32	28.4	28.5
No. of genes	3805	3819	3782	3767	3728	3783	3937	3905
Total length of CDS (Mb)	6.83	6.81	6.91	6.94	6.69	6.94	6.93	6.86
GC content in CDS	31.9	31.8	32.4	33.0	25.6	33.6	30.0	30.1
Gene density (gene/Mb)	418.1	421.5	421.6	420.0	438.6	418	427.5	429.6
Percent coding (%)	75.0	75.2	77.0	77.4	78.7	76.7	75.2	75.5
No. of genes with intron	163	417	506	758	763	515	798	832
Genes with intron (%)	4.2	10.9	13.4	20.1	20.5	13.6	20.3	21.3
No. of tRNA	45	45	45	45	46	45	45	44
No. of $tRNA^{met}$	2	2	2	2	2	2	2	2
Proteins with signal peptide	397	391	397	399	344	396	323	309
Proteins with transmembrane domain	832	817	805	772	813	793	836	839
Proteins with GPI anchor	63	54	55	50	57	57	52	47

*Based on data from CryptoDB and NCBI.

Genomic differences among host-adapted *C. parvum* subtype families

Host adaptation is known to occur in *C. parvum*, with IIa and IId subtype families preferentially infecting calves and lambs in some European countries, respectively (Fig. 3a). In addition to the highly divergent genes previously identified between *C. parvum* IIa (zoonotic) and IIc (anthroponotic) subtype families (Widmer et al. 2012), recent comparative genomics analyses of two IId isolates from China and Egypt and one IIa isolate from Egypt have shown differences in the number of three subtelomeric gene families among these isolates: SKSR secretory proteins, MEDLE family of secretory proteins, and insulinase-like proteases. It has further identified additional highly polymorphic genes in subtelomeric regions. The outcome of the analysis supports the view mentioned above that proteins encoded by these genes could be attributed to host adaptation in *Cryptosporidium* spp. (Feng et al. 2017).

A recent study by Nader et al. (2019) explored the population structures and genetics of *C. parvum*. In this comparative genomics analysis, *C. parvum* was separated into two subspecies based on genetic and biological differences, namely, *C. p. parvum* (zoonotic IIa and IId lineages) and *C. p. anthroponosum* (anthroponotic IIc lineage). They discovered that these lineages have undergone genetic recombination and that regions of the *C. hominis* genome were introgressed into each *C. parvum* lineage. In addition, some of these introgressed regions have shown signs of positive selection, especially in *C. p. anthroponosum*. As expected, genetic exchanges also exist between *C. p. parvum* and *C. p. anthroponosum*, with the majority of recombination events occurring in subtelomeric regions. Thus, genetic exchange or recombinant may play a potential role throughout the adaptive evolution of *Cryptosporidium* spp. (Nader et al. 2019).

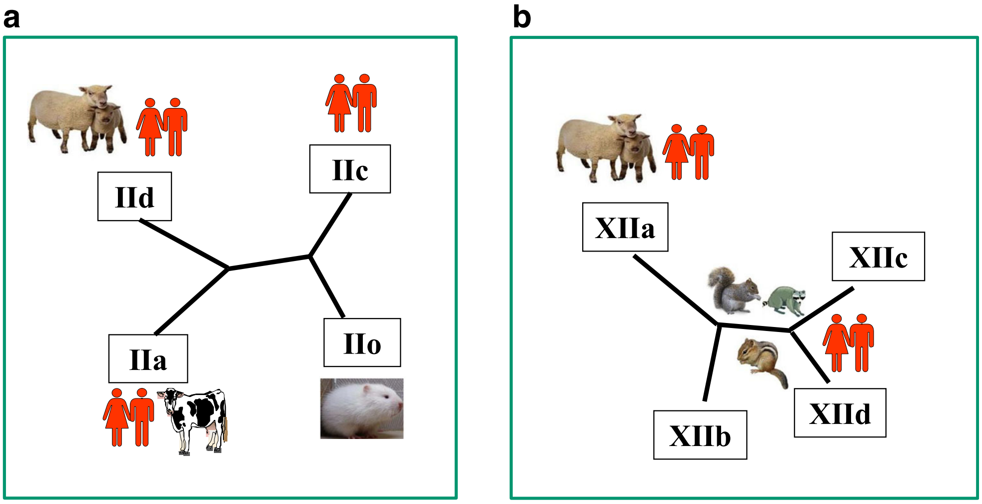

Fig. 3 **a** Major host-adapted *Cryptosporidium parvum* subtype families (with IIa infects cattle, IIc infects humans, IId infects sheep and goats, IIo infects bamboo rats). **b** A broad range of *C. ubiquitum* (with XIIa infects ruminants and humans, other subtype families infect rodents and humans)

