MICROBIAL GENOMICS

RESEARCH ARTICLE

Xu *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000379

Comparative genomic analysis of three intestinal species reveals reductions in secreted pathogenesis determinants in bovine-specific and non-pathogenic *Cryptosporidium* species

Zhixiao Xu[1], Na Li[2], Yaqiong Guo[2], Yaoyu Feng[1,2,*] and Lihua Xiao[2,*]

Abstract

The three common intestinal *Cryptosporidium* species in cattle differ significantly in host range, pathogenicity and public health significance. While *Cryptosporidium parvum* is pathogenic in pre-weaned calves and has a broad host range, *C. bovis* and *C. ryanae* are largely non-pathogenic and bovine-specific species in post-weaned calves. Thus far, only the genome of *C. parvum* has been sequenced. To improve our understanding of the genetic determinants of biological differences among *Cryptosporidium* spcies, we sequenced the genomes of *C. bovis* and *C. ryanae* and conducted a comparative genomics analysis. The genome of *C. bovis* has a gene content and organization more similar to *C. ryanae* than to other *Cryptosporidium* species sequenced to date; the level of similarity in amino acid and nucleotide sequences between the two species is 75.2 and 69.4%, respectively. A total of 3723 and 3711 putative protein-encoding genes were identified in the genomes of *C. bovis* and *C. ryanae*, respectively, which are fewer than the 3981 in *C. parvum*. Metabolism is similar among the three species, although energy production pathways are further reduced in *C. bovis* and *C. ryanae*. Compared with *C. parvum*, *C. bovis* and *C. ryanae* have lost 14 genes encoding mucin-type glycoproteins and three for insulinase-like proteases. Other gene gains and losses in the two bovine-specific and non-pathogenic species also involve the secretory pathogenesis determinants (SPDs); they have lost all genes encoding MEDLE, FLGN and SKSR proteins, and two of the three genes for NFDQ proteins, but have more genes encoding secreted WYLE proteins, secreted leucine-rich proteins and GPI-anchored adhesin PGA18. The only major difference between *C. bovis* and *C. ryanae* is in nucleotide metabolism. In addition, half of the highly divergent genes between *C. bovis* and *C. ryanae* encode secreted or membrane-bound proteins. Therefore, *C. bovis* and *C. ryanae* have gene organization and metabolic pathways similar to *C. parvum*, but have lost some invasion-associated mucin glycoproteins, insulinase-like proteases, MEDLE secretory proteins and other SPDs. The multiple gene families under positive selection, such as helicase-associated domains, AMP-binding domains, protein kinases, mucins, insulinases and TRAPs could contribute to differences in host specificity and pathogenicity between *C. parvum* and *C. bovis*. Biological studies should be conducted to assess the contribution of these copy number variations to the narrow host range and reduced pathogenicity of *C. bovis* and *C. ryanae*.

DATA SUMMARY

1. All sequencing reads of *Cryptosporidium bovis* and *Cryptosporidium ryanae* have been submitted to the NCBI Sequence Read Archive (SRA) under accessions SRR9329505 and SRR9329807, respectively. The assemblies of *C. bovis* and *C. ryanae* are deposited in GenBank under accession VHIT00000000 and VHLK00000000, respectively.

INTRODUCTION

Cryptosporidiosis is well recognized as an important cause of diarrhoea and enteric diseases in humans and domestic animals [1]. In addition to moderate-to-severe diarrhoea, it can cause weight loss and death in neonatal animals, children and immunocompromised persons [2, 3]. As *Cryptosporidium* infections are common in cattle, calves are considered major reservoir hosts [4].

Received 09 January 2020; Accepted 27 April 2020; Published 14 May 2020

Author affiliations: [1]State Key Laboratory of Bioreactor Engineering, School of Resources and Environmental Engineering, East China University of Science and Technology, Shanghai 200237, PR China; [2]Key Laboratory of Zoonosis of Ministry of Agriculture, College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, PR China.

***Correspondence:** Yaoyu Feng, yyfeng@scau.edu.cn; Lihua Xiao, lxiao1961@gmail.com

Keywords: *Cryptosporidium bovis*; *Cryptosporidium ryanae*; comparative genomics; host specificity; pathogenicity.

Abbreviations: ML, maximum-likelihood; MQO, malate quinone oxidoreductase; SPD, secreted pathogenesis determinant; TRAP, thrombospondin-related adhesive protein.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary tables are available with the online version of this article.

Cryptosporidium species vary in host range and public health significance. Thus far, over 40 *Cryptosporidium* species have been recognized [5]. Among them, *Cryptosporidium parvum* and *Cryptosporidium hominis* are two dominant species in humans. The former is also commonly found in cattle. In addition, cattle are frequently infected with *Cryptosporidium bovis*, *Cryptosporidium ryanae* and *Cryptosporidium andersoni* [5]. Among the four bovine *Cryptosporidium* species, *C. parvum*, which infects the small intestine of pre-weaned calves, is the only major species responsible for diarrhoea [3, 6]. *C. bovis* and *C. ryanae* often infect the small intestine of post-weaned calves and yearlings mostly without any clinical signs of disease [7, 8]. In contrast, *C. andersoni* infects the abomasum of mature cattle, leading to poor weight gain and reduced milk production [9]. Among the three intestinal species, *C. parvum* has a broad host range, while *C. bovis* and *C. ryanae* infect exclusively bovine animals [4].

Comparative genomics analysis of human-pathogenic *Cryptosporidium* species has revealed significant diversification in secretory pathogenesis determinants (SPDs), which include MEDLE proteins, insulinase-like proteases and mucin-type glycoproteins. Therefore, SPDs are suggested to be involved in differences in host range, tissue tropism and pathogenicity among *Cryptosporidium* species [10, 11]. Among them, MEDLE proteins were named after a conserved sequence motif at the C terminus and are expressed in the invasion stages of *C. parvum* [12, 13]. Insulinase-like proteases are widespread in apicomplexans, and are known to be involved in processing invasion-related proteins or modifying host cell activities [14]. Mucin-type glycoproteins are a large family of secreted proteins in micronemes and could be involved in the initial attachment and invasion of *Cryptosporidium* spcies [15].

Genes encoding SPDs are often arranged in the genome as clusters in the subtelomeric regions, which facilitates gene duplication, deletion and genetic recombination [11]. For example, compared with *C. parvum*, one gene encoding insulinase-like protease was lost in the 3′ subtelomeric region of chromosome 6 of *C. hominis*. In contrast, the gastric species *C. andersoni* has lost the subtelomeric regions encoding MEDLE proteins and insulinases entirely [11]. Similarly, a major difference between *C. parvum* and *Cryptosporidium* chipmunk genotype I is the loss of four subtelomeric genes encoding MEDLE proteins and one subtelomeric gene encoding an insulinase-like protease in the latter [16]. Copy number variations in the genes encoding MEDLE and insulinase-like proteases have also been seen among subtype families of *C. parvum*, which have different host preferences [17, 18]. In addition, an enrichment of positively selected genes encoding SPDs was observed in subtelomeric regions between zoonotic and anthroponotic *C. parvum* subtypes [19]. Differences in the number and sequences of genes encoding mucin-type glycoproteins could also be partially responsible for the tissue tropism between the intestinal and gastric *Cryptosporidium* species [11].

In this study, to improve our understanding of potential genetic determinants of the host range and pathogenicity in *Cryptosporidium* species, we sequenced the genomes of *C. bovis* and *C. ryanae* and performed a comparative genomics analysis of the three intestinal species infecting cattle and available whole genome sequence data from other *Cryptosporidium* species [10, 11, 20–22].

Impact Statement

Cryptosporidium species are important apicomplexan parasites, causing diarrhoea and enteric diseases in humans and domestic animals. *Cryptosporidium parvum*, *Cryptosporidium bovis* and *Cryptosporidium ryanae* are three common intestinal *Cryptosporidium* species in cattle. As a zoonotic pathogen, *C. parvum* is the only major species responsible for diarrhoea in pre-weaned calves. As bovine-specific species, *C. bovis* and *C. ryanae* often infect post-weaned calves and yearlings mostly without any clinical signs of disease. We sequenced the genomes of *C. bovis* and *C. ryanae* for the first time and conducted a comparative genomic analysis. We found that *C. bovis* and *C. ryanae* have lost many secretory pathogenesis determinants, such as mucin-type glycoproteins, insulinase-like proteases, secreted MEDLE proteins, FLGN, SKSR and NFDQ proteins, which could potentially contribute to the reduced host range and pathogenicity of *C. bovis* and *C. ryanae*. The results of our study are useful in understanding differences in pathogenicity of various *Cryptosporidium* species within the same host.

METHODS

Specimen collection and whole-genome sequencing

C. bovis isolate 42482 and *C. ryanae* isolate 45019 were collected from dairy calves in Shanghai and Guangdong, China, respectively. They were diagnosed by sequence analysis of the small subunit rRNA gene [23]. Sucrose and caesium chloride density gradient centrifugations and immunomagnetic separation were used to purify the oocysts from the specimens [24]. The purified oocysts were subjected to five freeze–thaw cycles and digested with proteinase K overnight. The QIAamp DNA Mini Kit (Qiagen Sciences) was used in extracting genomic DNA from the oocysts. The REPLI-g Midi Kit (Qiagen) was used to amplify the DNA harvested. The genomes were sequenced using Illumina HiSeq 2500 analysis of 250 bp paired-end libraries constructed using the Illumina TruSeq (v3) library preparation kit (Illumina). The sequence reads were trimmed to remove adapter sequences and regions of poor sequence quality (Phred score <25) and assembled *de novo* using the CLC Genomics Workbench Version 9.0 with word size of 63 and bulb size of 500.

Genome structure analysis and gene prediction

The *C. bovis* and *C. ryanae* genomes obtained were aligned with the published genomes of the *C. parvum* IOWA isolate [20], *C. ubiquitum* [11] and *C. andersoni* [11] using Mauve 2.3.1 [25] with default parameters. The syntenic relationship

(regions with orthologous genes) among the *C. bovis* genome and the other four genomes was illustrated using Circos 0.69 [26]. We used Bowtie2 to map the reads on the *C. bovis* genome, and the Integrative Genomics Viewer was used to check the coverage of the regions which connect large rearrangements between *C. bovis* and *C. parvum*.

