Review

Investigating *Cryptosporidium* spp. Using Genomic and Proteomic Techniques: Current Progress and Future Directions

Joanna Dąbrowska * , Jacek Sroka , Tomasz Cencek

Posted Date: 13 July 2023

doi: 10.20944/preprints202307.0870.v1

Keywords: Cryptosporidium spp.; genomics; proteomics

Review

Investigating *Cryptosporidium* spp. Using Genomic and Proteomic Techniques: Current Progress and Future Directions

Joanna Dąbrowska *, Jacek Sroka and Tomasz Cencek

Deparment of Parasitology and Invasive Disease, National Veterinary Research Institute Partyzantów 57 Avenue, 24-100 Puławy; joanna.dabrowska@piwet.pulawy.pl; jacek.sroka@piwet.pulawy.pl; tcencek@piwet.pulawy.pl

* Correspondence: joanna.dabrowska@piwet.pulawy.pl; Tel.: +48-81-889-30-25

Abstract: Cryptosporidiosis is a widespread disease caused by the parasitic protozoan *Cryptosporidium* spp., which infects various vertebrate species, including humans. Once unknown as a gastroenteritis-causing agent, *Cryptosporidium* spp. is now recognized as a pathogen causing life-threatening disease, especially in immunocompromised individuals such as AIDS patients. Advances in diagnostic methods and increased awareness have led to a significant shift in the perception of *Cryptosporidium* spp. as a pathogen. Nowadays, genomic and proteomic studies have played a main role in understanding the molecular biology of this complex-life-cycle parasite. Genomics has enabled the identification of numerous genes involved in the parasite's development and interaction with hosts. Proteomics has allowed for the identification of protein interactions, their function, structure, and cellular activity. The combination of these two approaches has significantly contributed to the development of new diagnostic tools, vaccines, and drugs for cryptosporidiosis. In this review, we summarize the major accomplishments in the field of *Cryptosporidium* research using genomics and proteomics.

Keywords: *Cryptosporidium* spp.; genomics; proteomics

1. Introduction

Cryptosporidium spp. is a protozoan parasite capable of infecting various vertebrates, including humans, causing gastroenteritis with symptoms such as watery diarrhea, abdominal pain, and weakness [1]. Immunocompromised individuals are particularly vulnerable to this parasite, which has a complex life cycle involving asexual and sexual reproduction within a single host and transmission through direct and indirect contact - animal-to-human, water, and food [2]. Of the over 39 species within the genus *Cryptosporidium*, *C. parvum* and *C. hominis* are the most common in humans, responsible for 90% of cases worldwide. The distribution of *Cryptosporidium* species varies across different regions globally, with *C. hominis* being the main causative agent in developing and industrialized nations, and *C. parvum* more frequently detected in humans from Middle Eastern countries. Some species, such as *C. parvum* and *C. meleagridis*, can infect both mammals and birds, while others, including *C. parvum*, *C. ubiquitum*, and *C. muris*, can cause zoonotic mixed infections [3].

Cryptosporidium was first identified in mice in 1907, and it emerged as a significant causative agent of severe diarrhea in patients with AIDS in the 1980s and 1990s [4]. Cryptosporidiosis is a rising infectious disease of global public health significance, with prevalence estimates ranging from 2.6 to 21.3% in Africa and from 0.3 to 4.3% in North America [5]. Diagnosis of cryptosporidiosis is based on microscopic identification of parasite oocysts, oocyst antigens, or oocyst DNA in fecal samples. However, due to the lack of specific morphological differences between parasite species, molecular tests using genetic markers such as SSU rRNA, COWP, hsp70, and gp60 are widely accepted as the standard for species or genotype identification. It is important to note that relying solely on a few genetic markers for diagnosis may be insufficient and inaccurate, especially when identifying mixed

 doi:10.20944/preprints202307.0870.v1

infections with different species and genotyping. Therefore, newer molecular tools such as whole-genome sequencing are being employed to better understand the genetic structure of *Cryptosporidium* and develop new diagnostic methods [6].

In addition, identifying proteins involved in the host-pathogen interaction and discovering potential drug targets can be helpful in reducing the parasite's presence in the environment. Despite extensive efforts to develop experimental models and evaluate nearly 1,000 chemotherapeutic agents, there is still no fully effective anti-*Cryptosporidium* therapy. [7]. The reason for the lack of drug efficacy is likely multifaceted and may include the parasite's location within the host cell, distinct structural and biochemical composition of drug targets, or its ability to block the import or rapidly efflux drug molecules. Therefore, understanding the basic mechanisms by which drugs are transported to the parasite and identifying unique targets are critical steps in developing effective therapeutic agents. In this regard, proteomic studies can be beneficial.