Genomic uniqueness of *Cryptosporidium* chipmunk genotype I, *Cryptosporidium ubiquitum*, and gastric species *Cryptosporidium andersoni*

Comparative analyses have also been conducted recently on *Cryptosporidium* chipmunk genotype I (Guo et al. 2015a), a newly emerged zoonotic pathogen in some industrialized nations, *C. ubiquitum*, a zoonotic species with a broad host range (Fayer et al. 2010) (Fig. 3b), and *C. andersoni*, a gastric species mainly infecting bovine animals (Xiao et al. 2004). They have revealed conservation in genome organization among intestinal *Cryptosporidium* species and sequence divergence in potential secreted pathogenesis determinants (PSDs) among major human-infecting species (Liu et al. 2016; Xu et al. 2019). These PSDs are mostly subtelomeric in chromosomal location, often members of multigene families, and under positive selection. These observations reaffirm that the loss of some subtelomeric genes encoding insulinase-like proteases and secreted MEDLE proteins and high sequence divergence in PSDs could contribute to the biological differences among human-pathogenic *Cryptosporidium* species (Xu et al. 2019).

These comparative genomics analyses have further shown significant genetic differences between gastric and intestinal *Cryptosporidium* species, not only in mitosome metabolism as observed previously in comparative genomics analysis of another gastric species, *C. muris* (Mogi and Kita 2010), but also in genome organization and invasion-related proteins and peptidases (Liu et al. 2016; Xu et al. 2019). Therefore, *C. andersoni* and *C. muris* possess some aerobic metabolism and a conventional electron transport chain, whereas *C. ubiquitum* has lost both of them (Liu et al. 2016). There appears to be a progressive reductive evolution of metabolism in the *Cryptosporidium* lineage within Apicomplexa; compared with *C. parvum* and *C. hominis*, *C. ubiquitum* has lost the entire electron transport chain and the associated biosynthesis of ubiquinone and polyisoprenoids. These findings suggest that there a need of different strategies for the development of intervention measures against gastric and intestinal *Cryptosporidium* spp. (Liu et al. 2016).

New subtyping tools for *Cryptosporidium* spp. based on WGS data

The WGS analysis of *Cryptosporidium* spp. has provided much needed data for the development of advanced typing tools for human-pathogenic species or genotypes that are genetically divergent from *C. parvum* and *C. hominis*. For example, gp60-based subtyping tools have been developed based on newly acquired nucleotide sequences of the target gene for *C. ubiquitum* and chipmunk genotype I (Guo et al. 2015a; Li et al. 2014). This genetic locus is widely used in subtype analysis of *C. parvum* and *C. hominis* because of its high sequence polymorphism (Xiao and Feng 2017). The high sequence heterogeneity, however, has also made the design of PCR primers challenging, especially for species or genotypes divergent from *C. parvum* and *C. hominis* (Feng et al. 2011a). The availability of WGS data from other *Cryptosporidium* spp. has greatly facilitated the development of new gp60-based subtyping tools.

The WGS data have also been used in the development of high-resolution multilocus sequence typing (MLST) tools for *C. parvum* and *C. hominis*. Currently, the molecular tracking of some hyper-transmissible and virulent subtypes of *C. parvum* and *C. hominis* is challenging, making the investigation of outbreaks associated with them difficult. MLST tools thus have been developed based on other polymorphic markers in the *C. parvum* and *C. hominis* genomes, especially those with simple tandem repeats (micro- or minisatellites) (Feng et al. 2014; Li et al. 2013; Robinson and Chalmers 2012; Tanriverdi et al. 2008). MLST tools are now available for another major zoonotic species *C. ubiquitum* and gastric species *C. andersoni*/*C. muris* (Feng et al. 2011b; Tang et al. 2016).