After training the software with the gene model of the *C. parvum* IOWA genome, protein-encoding genes in *C. bovis* and *C. ryanae* were predicted using GeneMark-ES [27], AUGUSTUS 3.2.1 [28] and Geneid 1.4 [29] with default settings, as described previously [16]. The final gene set was generated by consensus predictor EVidence Modeler [30] based on the prediction outcomes using the three software packages.

Functional annotation

BLAST P [31] and Hidden Markov Model (HMMER) analysis (http://hmmer.org) were used to annotate the predicted genes of *C. bovis* and *C. ryanae* by searching in the GenBank NR and Pfam databases. SignalP 4.1 [32], TMHMM 2.0 [33] and the GPI-SOM webserver [34] were used to identify signal peptides, transmembrane domains and GPI anchor sites, respectively. The KAAS web server [35] was used to analyse the metabolism with the BBH (Bi-directional Best Hit) method and eukaryote gene model. The annotations of functional proteins, catalytic enzymes and metabolic pathways within the genomes were conducted using Pfam (http://pfam.xfam.org/) [36], the online database KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.genome.jp/kegg/) and LAMP (Library of Apicomplexan Metabolic Pathways, release-2) [37], respectively.

Comparative genomics analysis

Sequence similarities among *C. bovis*, *C. ryanae*, *C. parvum* and other *Cryptosporidium* genomes in CryptoDB (http://cryptodb.org/cryptodb/) were assessed by using BLAST P and HMMER with e-value thresholds of 1e-3. OrthoMCL [38] was used to identify homologous gene families among *Cryptosporidium* spcies with e-value thresholds of 1e-5. VennPainter (https://github.com/linguoliang/VennPainter) was used to draw the Venn diagram of shared orthologues and species-specific genes in *C. parvum*, *C. ubiquitum*, *C. andersoni*, *C. bovis* and *C. ryanae*. Based on results of BLAST P homology analysis (threshold of protein pairs sharing 30% identity over 100 amino acids), the relationship among proteins in *C. bovis*, *C. parvum* and *C. ryanae* was visualized using Gephi (https://gephi.org/) with the Fruchterman–Reingold layout. The data of KAAS and LAMP were used in comparative analyses of metabolism in *Cryptosporidium* species. Comparisons of transporter proteins and invasion-related proteins among *Cryptosporidium* species were based on results of Pfam searches.

Phylogenetic analysis

Amino acid sequences encoded by single-copy orthologous genes shared among *Cryptosporidium* species and *Gregarina niphandrodes* were concatenated and aligned with each other using MUSCLE [39]. Poorly aligned positions were eliminated from the sequence alignments using Gblocks [40]. RAxML was used to reconstruct maximum-likelihood (ML) trees with 1000 bootstrap replications [41]. The concatenated sequence from *G. niphandrodes* was used as the outgroup in the phylogenetic analysis.

RESULTS

Genome features

A total of 7.08 million and 5.13 million of 250 bp paired-end reads were obtained from *C. bovis* isolate 42482 (from a dairy calf in Shanghai) and *C. ryanae* isolate 45019 (from a dairy calf in Guangdong) using Illumina sequencing, respectively. The reads were assembled into a 9.11 Mb *C. bovis* genome of 55 contigs and a 9.06 Mb *C. ryanae* genome of 93 contigs after removing contigs from contaminants. We identified 3723 protein-encoding genes in *C. bovis* and 3711 in *C. ryanae* by combining the gene prediction results from GeneMark, Augustus and Geneid. The gene content of *C. bovis* and *C. ryanae* is similar to that of *C. baileyi* but smaller than the genomes of *C. parvum* and *C. hominis* (Table 1). Compared with eight other *Cryptosporidium* species, *C. bovis* has a relatively high similarity in amino acid and nucleotide sequences to *C. ryanae* (75.2 and 69.4%, respectively). The GC content of *C. ryanae* is slightly higher than that of *C. bovis* in the overall genome and coding regions (32.9 and 33.9% versus 30.7 and 31.8%, respectively) (Table 1).

A complete synteny in gene organization was observed between the *C. bovis* and *C. ryanae* genomes, but some large rearrangements were observed between the *C. bovis* and *C. parvum* genomes (Fig. 1a). For example, in a rearrangement of ~150 kb in chromosome 1 of *C. parvum* that contains 52 genes (*cgd1_500* to *11_1010*), the homologous region in *C. bovis* is fragmented into different contigs, including contig_1 (chromosome 5), contig_16 (chromosome 7) and contig_32 (chromosome 6). Similarly, an ~480 kb fragment containing 175 genes at the 5′ region of chromosome 3 of *C. parvum* is translocated to chromosomes 1, 5 and 6 in *C. bovis*. In addition, an ~303 kb fragment containing 134 genes in chromosome 2 of *C. parvum* is translocated to the 5′ subtelomeric region (contig 6) of chromosome 8 in *C. bovis*. Several other rearrangements were seen in *C. bovis*, involving the 5′ region of chromosomes 5, 6 and 8 of *C. parvum*. We found that all junction regions of the large rearrangements in *C. bovis* were mapped to reads, and most of them had high coverage (50–794-fold). Two had lower coverage, including the regions in contig_23 (coverage: 6–70-fold) and contig_6 (coverage: 20–40-fold).

Based on orthology delineation, 3059 genes are shared by *C. parvum*, *C. bovis*, *C. ryanae*, *C. ubiquitum* and *C. andersoni*. Among the remaining genes, the genes shared between *C. bovis* and *C. ryanae* are different from those shared between *C. parvum* and *C. ubiquitum*. Thus, there are 126 genes shared only by *C. parvum* and *C. ubiquitum*, two human-pathogenic species with broad host ranges, while 114 other genes are

Table 1. Genomic features of *Cryptosporidium bovis* and *Cryptosporidium ryanae* in comparison with *Cryptosporidium parvum* and other *Cryptosporidium* species

	C. muris	*C. andersoni*	*C. parvum*	*C. hominis* UdeA01	*C. meleagridis*	*Cryptosporidium* chipmunk genotype I	*C. ubiquitum*	*C. bovis*	*C. ryanae*	*C. baileyi*
Total length (Mb)	9.21	9.09	9.1	9.06	8.97	9.05	8.97	9.11	9.06	8.5
No. of super contigs	45	135	8	97	57	50	27	55	93	153
GC content (%)	28.4	28.5	30.3	30.1	31	32	30.8	30.7	32.9	24.3
Nucleotide sequence similarity (%)	24.8	25.6	38.6	38.7	38.4	38.4	38.5	–	69.4	40.9
No. of genes	3937	3905	3981	3819	3782	3783	3767	3723	3711	3728
Total length of CDS (Mb)*	6.93	6.86	6.83	6.81	6.91	6.94	6.94	6.8	6.74	6.69
GC content in CDS (%)	30	30.1	31.9	31.8	32.4	33.6	33	31.8	33.9	25.6
Amino acid sequence similarity (%)	46.9	46.7	55.1	54.8	54.5	54.6	54.8	–	75.2	57.1
GC content at third position in codons (%)	17.8	18.1	22.5	23.5	24.1	26.9	24.5	25.4	30.2	12.6
Gene density (genes/Mb)	427.5	429.6	418.1	421.5	421.6	418	420	408.7	409.6	438.6
Percentage coding (%)	75.2	75.5	75	75.2	77	76.7	77.4	74.6	74.4	78.7
No. of genes with intron	798	832	163	417	506	515	758	571	602	763
Genes with intron (%)	20.3	21.3	4.2	10.9	13.4	13.6	20.1	15.3	16.2	20.5
No. of tRNAs	45	44	45	45	45	45	45	45	45	46
No. of $tRNA^{met}$	2	2	2	2	2	2	2	2	2	2
Proteins with signal peptide	323	309	397	391	397	396	399	366	329	344
Proteins with transmembrane domain	836	839	832	817	805	793	772	781	774	813
Proteins with GPI anchor	52	47	63	54	55	57	50	62	57	57

*CDS, coding sequences.

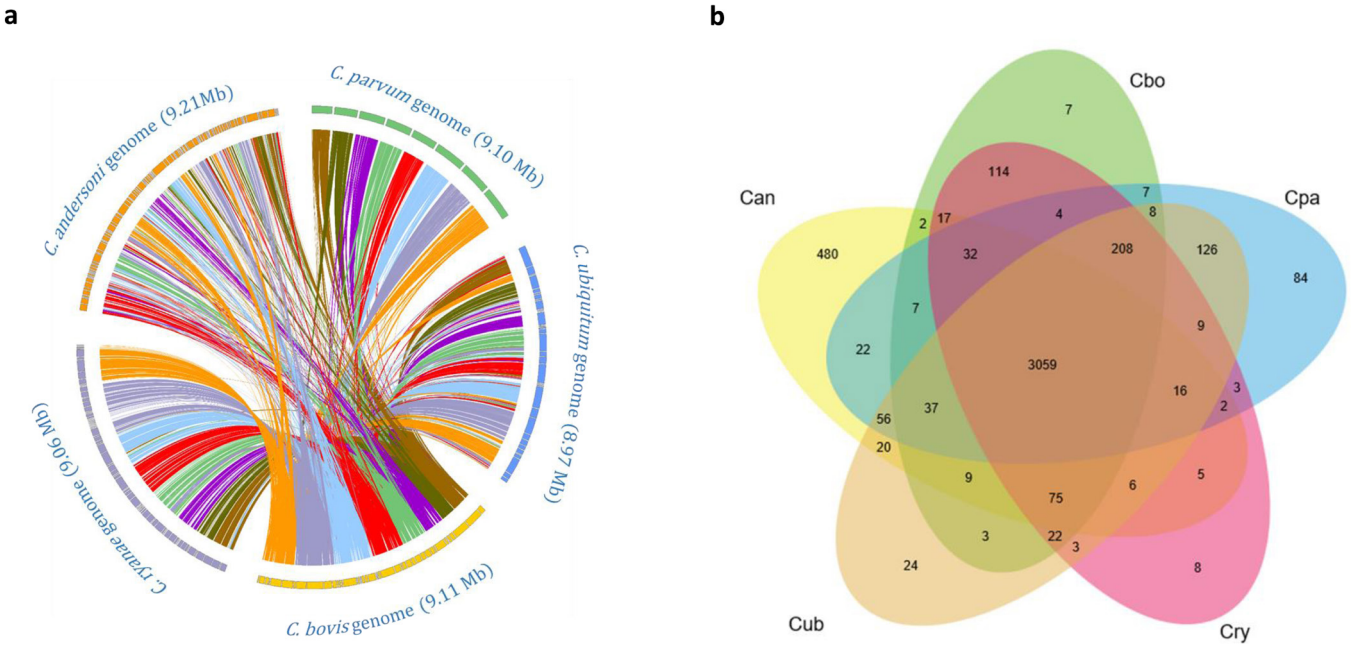

Fig. 1. Syntenic relationship and shared orthologous genes among *Cryptosporidium* species. (a) The syntenic relationship in gene organization among genomes of *Cryptosporidium parvum*, *C. ubiquitum*, *C. bovis*, *C. ryanae* and *C. andersoni*. Syntenic blocks (regions with orthologous genes) are connected with lines, with the different colours representing eight chromosomes of *C. parvum*. (b) Venn diagram of orthologous genes shared by five *Cryptosporidium* species. Abbreviations of taxa: *Cryptosporidium parvum* IOWA (Cpa); *C. ubiquitum* (Cub); *C. bovis* (Cbo); *C. ryanae* (Cry); *C. andersoni* (Can).