This review highlights the significant milestones in *Cryptosporidium* research that have been made possible through the application of genomics and proteomics techniques over the years.

2. Genomics of *Cryptosporidium* spp.

2.1. Exploring Cryptosporidium Diversity and Evolutionary History through Whole Genome Sequencing

The study of *Cryptosporidium* spp. was not a prominent research area until the 1980s, and as a result, genome studies began relatively late. The first attempts at genome analysis were made in 1999 by Liu et al.[8], this is considered to be an initial step but was not entirely successful in providing a comprehensive understanding of this parasite on the genomic level. The genome survey identified over 250 kb of the *C. parvum* Iowa (isolate originally obtained from C. Sterling, University of Arizona, Tucson) genome by identifying genes independent of their developmental expression using a random sequence analysis. A total of 442 clones were selected and subjected to automated sequencing using one (230 clones) or two primers (212 clones) flanking the cloning site. This method generated 654 genomic survey sequences that were assembled into 408 contigs containing over 250 kb of unique sequence, representing approximately 2.5% of the *C. parvum* genomic sequence. The comparison of these sequences revealed that 107 contigs (26%) displayed similarity to previously identified proteins, rRNA and tRNA genes, such as putative genes involved in the glycolytic pathway, DNA, RNA, and protein metabolism, and signal transduction pathways. Additionally, diverse microsatellite sequences were identified, constituting less than 1% of the genome and having potential as genetic markers for strain typing. These studies provided first comprehensive molecular insights into the fundamental biology and cellular metabolism of this apicomplexan parasite, which has been traditionally difficult to study experimentally. Unfortunately, the authors' analysis was constrained by the limited availability of gene sequences in the public databases at the time, which restricted the amount of data they could analyze.

The first cDNA sequence survey of *C. parvum* sporozoites was published one year later by Strong et al. [9]. To expedite gene discovery and identify potential drug and vaccine targets, the authors constructed cDNA and genomic DNA sequencing libraries from the Iowa isolate of *C. parvum* and generated approximately 2000 sequence tags by single-pass sequencing of random clones. Analysis of the resulting 567 expressed sequence tags (ESTs) and 1507 genome survey sequences (GSSs) yielded one megabase (1 mb) of unique genomic sequence, representing approximately 10% of the 10.4 mb *C. parvum* genome. The results of the studies were crucial and provided a basis for further analysis, particularly as they were conducted before the entire genome was known. Despite this, BLAST analyses revealed that 32% of the ESTs exhibited similarity to sequences already present in publicly available databases, indicating a significant proportion of the sequences had
a known homology and function. Furthermore, some ESTs encoded proteins with signal peptides or multiple transmembrane spanning segments, indicating their potential of being located in membrane-bound compartments, the cell surface, or extracellular secretion. The authors identified several tags encoding proteins with clear therapeutic potential, including S-adenosylhomocysteine hydrolase, histone deacetylase, polyketide/fatty-acid synthases, various cyclophilins,

 doi:10.20944/preprints202307.0870.v1

thrombospondin-related cysteine-rich protein, and ATP-binding-cassette transporters. The identification of these molecules was groundbreaking because many of them e.g., transporters, may be involved in the rapid efflux of drugs as well as nutrient uptake.

To identify new genetic markers for future epidemiological studies and to expand molecular diagnostics of *Cryptosporidium* spp., whole genome sequencing projects were conducted in the following years. The first two projects involved harvesting oocyst DNA from infected germ-free neonatal calves and utilizing the random shotgun Sanger sequencing approach. Notably, both studies served as the starting point for the genomic era of *Cryptosporidium* research, paving the way for extensive future analyses.