The outcome of comparative genomics analysis has also lead to the development of an amplicon sequencing tool for the highly conserved IbA10G2 subtype in European countries. The assay was developed based on analysis of WGS data from 17 *C. hominis* isolates and involved a one-step PCR analysis of nine polymorphic loci (Beser et al. 2017; Sikora et al. 2017). It identified ten sequence types among 44 clinical isolates in Sweden, with isolates within an outbreak or cluster belonging to the same sequence type (Beser et al. 2017). Therefore, WGS has allowed the development of advanced typing tools for virulent or hyper-transmissible *C. parvum* and *C. hominis* subtypes for which gp60 subtype is inadequate for outbreak identification and trace-back.

The availability of these MLST tools has facilitated the examination of the association between some gp60 subtypes and virulence (Cama et al. 2008; Cama et al. 2007; Iqbal et al. 2011). For example, population genetic analyses of *C. hominis* specimens from a peri-urban community in Lima, Peru, at 32 polymorphic loci in chromosome 6 indicated that genetic recombination mostly occurred in the virulent subtype IbA10G2 (Li et al. 2013). Similarly, genetic recombination might have also played a role in the generation of the virulent *C. hominis* subtype IaA28R4, which emerged in the USA in recent years (Feng et al. 2014). Genetic recombination is apparently also common in the hyper-transmissible *C. parvum* subtype IIaA15G2R1 (Feng et al. 2013), which is the most common subtype in humans and dairy calves in industrialized nations and the dominant subtype for outbreaks of cryptosporidiosis in the USA (Xiao 2010).

Comparative genomics of cryptosporidiosis epidemiology

Comparative genomics analysis of IbA10G2 has shown a low genetic heterogeneity among isolates of this virulent subtype, which is responsible for almost all autochthonous *C. hominis* infections in Europe (Chalmers and Caccio 2016). It has further revealed several unique features in IbA10G2 isolates, such as the synonymous nature of most SNPs, a lower transition/transversion ration of these SNPs compared with other *C. hominis* subtypes, and the absence of a deleterious mutation in COWP9 (Sikora et al. 2017).

Comparative analysis indicates that the genomes of IbA10G2 and IaA28R4 in the USA are highly similar, with major sequence differences occurring in the 5′ (around cgd6_60) and 3′ (around cgd6_5270) subtelomeric regions and around gp60 (cgd6_1080) of chromosome 6. In these regions, the mosaic distribution of bi-allelic sequences has revealed the occurrence of genetic recombination, which might have played a role in the generation of virulent *C. hominis* subtypes (Guo et al. 2015c). In contrast, the genomes of *C. hominis* IbA10G2 in Europe seem to be highly conserved (Beser et al. 2017; Chalmers and Caccio 2016). A similar *in silico* analysis of WGS data from several IbA10G2 isolates has identified fewer than 50 SNPs among most isolates from Europe (Chalmers and Caccio 2016; Sikora et al. 2017). The high genome sequence conservation in IbA10G2 presents major challenges to the investigation and tracking of *C. hominis* in European countries using conventional gp60 subtyping.

In contrast, a recent comparative genomics analysis of *C. hominis* isolates has revealed high genetic variations and frequent genetic recombination among *gp60* subtypes in Bangladesh (Gilchrist et al. 2018), which could be a reflection of the high endemicity of cryptosporidiosis and extensive subtype diversity of *C. hominis* in the study area. A total of 367,810 SNPs were observed among 32 *C. hominis* isolates sequenced, most of which appeared evenly distributed throughout the genomes, except for seven regions. The latter included the gp60 (cgd6_1080), SKSR (cgd1_3850), and insulinase-like peptidases (cdg3_4260) regions. The recombination rate was so high that linkage disequilibrium in the genome was lost among isolates after only 300 bp, and the genomes did not segregated by *gp60* subtype (Gilchrist et al. 2018). Therefore, the frequency of genetic recombination within *C. hominis* appears to be very different between industrialized nations and developing countries.

Conclusions and future perspectives

As discussed, WGS analysis and comparative genomics are increasingly used in molecular characterizations of *Cryptosporidium* spp. (Khan et al. 2018). The outcome thus far has significantly improved our understanding of basic biology and transmission dynamics of these important zoonotic pathogens. We now have increased appreciation of the role of genetic recombination in the evolution and dispersal of virulent and hyper-transmissible *C. hominis* and *C. parvum* subtypes. In addition, they have facilitated the development of advanced molecular detection tools that can be routinely used in the investigation of cryptosporidiosis outbreaks, especially those caused by hyper-transmissible and virulent subtypes. The new insights provided could further promote the development of new intervention strategies against cryptosporidiosis.