shared only by *C. bovis* and *C. ryanae*, two bovine-specific species (Tables S1 and S2, available in the online version of this article). Among these five *Cryptosporidium* species, *C. parvum* has 84 species-specific genes, compared with only a few species-specific genes in *C. bovis* and *C. ryanae*. The latter was largely due to the fact that *C. bovis* and *C. ryanae* share a virtually identical set of genes (Fig. 1b). Phylogenetic analysis of amino acid sequences of 100 orthologous genes supported the close relationship of *C. bovis* to *C. ryanae* (Fig. 2a). This was confirmed by phylogenetic analysis of amino acid sequences of invasion-related protein families, including mucin-type glycoproteins, insulinase-like proteases and thrombospondin-related adhesive proteins (TRAPs) (Fig. 2b–d).

Network analysis of the *C. parvum*, *C. bovis* and *C. ryanae* proteomes based on sequence similarity identified multiple gene families in clusters (Fig. 3a). Members of AAA proteins formed cluster 1. *Cryptosporidium* species possess a large number of protein kinases, which were included in cluster 2. Clusters 3, 4 and 7 in the network consisted of helicases with the DEAD, HA2 and SNF2 domains, respectively. Ras proteins involved in signalling pathways formed cluster 5. The metallophos domain was found in a diverse range of phosphoesterases, which formed cluster 6. Ubiquitin-conjugating enzymes involved in the second step of ubiquitination formed cluster 8. There are seven members of peptidyl-prolyl *cis-trans* isomerases in each of the three *Cryptosporidium* species, forming cluster 10 in the network. Protein network analysis indicated conservation in the members of these major protein families among *C. parvum*, *C. bovis* and *C. ryanae* (Fig. 3b). We found three unique clusters in *C. parvum*, namely cluster K (FGLN), cluster L (insulinase-like proteases) and cluster M (MEDLE proteins). Proteins containing the RNA recognition motif (cluster C), IMCp domain (cluster E) and ketoacyl synthase domain (cluster O) were only found in *C. bovis* and *C. ryanae*.

Divergent metabolic pathways among intestinal bovine *Cryptosporidium* species

Terpenoid metabolism

In *C. parvum*, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMPP) are two important five-carbon isoprene substrates in terpenoid metabolism. They are synthesized by farnesyl diphosphate (FPP) synthase and polyprenyl synthase (encoded by *cgd4_2550* and *cgd7_3730*, respectively). The genes encoding these two enzymes were shown to have high expression in *C. parvum* according to data in CryptoDB (https://cryptodb.org/), but they are lost in the predicted proteomes of *C. bovis* and *C. ryanae* (Fig. 4b and d) as well as *C. ubiquitum* [11]. In other apicomplexans, IPP biosynthesis is one of the major metabolic pathways in the apicoplast. However, the apicoplast is lost in *Cryptosporidium* species, and the remaining IPP biosynthesis apparently has been further reduced in

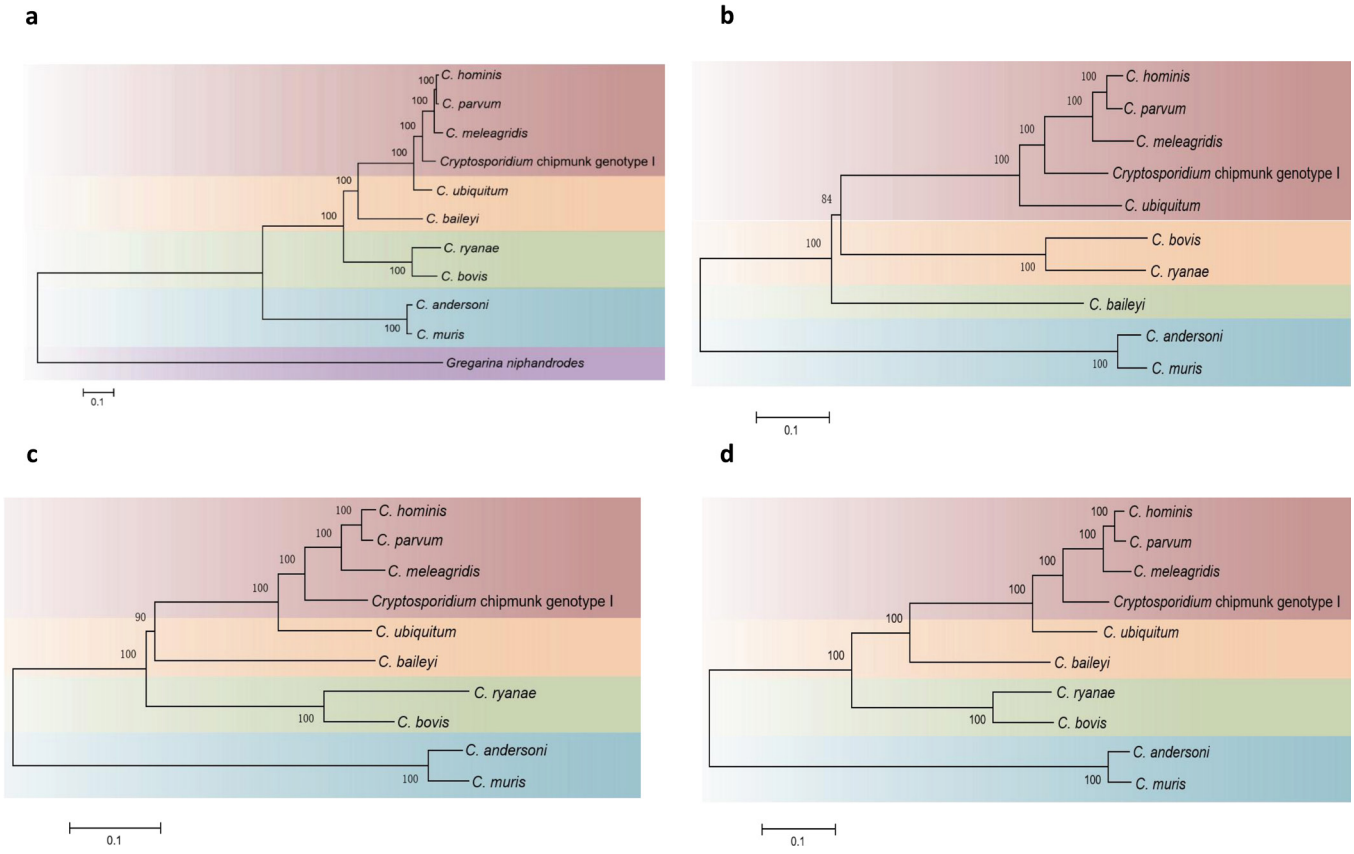

Fig. 2. Phylogenetic relationships among *Cryptosporidium* species. (a) Phylogenetic relationship of *Cryptosporidium* species based on ML analysis of sequences of 100 single-copy orthologous proteins. (b) Phylogenetic relationship of *Cryptosporidium* species based on ML analysis of insulinase-like proteases. (c) Phylogenetic relationship of *Cryptosporidium* species based on ML analysis of TRAP sequences. (d) Phylogenetic relationship of *Cryptosporidium* species based on ML analysis of mucin-type glycoproteins.

some species within the genus. The progressive loss of IPP biosynthesis pathways in *Cryptosporidium* species further confirms that the lipid metabolism in the parasites is not dependent on the apicoplast. Instead, they could salvage the nutrients from the host.

Electron transport chain

A further reduction in the electron transport chain was detected in *C. bovis* and *C. ryanae*. *C. bovis* and *C. ryanae* have lost all genes encoding ATP synthase and the alternative oxidase (AOX) (Table 2). In particular, the gene encoding malate quinone oxidoreductase (MQO) is lost in *C. bovis* and *C. ryanae*, whereas the orthologous genes are present in other *Cryptosporidium* species (Fig. 4). Similarly, the gene encoding the oxoglutarate/malate translocator protein (*cgd1_600* in *C. parvum*) is absent in *C. bovis* and *C. ryanae*.

Coenzyme Q (CoQ), also known as ubiquinone, is involved in transferring electrons from nicotinamide adenine dinucleotide (NADH) dehydrogenase (complex I), MQO and complex II to the cytochrome bc1 complex (complex III). In comparison with *C. parvum*, *C. ubiquitum* has lost four of the eight genes encoding enzymes in CoQ metabolism, while *C. bovis* and *C. ryanae* have lost one additional such gene.

The number of mitochondrial carrier proteins is reduced in *C. bovis* and *C. ryanae* due to simplification of the electron transport system. There are only three genes encoding mitochondrial carrier proteins in *C. bovis* and four in *C. ryanae* (Table 3). In comparison, *C. parvum* has nine such genes while *C. ubiquitum* has six (Table 3). Moreover, the number of triose phosphate transporters (six in *C. bovis* and seven in *C. ryanae*) and ABC transporters (22 in *C. bovis* and 20 in *C. ryanae*) is also different between *C. bovis* and *C. ryanae*.

Nucleotide metabolism

Compared with *C. parvum*, *C. bovis* possesses all 42 orthologous genes encoding enzymes involved in the interconversion of purines and pyrimidines, whereas five such genes are absent in *C. ryanae*. In purine metabolism, the genes encoding inosine monophosphate (IMP) dehydrogenase (*cgd6_20*), guanosine monophosphate (GMP) synthase (*cgd5_4520*) and nucleoside-triphosphate pyrophosphatase (*cgd4_4150*) are absent in *C. ryanae*. The three enzymes are involved in the conversion of IMP to xanthosine 5′-phosphate (XMP),

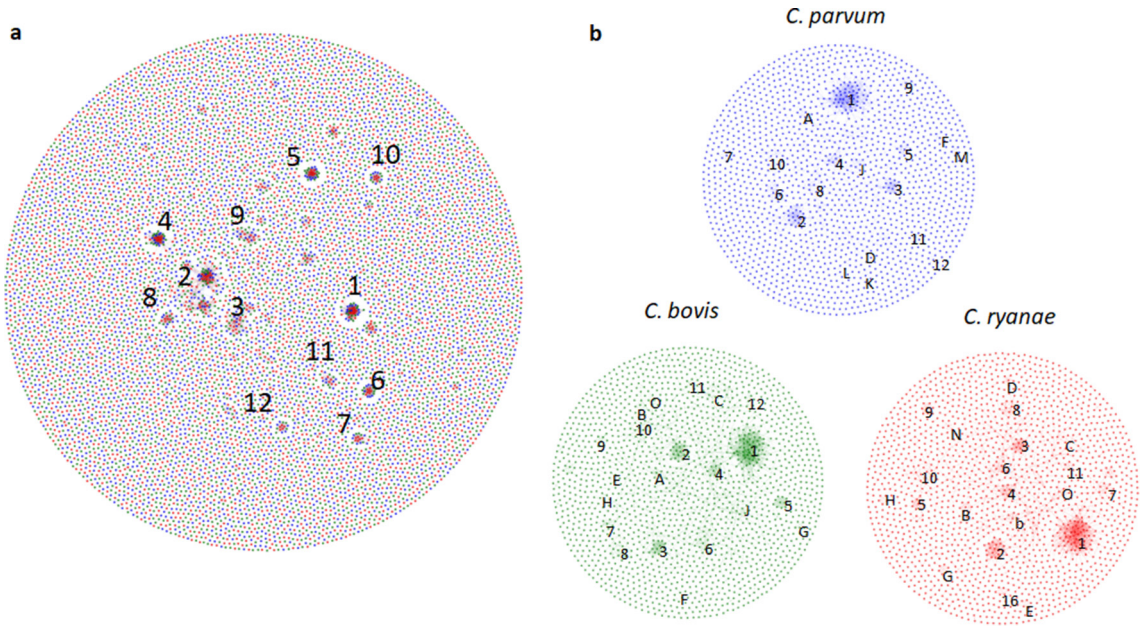

Fig. 3. Protein architecture network of *Cryptosporidium bovis*, *Cryptosporidium ryanae* and *Cryptosporidium parvum*. (a) Protein architecture network based on sequence similarity of all proteins in proteomes of *C. bovis* (green), *C. ryanae* (red) and *C. parvum* (blue). 1, AAA proteins; 2, protein kinase; 3, DEAD; 4, SNF2 family; 5, Ras protein; 6, metallophos; 7, HA2 helicase; 8, ubiquitin-conjugating enzyme; 9, ABC transporter; 10, cyclophilin type peptidyl-prolyl *cis-trans* isomerase; 11, TRAP; 12, zinc finger C3H1-type domain containing protein. (b) Protein architecture network based on sequence similarity of proteins in proteomes of *C. parvum* (blue), *C. bovis* (green) and *C. ryanae* (red). 1, protein kinase; 2, DEAD; 3, AAA proteins; 4, SNF2 family; 5, Ras protein; 6, ABC transporter; 7, TRAP; 8, DnaJ domain; 9, ubiquitin-conjugating enzyme; 10, WD40; 11, cyclophilin type peptidyl-prolyl *cis-trans* isomerase; 12 proteasome; A, metallophos; B, HA2 helicase; C, RNA recognition motif; D, minichromosome maintenance; E, inner membrane complex protein; F, DHHC domain; G, HSP60; H, CCCH type domain; J, ankyrin repeat; K, FGLN; L, IDE; M, MEDLE; N, Sec7 domain; O, beta-ketoacyl synthase.

XMP to GMP, and deoxyguanosine triphosphate (dGTP) to deoxyguanosine monophosphate (dGMP), respectively (Table 2). In pyrimidine metabolism, uracil phosphoribosyltransferase (*cgd4_4460*) and deoxycytidine monophosphate (dCMP) deaminase (*cgd2_2780*) are absent in *C. ryanae*. In other *Cryptosporidium* spcies, uracil is transported into the parasites by nucleobase transporter and catalysed to uridine monophosphate (UMP) by uracil phosphoribosyltransferase (Table 2). The loss of dCMP deaminase indicates that *C. ryanae* does not have the ability to convert dCMP to deoxyuridine monophosphate (dUMP).

Other losses in metabolic pathways

Compared with *C. bovis* and *C. ryanae*, *C. parvum* has 462 species-specific genes, 276 of which encode putative proteins with unknown functions. The genes lost in *C. bovis* and *C. ryanae* encode proteins involved in various metabolic pathways. In amino acid metabolism, the gene encoding tryptophan synthase is present in *C. parvum* (*cgd5_4560*), but absent in *C. bovis* and *C. ryanae*. A gene encoding asparagine synthase A, which could convert aspartate into asparagine, is also absent in *C. bovis* and *C. ryanae*. The orthologue of E3 ubiquitin ligase (*cgd6_2490*), which catalyses the transfer of ubiquitin from the E2 ubiquitin-conjugating enzyme to the protein substrate, was not detected in *C. bovis* or *C. ryanae*, indicating that the protein degradation ability is decreased in these two species. Dynamin is a GTPase involved in endocytosis, division of organelles, cytokinesis and microbial pathogen resistance in eukaryotic cells. The gene encoding dynamin (*cgd8_1990*) in *C. parvum* is absent in *C. bovis* and *C. ryanae*. In addition, three genes encoding ribosomal proteins (*cgd1_300*, *cgd3_2250* and *cgd7_4050*) are lost in *C. bovis* and *C. ryanae*. The gene (*cgd3_2840*) encoding a protein that has two C2H2 zinc fingers and is involved in RNA metabolism is also absent in *C. bovis* and *C. ryanae*. Furthermore, *C. bovis* and *C. ryanae* have lost one member of the polypeptide *N*-acetylgalactosaminyltransferase family and histidine phosphatase superfamily, which each possess two adjacent genes in other intestinal *Cryptosporidium* species.

Gains and losses in subtelomeric genes encoding invasion-related proteins

Compared with other *Cryptosporidium* species, the genes encoding mucin-type glycoproteins have high divergence in *C. bovis* and *C. ryanae* (Table S3). Among them, the gene encoding CP2 (*cgd6_5410*), which is involved in the invasion process and the integrity of the parasitophorous vacuole membrane (PVM), is absent in *C. bovis*, *C. ryanae* and *C. andersoni*. Similarly, the cluster of seven mucin genes (encoding Muc1–Muc7) in the 5′ subtelomeric regions of chromosome 2 in *C. parvum* were not detected in *C. bovis* or *C. ryanae*. In addition, the genes encoding Muc12, Muc14,

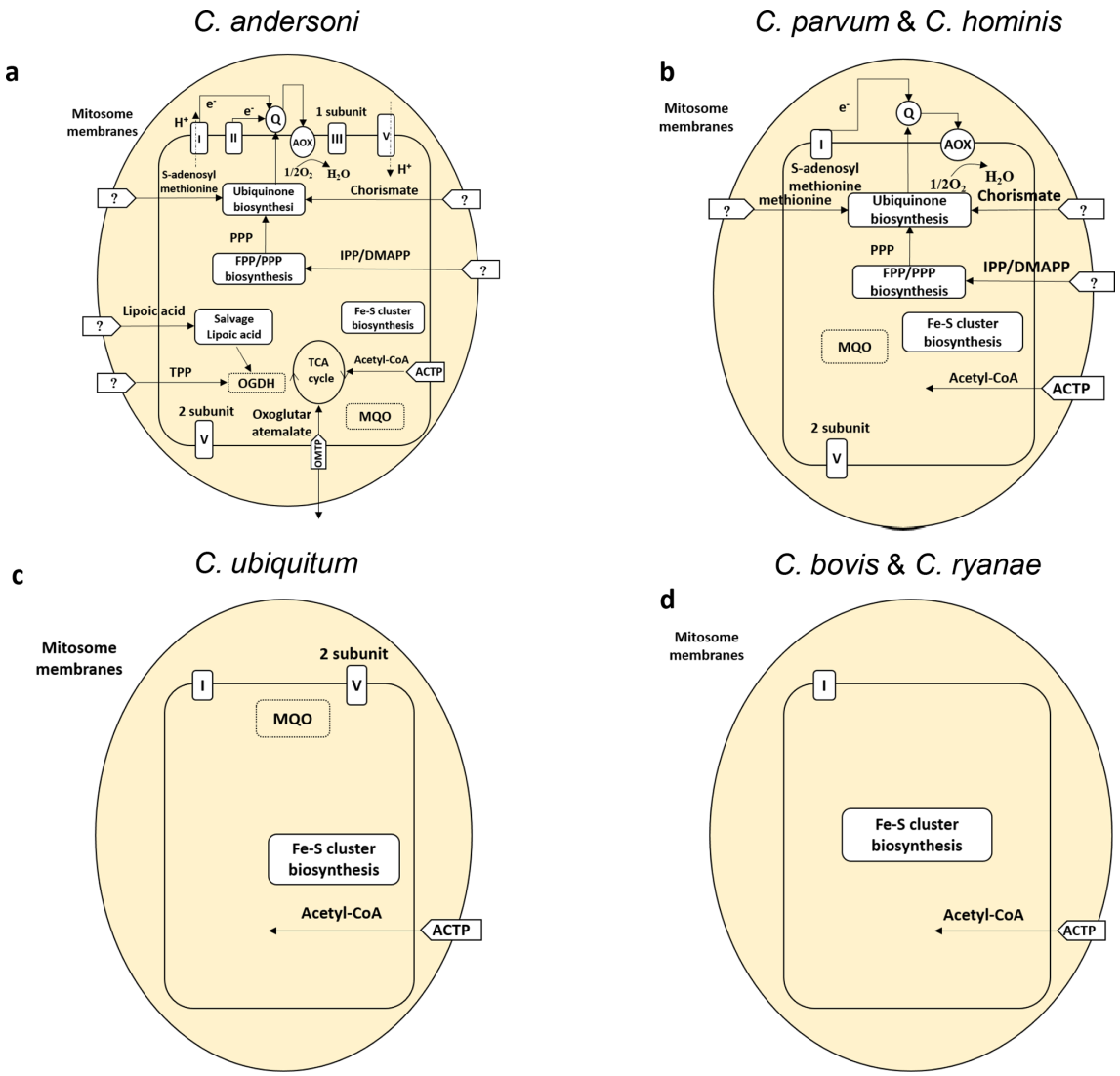

Fig. 4. Mitochondrial metabolism in several *Cryptosporidium* species. Abbreviations of enzymes: AOX, alternative oxidase; MAT, methionine adenosyl transferase; OGDH, oxoglutarate dehydrogenase complex; MQO, malate quinone oxidoreductase. Abbreviations of metabolites: Q, ubiquinone (coenzyme Q); CoA, coenzyme A; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; PPP, polyprenyl diphosphate. Abbreviations of transporter proteins: OMTP, oxoglutarate/malate transporter protein; AATP, amino acids transporter protein; ACTP, acetyl-CoA transporter protein.