We are still at the early stage of comparative genomics compared to progresses in other related apicomplexan pathogens such as *Plasmodium* spp. and *T. gondii*. Therefore, additional WGS data should be acquired from human-pathogenic and zoonotic *Cryptosporidium* spp. before we can have better understanding of population genetic structures of these pathogens and genetic determinants of virulence and hyper-transmissibility (Widmer et al. 2012). More systematic comparative genomics analyses of genetically related species with different host ranges would improve our understanding of the evolution and host adaptation within the *Cryptosporidium* genus (Jex and Gasser 2010; Khan et al. 2018).

Funding information This work was supported by the National Natural Science Foundation of China (31820103014 and 31630078).

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

References

Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science 304:441–445

Andersson S, Sikora P, Karlberg ML, Winiecka-Krusnell J, Alm E, Beser J, Arrighi RB (2015) It's a dirty job–a robust method for the purification and de novo genome assembly of *Cryptosporidium* from clinical material. J Microbiol Methods 113:10–12

Bankier AT, Spriggs HF, Fartmann B, Konfortov BA, Madera M, Vogel C, Teichmann SA, Ivens A, Dear PH (2003) Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. Genome Res 13:1787–1799

Beser J, Hallström BM, Advani A, Andersson S, Östlund G, Winiecka-Krusnell J, Lebbad M, Alm E, Troell K, Arrighi RBG (2017) Improving the genotyping resolution of *Cryptosporidium hominis* subtype IbA10G2 using one step PCR-based amplicon sequencing. Infect Genet Evol 55:297–304

Bouzid M, Hunter PR, Chalmers RM, Tyler KM (2013) *Cryptosporidium* pathogenicity and virulence. Clin Microbiol Rev 26:115–134

Cama VA, Ross JM, Crawford S, Kawai V, Chavez-Valdez R, Vargas D, Vivar A, Ticona E, Navincopa M, Williamson J, Ortega Y, Gilman RH, Bern C, Xiao L (2007) Differences in clinical manifestations among *Cryptosporidium* species and subtypes in HIV-infected persons. J Infect Dis 196:684–691

Cama VA, Bern C, Roberts J, Cabrera L, Sterling CR, Ortega Y, Gilman RH, Xiao L (2008) *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. Emerg Infect Dis 14:1567–1574

Chalmers RM, Caccio S (2016) Towards a consensus on genotyping schemes for surveillance and outbreak investigations of *Cryptosporidium*, Berlin, June 2016. Euro Surveill 21(37):30338

Fayer R, Santin M, Macarisin D (2010) *Cryptosporidium ubiquitum* n. sp. in animals and humans. Vet Parasitol 172:23–32

Fei J, Wu H, Su J, Jin C, Li N, Guo Y, Feng Y, Xiao L (2018) Characterization of MEDLE-1, a protein in early development of *Cryptosporidium parvum*. Parasit Vectors 11:312

Feng Y, Lal AA, Li N, Xiao L (2011a) Subtypes of *Cryptosporidium* spp. in mice and other small mammals. Exp Parasitol 127:238–242

Feng Y, Yang W, Ryan U, Zhang L, Kvác M, Koudela B, Modry D, Li N, Fayer R, Xiao L (2011b) Development of a multilocus sequence tool for typing *Cryptosporidium muris* and *Cryptosporidium andersoni*. J Clin Microbiol 49:34–41

Feng Y, Torres E, Li N, Wang L, Bowman D, Xiao L (2013) Population genetic characterisation of dominant *Cryptosporidium parvum* subtype IIaA15G2R1. Int J Parasitol 43:1141–1147

Feng Y, Tiao N, Li N, Hlavsa M, Xiao L (2014) Multilocus sequence typing of an emerging *Cryptosporidium hominis* subtype in the United States. J Clin Microbiol 52:524–530

Feng Y, Li N, Roellig DM, Kelley A, Liu G, Amer S, Tang K, Zhang L, Xiao L (2017) Comparative genomic analysis of the IId subtype family of *Cryptosporidium parvum*. Int J Parasitol 47:281–290