Muc17, Muc20 and Muc24 are lost in *C. bovis* and *C. ryanae*. In contrast, *C. bovis* and *C. ryanae* have several genes (*C_bov_6.3221*, *C_bov_8.3556*, *C_bov_4.2822*, *C_bov_4.2823*, *C_bov_42.2912*, *C_bov_6.3080*, *C_bov_8.3622*, *C_bov_8.3594*, *C_bov_1.182*, *C_bov_10.262*, *C_bov_20.1093*, *C_bov_3.2223*, *C_bov_8.3592*, *C_bov_8.3638*, *C_bov_1.152*, *C_rya_29.1908*, *C_rya_26.1661*, *C_rya_6.2899*, *C_rya_45.2592*, *C_rya_23.1311*, *C_rya_23.1284*, *C_rya_11.174*, *C_rya_19.991*, *C_rya_25.1585*, *C_rya_23.1281*, *C_rya_96.3702*) encoding novel mucin-type glycoproteins. Among them, *C_bov_6.3080* and *C_rya_45.2592* are subtelomeric, while *C_bov_4.2822* and *C_bov_4.2823* are adjacent to each other.

Compared with *C. parvum*, three of the 23 genes encoding insulinase-like proteases are lost in *C. bovis* and *C. ryanae* (Table S3). Two of them are *C. parvum*-specific genes located in 3′ subtelomeric regions of chromosomes 5 and 6. Furthermore, the gene (*cgd3_4270*) encoding INS16, which is a paralogue of *cgd3_4260* with 83% amino acid sequence similarity, is absent in *C. ubiquitum*, *C. bovis* and *C. ryanae*, but present in other *Cryptosporidium* species. As in *C. ubiquitum*, *C. baileyi* and *C. andersoni*, *C. bovis* and *C. ryanae* have lost all genes encoding MEDLE family proteins (Table S3).

Gene gains and losses in other multigene protein families

C. bovis and *C. ryanae* have gained members of several multigene families compared with *C. parvum* (Table S4). The WYLE protein family contains secreted proteins with the WYLE

Table 2. Comparison of essential metabolic pathways among *Cryptosporidium* species and other common apicomplexan parasites

Category	Metabolic pathway	Cpar	Chom	Cmel	Cchi	Cubi	Cbov	Crya	Cbai	Cand	Pfal	Tgon
Carbohydrate and energy metabolism	Glycolysis	+	+	+	+	+	+	+	+	+	+	+
	Methylcitrate cycle	–	–	–	–	–	–	–	–	–	–	+
	TCA cycle	–	–	–	–	–	–	–	–	+	+	+
	Pentose phosphate pathway	–	–	–	–	–	–	–	–	–	+	+
	Shikimate biosynthesis	–	–	–	–	–	–	–	–	–	+	+
	Folate biosynthesis	–	–	–	–	–	–	–	–	–	+	+
	Synthesis of pterin	–	–	–	–	–	–	–	–	–	–	+
	Galactose metabolism	–	–	–	–	–	–	–	–	–	–	+
	Synthesis of starch	+	+	+	+	+	+	+	+	+	–	+
	Synthesis of trehalose	+	+	+	+	+	+	+	+	+	–	+
	Synthesis of 1,3-beta-glucan	–	–	–	–	–	–	–	–	–	–	+
	Conversion between UDP-Glc and UDP-Gal	+	+	+	+	+	+	+	+	+	–	+
	Conversion between GDP-Man and GDP-Fuc	–	–	–	–	–	–	–	–	–	+	+
	Conversion from UDP-Glc to UDP-GlcA to UDP-Xyl	+	+	+	+	+	+	+	+	+	–	–
	Synthesis of mannitol from fructose	+	+	+	+	+	+	+	+	+	–	–
	Fatty acid biosynthesis in cytosol (FAS I)	+	+	+	+	+	+	+	+	+	–	+
	Fatty acid biosynthesis in apicoplast (FAS II)	–	–	–	–	–	–	–	–	–	+	+
	Fatty acid degradation	–	–	–	–	–	–	–	–	–	–	+
	Oxidative phosphorylation (NADH dehydrogenase)	+	+	+	+	+	+	+	+	+	+	+
	Oxidative phosphorylation (Complex II)	–	–	–	–	–	–	–	–	+	+	+
	Oxidative phosphorylation (Complex III)	–	–	–	–	–	–	–	–	one sub	+	+
	Oxidative phosphorylation (Complex IV)	–	–	–	–	–	–	–	–	–	+	+
	F-ATPase	two sub	two sub	two sub	two sub	two sub	–	–	two sub	+	+	+
	Alternative oxidase (AOX)	+	+	+	+	–	–	–	–	+	–	–
	Glyoxalase metabolism producing D-lactate	–	–	–	–	–	–	–	–	–	+	+
	Synthesis of isoprene (MEP/DOXP)	–	–	–	–	–	–	–	–	–	+	+
	Synthesis of farnesyl/polyprenyl diphosphate	+	+	+	+	–	–	–	–	+	+	+

Continued

Table 2. Continued

Category	Metabolic pathway	Cpar	Chom	Cmel	Cchi	Cubi	Cbov	Crya	Cbai	Cand	Pfal	Tgon
Nucleotide metabolism	Synthesis of purine rings *de novo*	–	–	–	–	–	–	–	–	–	–	–
	Conversion from IMP to XMP	+	+	+	+	+	+	–	–	–	+	+
	Conversion from XMP to GMP	+	+	–	–	–	+	–	–	–	+	+
	Synthesis of pyrimidine *de novo*	–	–	–	–	–	–	–	–	–	+	+
	Conversion from uracil to UMP	+	+	+	+	+	+	–	+	+	+	+
	Conversion from dCMP to dUMP	+	+	+	+	+	+	–	+	+	+	+
Amino acid metabolism	Synthesis of alanine from pyruvate	–	–	–	–	–	–	–	–	–	–	+
	Synthesis of glutamate from nitrite/nitrate	–	–	–	–	–	–	–	–	–	+	+
	Conversion from glutamate to glutamine	+	+	+	+	+	+	+	+	+	+	+
	Synthesis of aspartate from oxaloacetate and glutamate	–	–	–	–	–	–	–	–	–	+	+
	Conversion from aspartate to asparagine	+	+	+	+	+	–	–	–	–	+	+
	Conversion from glutamate to proline	+	+	+	+	+	+	+	+	+	–	+
	Synthesis of serine from glycerate/glycerol phosphate	–	–	–	–	–	–	–	–	–	–	+
	Conversion from serine to cysteine	–	–	–	–	–	–	–	–	–	–	+
	Conversion from serine to glycine	+	+	+	+	+	+	+	+	+	+	+
	Recycle homocysteine into methionine	–	–	–	–	–	–	–	–	–	+	+
	Synthesis of lysine from aspartate	–	–	–	–	–	–	–	–	–	–	+
	Synthesis of threonine from aspartate	–	–	–	–	–	–	–	–	–	–	+
	Synthesis of ornithine from arginine	–	–	–	–	–	–	–	–	–	+	–
	Synthesis of ornithine from proline	–	–	–	–	–	–	–	–	–	+	+
	Synthesis of polyamine from ornithine	–	–	–	–	–	–	–	–	–	+	–
	Polyamine pathway backward	+	+	+	+	+	+	+	+	+	–	+
	Degradation of branched-chain amino acids	–	–	–	–	–	–	–	–	–	–	+
	Synthesis of tryptophan	+	+	–	+	+	–	–	–	–	–	–
	Aromatic amino acid hydroxylases (AAAH)	–	–	–	–	–	–	–	–	–	–	+

Continued

Table 2. Continued

Category	Metabolic pathway	Cpar	Chom	Cmel	Cchi	Cubi	Cbov	Crya	Cbai	Cand	Pfal	Tgon
Vitamins and others	Synthesis of ubiquinone (coenzyme Q)	+	+	+	+	−	−	−	−	+	+	+
	Synthesis of Fe-S cluster	+	+	+	+	+	+	+	+	+	+	+
	Synthesis of haem	−	−	−	−	−	−	−	−	−	+	+
	Synthesis of thiamine (vitamin B1)	−	−	−	−	−	−	−	−	−	+	−
	Conversion from thiamine to thiamine pyrophosphate (TPP)	−	−	−	−	−	−	−	−	−	+	+
	Synthesis of FMN/FAD from riboflavin	−	−	−	−	−	−	−	−	−	+	+
	Synthesis of pyridoxal phosphate (vitamin B6) *de novo*	−	−	−	−	−	−	−	−	−	+	+
	Synthesis of NAD(P)+ *de novo* from nicotinate/nicotinamide	−	−	−	−	−	−	−	−	−	+	+
	Synthesis of pantothenate from valine	−	−	−	−	−	−	−	−	−	−	+
	Synthesis of CoA from pantothenate	+	+	+	+	+	+	+	+	+	+	+
	Synthesis of lipoic acid *de novo* in apicoplast	−	−	−	−	−	−	−	−	−	+	+
	Salvage of lipoic acid in mitochondria	−	−	−	−	−	−	−	−	+	+	+
	Synthesis of porphyrin/cytochrome proteins	−	−	−	−	−	−	−	−	−	+	+

Plus symbols denote that these metabolic pathways were identified in this apicomplexan parasite, whereas minus symbols denote that these metabolic pathways were absent. Abbreviations: *Cryptosporidium parvum* (Cpar); *C. hominis* (Chom); *C. meleagridis* (Cmel); *Cryptosporidium* chipmunk genotype I (Cchi); *C. ubiquitum* (Cubi); *C. bovis* (Cbov); *C. ryanae* (Crya); *C. baileyi* (Cbai); *C. andersoni* (Cand); *Plasmodium falciparum* (Pfal); *Toxoplasma gondii* (Tgon).
Sub, abbreviation of subunit. One sub means only one subunit of the protein was detected in the species.

Table 3. Putative transporters in *Cryptosporidium* species and other common apicomplexan parasites*

Substrate	Cellular location	Tgon	Pfal	Cand	Cmur	Cpar	ChomUde	Cmel	Cchi	Cubi	Cbov	Crya	Cbai
Hexose		5	2	2	3	2	2	2	2	2	2	2	2
Triose phosphate	Plasma/apicoplast membrane	4	4	8	8	8	8	8	7	8	6	7	7
Amino acids	Plasma membrane	6	1	12	12	10	10	10	10	10	10	10	10
Nucleobase/nucleoside	Plasma membrane	4	4	1	1	1	1	1	1	1	1	1	1
Nucleotide-sugar	Plasma membrane	4	1	2	2	3	3	3	3	3	3	3	2
Folate/pterine	Plasma membrane	7	2	1	1	1	2	1	1	1	1	1	1
Formate/nitrite		3	1	0	0	0	0	0	0	0	0	0	0
GABA (aminobutanoate)	Plasma/mitochondrial membrane	5	2	0	0	0	0	0	0	0	0	0	0
Acetyl-CoA		1	1	1	1	1	1	1	1	1	1	1	1
Chloride		2	0	0	0	0	0	0	0	0	0	0	0
Inorganic phosphate		1	1	0	0	0	0	0	0	0	0	0	0
Sulfate		4	1	1	1	1	1	1	1	1	1	1	1
Sodium/potassium/calcium		9	0	3	3	2	2	2	2	2	2	2	2
Zinc		4	2	2	2	2	2	2	2	2	2	2	2
Copper		3	2	1	1	1	1	1	1	1	1	1	1
Choline	Plasma membrane	2	1	0	0	0	0	0	0	0	0	0	0
Cadmium/zinc/cobalt (efflux)	Plasma membrane	1	1	1	1	1	1	1	1	1	1	1	1
Glycerol/water	Plasma membrane	2	2	0	0	0	0	0	0	0	0	0	0
ABC transporter	Plasma membrane	24	16	21	21	21	21	21	21	21	22	20	22
Mitochondrial carrier	Mitochondrial membrane	21	14	13	12	9	9	8	8	6	3	4	6

*The detection of these transporter proteins was based on the Pfam search.
Tgon: *Toxoplasma gondii*; Pfal: *Plasmodium falciparum*; Cand: *Cryptosporidium andersoni*; Cmur: *C. muris*; Cpar: *C. parvum*; ChomUde: *C. hominis* UdeA01; Cmel: *C. meleagridis*; Cchi: *Cryptosporidium* chipmunk genotype I; Cubi: *C. ubiquitum*; Cbov: *C. bovis*; Crya: *C. ryanae*; Cbai: *C. baileyi.*

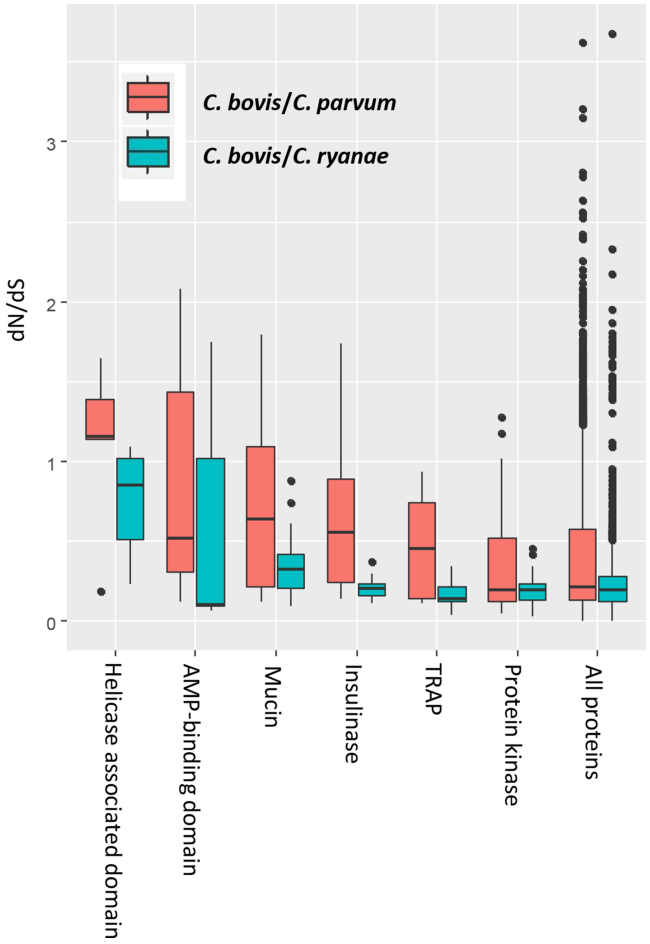

Fig. 5. Selective pressure in multigene families between *Cryptosporidium parvum* and *Cryptosporidium bovis* and between *C. bovis* and *Cryptosporidium ryanae* as indicated by the dN/dS ratio.

sequence in the middle of the proteins. In *C. parvum*, *C. hominis* and *C. meleagridis*, there are six genes encoding WYLE proteins, five of which form a cluster in chromosome 8. Interestingly, three and two additional genes encoding WYLE proteins were detected in the gene cluster in *C. bovis* and *C. ryanae*, respectively. In contrast, only four genes of the WYLE protein family were detected in the gastric species *C. andersoni* and *C. muris*. Furthermore, two genes (*C_bov_31.2447* and *C_bov_31.2452*) encoding secreted leucine-rich proteins form a new gene family in *C. bovis*. One orthologue of the gene, *C_bov_31.2447*, was found in *C. ryanae*. Similarly, two genes (*C_bov_11.434* and *C_bov_18.914*) encoding a new protein family annotated as GPI-anchored adhensin were detected in *C. bovis* with only one orthologue in *C. ryanae*.

More often, members of multigene families are lost in *C. bovis* and *C. ryanae*. The FLGN and SKSR families of secreted proteins are present in all major human-infecting *Cryptosporidium* species. Between them, the FLGN protein family has six, six, six, six and four members in *C. parvum*, *C. hominis*, *C. meleagridis*, *Cryptosporidium* chipmunk genotype I and *C. ubiquitum*, respectively. Similarly, the SKSR protein family has nine, 11, 10, nine and seven members in *C. parvum*, *C. hominis*, *C. meleagridis*, *Cryptosporidium* chipmunk genotype I and *C. ubiquitum,* respectively. None of these FLGN and SKSR genes were detected in *C. bovis* or *C. ryanae*. The NFDQ protein family has three subtelomeric genes (*cgd6_5500*, *cgd5/6_5500* and *cgd8_10*) in *C. parvum*, six in *C. hominis*, four in *Cryptosporidium* chipmunk genotype I, two in *C. meleagridis* and one in *C. ubiquitum*. Among them, only the orthologue of *cgd6_5500* was detected in *C. bovis* (*C_bov_16.739*) and *C. ryanae* (*C_rya_14.480*). Similar to other *Cryptosporidium* species, *C. bovis* and *C. ryanae* have only one orthologue of *cgd8_680_90*, which encodes a large low-complexity protein; a paralogue (*cgd8_660_70*) of *cgd8_680_90* is present in *C. parvum*.

Highly divergent genes between *C. bovis* and *C. ryanae*

We compared the genomes of *C. bovis* and *C. ryanae* and found 46 highly divergent genes encoding proteins with an amino acid identity below 65% (Table S5). Among them, 22 (47.8%) genes encode secreted proteins, 18 (39.8%) encode membrane-bound proteins, 17 (37.0%) are located in the subtelomeric regions and 21 (45.7%) have paralogous genes in *C. bovis*. Notably, *C_bov_10.237* encodes a secreted mucin-like glycoprotein that has only 51.3% sequence identity to the protein encoded by *C_rya_24.1464*; *C_bov_21.1320* encodes a secreted insulinase-like peptidase, which has only 47.5% sequence identity to the homologue in *C. ryanae*; and *C_bov_5.3046* encodes a membrane-associated aspartyl protease with three paralogous genes, and has 59.8% sequence similarity to the homologe in *C. ryanae*. The same is also true for genes encoding oocyst wall protein (*C_bov_26.1848*), ubiquitin-activating enzyme E1 (*C_bov_6.3147*) and secreted low-complexity containing protein (*C_bov_8.3456*). The functions of other proteins involved are unknown.

Genes under selection pressure

The orthologous genes between *C. bovis* and *C. parvum* exhibited elevated dN/dS ratios compared with those between *C. bovis* and *C. ryanae*, especially in the gene families that encode proteins involved in host–pathogen interactions. We found that the gene families encoding helicase-associated domains, AMP-binding domains, protein kinases, mucins, insulinases and TRAPs have higher dN/dS ratios between *C. bovis* and *C. parvum* than between *C. bovis* and *C. ryanae* (Fig. 5). The genes under positive selection between *C. bovis* and *C. parvum* include six helicases, four RNA polymerases, four protein kinases, three insulinase-like peptidases and two ABC transporters (Table 4). The three insulinase-like peptidases under positive selection are in a gene cluster within chromosome 3 in *C. parvum*. The gene *cgd3_4270* also is among them but is lost in *C. bovis* and *C. ryanae*.