Feng Y, Ryan UM, Xiao L (2018) Genetic diversity and population structure of *Cryptosporidium*. Trends Parasitol 34:997–1011

Gilchrist CA, Cotton JA, Burkey C, Arju T, Gilmartin A, Lin Y, Ahmed E, Steiner K, Alam M, Ahmed S, Robinson G, Zaman SU, Kabir M, Sanders M, Chalmers RM, Ahmed T, Ma JZ, Haque R, Faruque ASG, Berriman M, Petri WA (2018) Genetic diversity of *Cryptosporidium hominis* in a bangladeshi community as revealed by whole-genome sequencing. J Infect Dis 218:259–264

Guo Y, Cebelinski E, Matusevich C, Alderisio KA, Lebbad M, McEvoy J, Roellig DM, Yang C, Feng Y, Xiao L (2015a) Subtyping novel zoonotic pathogen *Cryptosporidium* chipmunk genotype I. J Clin Microbiol 53:1648–1654

Guo Y, Li N, Lysén C, Frace M, Tang K, Sammons S, Roellig DM, Feng Y, Xiao L (2015b) Isolation and enrichment of *Cryptosporidium* DNA and verification of DNA purity for whole-genome sequencing. J Clin Microbiol 53:641–647

Guo Y, Tang K, Rowe LA, Li N, Roellig DM, Knipe K, Frace M, Yang C, Feng Y, Xiao L (2015c) Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. BMC Genomics 16:320

Hadfield SJ, Pachebat JA, Swain MT, Robinson G, Cameron SJ, Alexander J, Hegarty MJ, Elwin K, Chalmers RM (2015) Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples. BMC Genomics 16:650

Heiges M, Wang H, Robinson E, Aurrecoechea C, Gao X, Kaluskar N, Rhodes P, Wang S, He CZ, Su Y, Miller J, Kraemer E, Kissinger JC (2006) CryptoDB: a *Cryptosporidium* bioinformatics resource update. Nucleic Acids Res 34:D419–D422

Ifeonu OO, Chibucos MC, Orvis J, Su Q, Elwin K, Guo F, Zhang H, Xiao L, Sun M, Chalmers RM, Fraser CM, Zhu G, Kissinger JC, Widmer G, Silva JC (2016) Annotated draft genome sequences of three species of *Cryptosporidium*: *Cryptosporidium meleagridis* isolate UKMEL1, *C. baileyi* isolate TAMU-09Q1 and *C. hominis* isolates TU502_2012 and UKH1. Pathog Dis 74:ftw080

Iqbal J, Khalid N, Hira PR (2011) Cryptosporidiosis in Kuwaiti children: association of clinical characteristics with *Cryptosporidium* species and subtypes. J Med Microbiol 60:647–652

Isaza JP, Galván AL, Polanco V, Huang B, Matveyev AV, Serrano MG, Manque P, Buck GA, Alzate JF (2015) Revisiting the reference genomes of human pathogenic *Cryptosporidium* species: reannotation of *C. parvum* Iowa and a new *C. hominis* reference. Sci Rep 5:16324

Jex AR, Gasser RB (2010) Genetic richness and diversity in *Cryptosporidium hominis* and C. parvum reveals major knowledge gaps and a need for the application of "next generation" technologies–research review. Biotechnol Adv 28:17–26

Jex AR, Gasser RB (2014) *Cryptosporidium*: Current state of genomics and systems biological research. *Cryptosporidium*: parasite and disease:327–344

Khan A, Shaik JS, Grigg ME (2018) Genomics and molecular epidemiology of *Cryptosporidium* species. Acta Trop 184:1–14

Kotloff KL (2017) The burden and etiology of diarrheal illness in developing countries. Pediatr Clin N Am 64:799–814

Li N, Xiao L, Cama VA, Ortega Y, Gilman RH, Guo M, Feng Y (2013) Genetic recombination and *Cryptosporidium hominis* virulent subtype IbA10G2. Emerg Infect Dis 19:1573–1582

Li N, Xiao L, Alderisio K, Elwin K, Cebelinski E, Chalmers R, Santin M, Fayer R, Kvac M, Ryan U, Sak B, Stanko M, Guo Y, Wang L, Zhang L, Cai J, Roellig D, Feng Y (2014) Subtyping *Cryptosporidium ubiquitum*, a zoonotic pathogen emerging in humans. Emerg Infect Dis 20:217–224