DISCUSSION

The results of this study have shown significant differences among the genomes of the three common intestinal

Table 4. Multigene protein families under positive selective pressure between *Cryptosporidium bovis* and *Cryptosporidium ryanae*

Gene family	Gene in *C. parvum*	Gene in *C. bovis*	dN/dS ratio	Annotation
Helicase	*cgd1_2650*	*C_bov_13.593*	1.64629	SNF2 helicase
	cgd6_1410	*C_bov_13.593*	1.64629	Pre-mRNA splicing factor ATP-dependent RNA helicase
	cgd6_3860	*C_bov_25.1726*	1.12968	SNF2 helicase
	cgd7_640	*C_bov_4.2704*	1.38836	Prp16p pre-mRNA splicing factor. HrpA family SFII helicase
	cgd8_2770	*C_bov_42.2905*	1.08081	SNF2L orthologue with an SWI/SNF2 like ATPase and an Myb domain
	cgd8_4100	*C_bov_13.593*	1.64629	PRP43 involved in spliceosome disassembly mRNA splicing
Insulinase-like peptidase	*cgd3_4250*	*C_bov_21.1320*	1.73579	Secreted insulinase-like peptidase
	cgd3_4260	*C_bov_21.1321*	1.28419	Secreted insulinase-like peptidase
	cgd3_4280	*C_bov_21.1322*	1.11455	Secreted insulinase-like peptidase
Protein kinase	*cgd5_250*	*C_bov_24.1656*	1.01196	Ser/Thr protein kinase
	cgd5_3180	*C_bov_17.879*	1.27347	Ser/Thr protein kinase
	cgd6_4960	*C_bov_30.2379*	1.01594	Ser/Thr protein kinase
	cgd6_540	*C_bov_23.1582*	1.17354	Ser/Thr protein kinase
ABC transporter	*cgd2_90*	*C_bov_6.3084*	1.80151	ABC transporter with 9× transmembrande domains and 2× AAA
	cgd4_4440	*C_bov_27.1928*	1.19132	ABC transporter with 9× transmembrande domains and 2× AAA
RNA polymerase	*cgd7_3720*	*C_bov_6.3158*	1.75622	RNA polymerase beta subunit
	cgd8_170	*C_bov_10.307*	1.28553	DNA-directed RNA polymerase beta subunit
	cgd3_2620	*C_bov_20.1075*	1.60633	DNA-directed RNA polymerase, possible RNA polymerase
	cgd6_3290	*C_bov_36.2567*	1.60829	DNA-directed RNA polymerase III C1 subunit
Acyl transferase domain	*cgd3_2180*	*C_bov_14.664*	2.0781	Type I fatty acid synthase
	cgd4_2900	*C_bov_36.2532*	2.03806	Polyketide synthase

Cryptosporidium species in bovine animals. The nucleotide and amino acid sequence identities between *C. bovis* and *C. parvum* are 38.6 and 55.1%, respectively, while those between *C. bovis* and *C. ryanae* are 69.4 and 75.2%, respectively. In contrast, the nucleotide and amino acid sequence identities between *Cryptosporidium* chipmunk genotype I and other major human-pathogenic species such as *C. hominis*, *C. parvum*, *C. meleagridis* and *C. ubiquitum* are 78.7–82.5 and 79.0–84.0%, respectively [16]. These genomic differences among *Cryptosporidium* species are in agreement with their phylogenetic relationship (Fig. 2). They could contribute to the differences in human infectivity and pathogenicity among intestinal *Cryptosporidium* species.

Accompanying the significant sequence differences is a reduction in synteny in gene organization between the *C. bovis*/*C. ryanae* and *C. parvum* genomes. Compared with the large syntenic regions among *C. hominis*, *C. parvum* and *Cryptosporidium* chipmunk genotype I, the syntenic regions between *C. bovis*/*C. ryanae* and *C. parvum* are more fragmented. Blocks of rearrangements and deletions were observed in some chromosomes between *C. bovis* and *C. parvum*, especially in the subtelomeric regions, leading to losses in the former of some subtelomeric genes encoding secreted proteins. Breaks in genome synteny are common in other apicomplexans, leading to the losses of multigene families and species-specific genes [42].

Compared with *C. parvum* and other human-pathogenic intestinal *Cryptosporidium* species, *C. bovis* and *C. ryanae* appear to have more streamlined metabolism. The gene content of the *C. bovis* and *C. ryanae* genomes is smaller than that of the *C. hominis* and *C. parvum* genomes. There are nearly 3300 genes shared by all intestinal *Cryptosporidium* species. Compared with *C. parvum*, the genes lost in *C. bovis* mostly encode metabolism-related enzymes and secreted proteins. The loss of enzymes involved in the metabolic pathways leads to further reduced biosynthesis capacity and energy production in *C. bovis* and *C. ryanae*. As a result, these two parasites could be more dependent on specific hosts to salvage nutrients. Previous studies have shown a progressive reduction in the electron transport chain in *Cryptosporidium* species [11]. The loss of the genes encoding ATP synthase and MQO in *C. bovis* and *C. ryanae* has provided new evidence for progressive reduction in energy production within the genus *Cryptosporidium*. Variations in metabolism are thought to contribute to lineage-specific adaptation to the host environment and virulence of apicomplexan parasites. In *Toxoplasma*

gondii, altered capacity for energy production is associated with strain-specific differences in growth rates and virulence across different hosts, organs and cell types [43]. Because of the importance of some metabolic pathways in pathogen growth and survival, they could be potential targets for drug development, such as isoprenoid biosynthesis [44] and the shikimate pathway [45]. MQO could be such a potential target against *C. parvum*, but not against *C. bovis* or *C. ryanae*.

A major difference among the three bovine intestinal *Cryptosporidium* species is in the number of mucin-type glycoproteins, which are important SPDs involved in the attachment of sporozoites to the host cells [46]. *C. bovis* and *C. ryanae* have lost a series of mucin-type glycoproteins, including CP2, Muc1–Muc7, Muc12, Muc14, Muc17 and Muc20. In addition to the loss of mucin-type glycoproteins, several novel mucin-type glycoproteins were observed in *C. bovis* and *C. ryanae*. Thus, Muc25–Muc39 have no orthologues in *C. parvum* and most of them are present in both *C. bovis* and *C. ryanae*. These copy number variations in mucin-type glycoproteins could potentially contribute to the phenotypic differences among intestinal *Cryptosporidium* species, such as variations in growth rate of the pathogens and duration and intensity of infections [46].

Similarly, subtelomeric genes encoding other invasion-associated proteins, such as secreted MEDLE proteins and insulinase-like proteases, are also divergent among *C. parvum*, *C. bovis* and *C. ryanae*. Three insulinase-like proteases are lost in *C. bovis* and *C. ryanae*, two of which are in the subtelomeric region and one is in the multigene cluster. Similarly, genes encoding MEDLE proteins located in the subtelomeric region are completely absent in *C. bovis* and *C. ryanae*. The number of invasion-related proteins is known to be different among apicomplexans. For example, *Neospora caninum* and *Sarcocystis neurona* have 227 and 23 SAG1-related sequences (SRS), respectively, which are involved in modulation of host immune responses [47]. Similarly, *Toxoplasma gondii* strains Me49 (less virulent) and GT1 (more virulent) have 109 and 90 such genes, respectively [48]. *Theileria parva* and *Theileria annulata* are known to have different numbers (85 and 51 members, respectively) of genes encoding subtelomeric variable secreted proteins (SVSPs) [47], which could contribute to differences in host range and pathogenicity between the two species. *Cryptosporidium* species do not have homologous proteins of these families, but subtelmoeric genes encoding secreted proteins account for the majority of multigene families in their genomes. They were previously suggested to be SPDs in *Cryptosporidium* species [16].

Our comparative genomics analysis has revealed some gains and losses of other potential SPDs among the three bovine *Cryptosporidium* species. They include genes encoding secreted WYLE, FLGN, SKSR and NFDQ proteins. Previous studies have suggested that differences in pathogenicity, transmission modes and host range among *Toxoplasma gondii* strains could have been caused by differences in copy numbers of genes encoding SRS proteins and secretory proteins from micronemes (MICs), dense granules (GRAs) and rhoptries (ROPs), which appear to be SPDs in *Toxoplasma gondii* [49]. In *Cryptosporidium* species, differences in copy numbers of genes encoding SKSR proteins have been observed between *C. parvum* IIa and IId subtype families [17]. The subtelomeric genes encoding these SPDs, except for those encoding WYLE proteins, are mostly lost in *C. bovis* and *C. ryanae*. In contrast, the latter two species have additional members of WYLE proteins, which could contribute to the biological uniqueness of these two bovine *Cryptosporidium* species.

Compared with other *Cryptosporidium* species, *C. bovis* and *C. ryanae* have similar gene contents and the closest genetic relationship. Minor differences in gene content between the two species include genes encoding enzymes in nucleotide metabolism, ABC transporters, mitochondrial carriers, mucin-type glycoproteins and several hypothetical proteins. However, 46 genes with highly divergent sequences are present between *C. bovis* and *C. ryanae*. Half of the highly divergent genes between *C. bovis* and *C. ryanae* encode secreted proteins or membrane-bound proteins and one-third of the highly divergent genes are located in the subtelomeric regions. While most of the genes encode proteins with unknown functions, some are specific to *C. bovis* and *C. ryanae*, including members of invasion-related protein families, ubiquitin-activating enzymes and oocyst wall proteins. More functional studies on these proteins are needed to understand the importance of the sequence divergence between these two species.

The elevated dN/dS ratios in the orthologous genes between *C. bovis* and *C. parvum* reveal a divergence in the evolution between these two species. The positive selection identified in some multigene families could be a reflection of the proteins encoded by the genes in host specificity and pathogenicity between the two species. In addition to the gains and losses of invasion-related protein families between *C. parvum* and *C. bovis*, some members of these families are also under positive selection, including three insulinase-like peptidases located in a gene cluster within chromosome 3. Previous studies have shown that only a few orthologous genes are under positive selection among closely related *Cryptosporidium* species [50, 51], and most of them are located in the subtelomeric regions. Between *C. parvum* and *C. bovis*, however, some of the positively selected genes are distributed in various parts of the chromosomes. Furthermore, multiple gene families encoding helicases and polymerases are also among those with high dN/dS ratios. These genes are relatively conservative between *C. parvum* and other intestinal *Cryptosporidium* species sequenced thus far. Sequence polymorphisms in these genes could affect the efficiency of transcription and translation, leading to the divergence in biological characteristics between *C. parvum* and *C. bovis*. More transcriptomic and proteomic studies of *Cryptosporidium* species are needed to understand the significance of this finding. As expected, two genes encoding ABC transporters are also under positive selection, which could be involved in endobiotic and xenobiotic detoxification [52]. They could be potential targets for drug development.

In conclusion, *C. bovis* and *C. ryanae* apparently have high similarities in gene organization, metabolic pathways and SPDs. They have reduced metabolic capacity compared with *C. parvum* and other *Cryptosporidium* species. The loss of some mucin-type glycoproteins and insulinase-like proteases and all six secreted MEDLE proteins could potentially be responsible for the narrowed host range of *C. bovis* and *C. ryanae*. The loss of some other SPDs such as FLGN, SKSR and NFDQ proteins might contribute to the reduced pathogenicity of *C. bovis* and *C. ryanae*. Highly divergent genes encoding secreted and surface-associated proteins could contribute to the biological differences between *C. bovis* and *C. ryanae*. These hypotheses, however, should be examined in future studies using the functional genomics approach to confirm the findings from comparative genomics. Multiple isolates of *C. bovis* and *C. ryanae* should be sequenced and analysed to support some of the conclusions. This will probably lead to improved understanding of determinants of the host specificity and pathogenicity of different *Cryptosporidium* species.

Funding information
This work was supported in part by the National Natural Science Foundation of China (31 630 078 and 31602042), National Key R and D Program of China (2017YFD0500404), and the 111 Project (D20008).

Author contributions
Conceptualization: Y.F. and L.X.; methodology: Z.X. and L.X.; formal analysis: Z.X. and L.X.; investigation and resources: N.L. and Y.G.; writing – original draft preparation: Z.X.; writing – review and editing: L.X. and Y.F.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
This study was approved by the Ethics Committee of the East China University of Science and Technology. Faecal specimens from dairy cattle were collected with the permission of the farm manager. During specimen collection, cattle were handled in accordance with the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Data Bibliography
1. Lihua Xiao. NCBI BioProject PRJNA545588 and PRJNA545579 (2019).

2. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, *et al*. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441-445 NCBI BioProject PRJNA15586 (2004).

3. Isaza JP, Galvan AL, Polanco V, Huang B, Matveyev AV, Serrano MG, Manque P, Buck GA, Alzate JF. Revisiting the reference genomes of human pathogenic *Cryptosporidium* species: reannotation of *C. parvum* Iowa and a new *C. hominis* reference. *Sci Rep* 5, 16324 European *Nucleotide Archive* PRJEB10000 (2015).

4. Ifeonu OO, Chibucos MC, Orvis J, Su Q, Elwin K, Guo F, Zhang H, Xiao L, Sun M, Chalmers RM *et al*. 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. NCBI BioProject PRJNA222838 and PRJNA22283574 (2016).

5. Liu S, Roellig DM, Guo Y, Li N, Frace MA, Tang K, Zhang L, Feng Y, Xiao L. Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. *BMC Genomics* 17, 1006 NCBI BioProject PRJNA246478 and PRJNA308889 (2016).

6. Xu, Z., Guo, Y., Roellig, D.M., Feng, Y., Xiao, L. 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. NCBI BioProject PRJNA511361 2019.

References

1. Checkley W, White AC, Jaganath D, Arrowood MJ, Chalmers RM *et al*. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. *Lancet Infect Dis* 2015;15:85–94.
2. Chalmers RM, Davies AP. Minireview: clinical cryptosporidiosis. *Exp Parasitol* 2010;124:138–146.
3. Santín M. Clinical and subclinical infections with *Cryptosporidium* in animals. *N Z Vet J* 2013;61:1–10.
4. Xiao L. Molecular epidemiology of cryptosporidiosis: an update. *Exp Parasitol* 2010;124:80–89.
5. Feng Y, Ryan UM, Xiao L. Genetic diversity and population structure of *Cryptosporidium*. *Trends Parasitol* 2018;34:997–1011.
6. Rieux A, Paraud C, Pors I, Chartier C. Molecular characterization of *Cryptosporidium* isolates from beef calves under one month of age over three successive years in one herd in western France. *Vet Parasitol* 2014;202:171–179.
7. Fayer R, Santin M, Trout JM. Prevalence of *Cryptosporidium* species and genotypes in mature dairy cattle on farms in eastern United States compared with younger cattle from the same locations. *Vet Parasitol* 2007;145:260–266.
8. Santín M, Trout JM, Fayer R. A longitudinal study of cryptosporidiosis in dairy cattle from birth to 2 years of age. *Vet Parasitol* 2008;155:15–23.
9. Ralston B, Thompson RCA, Pethick D, McAllister TA, Olson ME. *Cryptosporidium andersoni* in Western Australian feedlot cattle. *Aust Vet J* 2010;88:458–460.
10. Guo Y, Tang K, Rowe LA, Li N, Roellig DM *et al*. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *BMC Genomics* 2015;16:320.
11. Liu S, Roellig DM, Guo Y, Li N, Frace MA *et al*. Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. *BMC Genomics* 2016;17:1006.
12. Fei J, Wu H, Su J, Jin C, Li N *et al*. Characterization of MEDLE-1, a protein in early development of *Cryptosporidium parvum*. *Parasit Vectors* 2018;11:312.
13. Su J, Jin C, Wu H, Fei J, Li N *et al*. Differential expression of three *Cryptosporidium* species-specific MEDLE proteins. *Front Microbiol* 2019;10:1177.
14. Zhang S, Wang Y, Wu H, Li N, Jiang J *et al*. Characterization of a species-specific insulinase-like protease in *Cryptosporidium parvum*. *Front Microbiol* 2019;10:354.
15. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. *Cryptosporidium* pathogenicity and virulence. *Clin Microbiol Rev* 2013;26:115–134.
16. Xu Z, Guo Y, Roellig DM, Feng Y, Xiao L. 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* 2019;20:406.
17. Feng Y, Li N, Roellig DM, Kelley A, Liu G *et al*. Comparative genomic analysis of the IId subtype family of *Cryptosporidium parvum*. *Int J Parasitol* 2017;47:281–290.
18. Zhang S, Chen L, Li F, Li N, Feng Y *et al*. Divergent Copies of a *Cryptosporidium parvum*-Specific Subtelomeric Gene. *Microorganisms* 2019;7:E366:366.
19. Nader JL, Mathers TC, Ward BJ, Pachebat JA, Swain MT *et al*. Evolutionary genomics of anthroponosis in *Cryptosporidium*. *Nat Microbiol* 2019;4:826–836.
20. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G *et al*. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 2004;304:441–445.

21. Ifeonu OO, Chibucos MC, Orvis J, Su Q, Elwin K *et al*. 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* 2016;74:ftw080.
22. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM *et al*. The genome of *Cryptosporidium hominis*. *Nature* 2004;431:1107–1112.
23. Xiao L, Escalante L, Yang C, Sulaiman I, Escalante AA *et al*. Phylogenetic analysis of *Cryptosporidium* parasites based on the small-subunit rRNA gene locus. *Appl Environ Microbiol* 1999;65:1578–1583.
24. Guo Y, Cebelinski E, Matusevich C, Alderisio KA, Lebbad M *et al*. Subtyping novel zoonotic pathogen *Cryptosporidium* chipmunk genotype I. *J Clin Microbiol* 2015;53:1648–1654.
25. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;5:e11147.
26. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–1645.
27. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 2005;33:6494–6506.
28. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 2004;32:W309–W312.
29. Parra G, Blanco E, Guigó R. Geneid in Drosophila. *Genome Res* 2000;10:511–515.
30. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* 2008;9:R7.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
32. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 2011;8:785–786.
33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001;305:567–580.
34. Fankhauser N, Mäser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* 2005;21:1846–1852.
35. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35:W182–W185.
36. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY *et al*. Pfam: the protein families database. *Nucleic Acids Res* 2014;42:D222–D230.
37. Shanmugasundram A, Gonzalez-Galarza FF, Wastling JM, Vasieva O, Jones AR. Library of apicomplexan metabolic pathways: a manually curated database for metabolic pathways of apicomplexan parasites. *Nucleic Acids Res* 2013;41:D706–D713.
38. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13:2178–2189.
39. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
40. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;17:540–552.
41. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005;21:456–463.
42. DeBarry JD, Kissinger JC. Jumbled genomes: missing apicomplexan synteny. *Mol Biol Evol* 2011;28:2855–2871.
43. Song C, Chiasson MA, Nursimulu N, Hung SS, Wasmuth J *et al*. Metabolic reconstruction identifies strain-specific regulation of virulence in *Toxoplasma gondii*. *Mol Syst Biol* 2013;9:708.
44. Moreno SNJ, Li Z-H. Anti-infectives targeting the isoprenoid pathway of *Toxoplasma gondii*. *Expert Opin Ther Targets* 2008;12:253–263.
45. McConkey GA, Pinney JW, Westhead DR, Plueckhahn K, Fitzpatrick TB *et al*. Annotating the *Plasmodium* genome and the enigma of the shikimate pathway. *Trends Parasitol* 2004;20:60–65.
46. Lendner M, Daugschies A. *Cryptosporidium* infections: molecular advances. *Parasitology* 2014;141:1511–1532.
47. Reid AJ. Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa. *Parasitology* 2015;142 Suppl 1:S57–S70.
48. Wasmuth JD, Pszenny V, Haile S, Jansen EM, Gast AT *et al*. Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of *Toxoplasma* virulence. *mBio* 2012;3:e00321-12.
49. Lorenzi H, Khan A, Behnke MS, Namasivayam S, Swapna LS *et al*. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nat Commun* 2016;7:10147.
50. Mazurie AJ, Alves JM, Ozaki LS, Zhou S, Schwartz DC *et al*. Comparative genomics of *Cryptosporidium*. *Int J Genomics* 2013;2013:832756–.
51. Isaza JP, Galván AL, Polanco V, Huang B, Matveyev AV *et al*. Revisiting the reference genomes of human pathogenic *Cryptosporidium* species: reannotation of *C. parvum* Iowa and a new *C. hominis* reference. *Sci Rep* 2015;5:16324.
52. Zapata F, Perkins ME, Riojas YA, Wu TW, Le Blancq SM. The *Cryptosporidium parvum* ABC protein family. *Mol Biochem Parasitol* 2002;120:157–161.