Li B, Wu H, Li N, Su J, Jia R, Jiang J, Feng Y, Xiao L (2017) Preliminary characterization of MEDLE-2, a protein potentially involved in the invasion of *Cryptosporidium parvum*. Front Microbiol 8:1647

Liu S, Roellig DM, Guo Y, Li N, Frace MA, Tang K, Zhang L, Feng Y, Xiao L (2016) Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. BMC Genomics 17:1006

Mazurie AJ, Alves JM, Ozaki LS, Zhou S, Schwartz DC, Buck GA (2013) Comparative genomics of *Cryptosporidium*. Int J Genomics 2013:832756

Mogi T, Kita K (2010) Diversity in mitochondrial metabolic pathways in parasitic protists *Plasmodium* and *Cryptosporidium*. Parasitol Int 59:305–312

Nader JL, Mathers TC, Ward BJ, Pachebat JA, Swain MT, Robinson G, Chalmers RM, Hunter PR, van Oosterhout C, Tyler KM (2019) Evolutionary genomics of anthroponosis in *Cryptosporidium*. Nat Microbiol 4:826–836

Nash JHE, Robertson J, Elwin K, Chalmers RA, Kropinski AM, Guy RA (2018) Draft genome assembly of a potentially zoonotic *Cryptosporidium parvum* isolate, UKP1. Microbiol Resour Announc 7:e01291–e01218

Robinson G, Chalmers RM (2012) Assessment of polymorphic genetic markers for multi-locus typing of *Cryptosporidium parvum* and *Cryptosporidium hominis*. Exp Parasitol 132:200–215

Ryan U, Zahedi A, Paparini A (2016) *Cryptosporidium* in humans and animals-a one health approach to prophylaxis. Parasite Immunol 38: 535–547

Sikora P, Andersson S, Winiecka-Krusnell J, Hallström B, Alsmark C, Troell K, Beser J, Arrighi RB (2017) Genomic variation in IbA10G2 and other patient-derived *Cryptosporidium hominis* subtypes. J Clin Microbiol 55:844–858

Tang Y, Li N, Song M, Roellig DM, Feng Y, Xiao L (2016) Development of a multilocus sequence typing tool for high-resolution subtyping and genetic structure characterization of *Cryptosporidium ubiquitum*. Infect Genet Evol 45:256–261

Tanriverdi S, Grinberg A, Chalmers RM, Hunter PR, Petrovic Z, Akiyoshi DE, London E, Zhang L, Tzipori S, Tumwine JK, Widmer G (2008) Inferences about the global population structures of *Cryptosporidium parvum* and *Cryptosporidium hominis*. Appl Environ Microbiol 74:7227–7234

Troell K, Hallström B, Divne AM, Alsmark C, Arrighi R, Huss M, Beser J, Bertilsson S (2016) *Cryptosporidium* as a testbed for single cell genome characterization of unicellular eukaryotes. BMC Genomics 17:471

Widmer G, Lee Y, Hunt P, Martinelli A, Tolkoff M, Bodi K (2012) Comparative genome analysis of two *Cryptosporidium parvum* isolates with different host range. Infect Genet Evol 12:1213–1221

Xiao L (2010) Molecular epidemiology of cryptosporidiosis: an update. Exp Parasitol 124:80–89

Xiao L, Feng Y (2008) Zoonotic cryptosporidiosis. FEMS Immunol Med Microbiol 52:309–323

Xiao L, Feng Y (2017) Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. Food and Waterborne Parasitology 8-9:14–32

Xiao L, Fayer R, Ryan U, Upton SJ (2004) *Cryptosporidium* taxonomy: recent advances and implications for public health. Clin Microbiol Rev 17:72–97

Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA (2004) The genome of *Cryptosporidium hominis*. Nature 431:1107–1112

Xu Z, Guo Y, Roellig DM, Feng Y, Xiao L (2019) Comparative analysis reveals conservation in genome organization among intestinal *Cryptosporidium* species and sequence divergence in potential secreted pathogenesis determinants among major human-infecting species. BMC Genomics 20:406

Zhang S, Wang Y, Wu H, Li N, Jiang J, Guo Y, Feng Y, Xiao L (2019) Characterization of a species-specific insulinase-like protease in *Cryptosporidium parvum*. Front Microbiol 10:354