In 2004, Abrahamsen et al. [10] conducted a comprehensive survey of the *Cryptosporidium parvum* Iowa type II full genome using a shotgun approach with Sanger sequencing. Notably, the genome of the parasite was fully assembled at the chromosome level by using a HAPPY map to create scaffold and ordering the sequence contigs. The analysis revealed a highly compact genome with a total size estimated at 9.1 Mb, which was comprised of eight chromosomes ranging from ~0.9 to 1.4 Mb. The genome had a 13-genome coverage with five gaps. Gene prediction and annotation of open reading frames (ORFs) facilitated the identification of 3.807 protein-encoding genes in the *C. parvum* genome. Interestingly, only about 5% of genes possessed introns, and the parasite had smaller intergenic spaces than related microorganisms. One of the key takeaways from this study was that the parasite does not have extranuclear genomes, which was confirmed for the first time. Furthermore, the sequence analysis strongly suggested that *Cryptosporidium* lacks an apicoplast, which differs from other apicomplexans. Similarly, a degenerate mitochondrion that had lost its genome was identified. Moreover, the analysis revealed novel classes of cell-surface and secreted proteins with potential roles in pathogenesis, such as var and rifin molecules, SKSR, MEDLE, WYLE, FGLN, and GGC. The studies demonstrated that the parasite have a unique metabolic profile characterized by a limited number of genes associated with anabolic and catabolic processes. These rely solely on anaerobic glycolysis to generate ATP through the degradation of simple sugars and lack both a mitochondrial genome and numerous nuclear genes related to the Krebs cycle or electron transport chain. Additionally, the parasite has largely lost the ability to synthesize amino acids de novo. The authors also identified a set of four paralogous proteins belonging to the sbmA family that are involved in the transport of peptide antibiotics in bacteria and which can be potential novel routes for drug target and drug discovery. Furthermore, *C. parvum* lacked the group of variant surface antigens used for immune evasion by other Apicomplexa.

In 2004, Xu et al. [11] published a results of a genomic analysis of *Cryptosporidium hominis* isolate TU502, which revealed striking similarities in gene complement with *C. parvum*, exhibiting only 3-5% sequence divergence at the level of micro- and minisatellite repeat lengths and differential gene regulation. The *C. hominis* genome comprises eight chromosomes, with a higher GC content of 31.7% compared with 30.3% for *C. parvum*. Gene prediction identified 3994 genes, of which 60% exhibit similarity to known nucleotide sequences and 5-20% possess introns. Repeated DNA sequences of 2-50-base-pair (bp) were found towards the chromosome ends. Like *C. parvum*, *C. hominis* uses a glycolysis-based metabolism, but it lacks enzymes for the synthesis of key biochemical processes such as simple sugars, amino acids, and nucleotides. However, the parasite can produce more complex compounds such as amylopectin, fatty acids, and starch from precursor molecules, utilizing enzymes that have minimal similarity to known biosynthetic enzymes and are thus potential therapeutic targets. It is worth to noticed that Abrahamsen et al. and Xu et al. studies showed that *C. parvum* and *C. hominis* have reduced or absent genes from metabolic pathways, organelles, and genes that are common in eukaryotes generally, or Apicomplexa specifically. This reduction in metabolic capacity forces the parasite to rely on the host for nutrient acquisition. The study conducted by Xu et al. confirmed Abrahamsen et al's data that both *C. parvum* and *C. hominis* have a diminished spliceosomal pathway, characterized by a reduced U6 spliceosomal complex and the absence of identifiable homologues to DICER and argonaute. This suggests that small interfering RNAs may play a distinct role in gene regulation in these microorganisms, as noted by Xu et al.

 doi:10.20944/preprints202307.0870.v1

In 2003, CryptoDB was created as an online database of known *Cryptosporidium* genomes, following the completion of two ground breaking genomic projects by Abrahamsen et al. and Xu et al. These projects provided the foundation for the development of CryptoDB, which currently includes fifteen genomic sequences from nine different *Cryptosporidium* genotypes. One of the genomes deposited in CryptoDB is *Cryptosporidium muris*, whose genome was sequenced using a hybrid sequencing approach. Although the results of the study are unpublished, it is known that the *C. muris* genome is similar in nucleotide composition and gene content to *C. parvum* and *C. hominis*, and encodes around 3900 protein-coding genes. Unlike *C. parvum* and *C. hominis*, *C. muris* relies on anaerobic metabolism for energy production as it lacks a cytochrome-based respiratory chain and a mitochondrial genome [12].

Over time, as techniques for sequencing have become more sophisticated and accessible, new whole-genome sequencing data from additional isolates have been published. Several recent studies have utilized draft genome sequencing to improve our understanding of *Cryptosporidium* parasites.

For instance, Ifeonu et al. [13] conducted a study on *C. hominis* isolates TU502 2012 and UKH1, *C. meleagridis* UKMEL1, and the avian parasite *C. baileyi* TAMU-09Q1. The authors resequenced the TU502 2012 genome of and improved the genome assembly by reducing the number of contigs by 119, compared with previous results. Although the genome size of 9.1 Mb remained unchanged, gene length increased by 500 bp compared with the 2004 annotation. However, the current gene set in genome TU502 2012 was found to be biased in the opposite direction. Although it contains 3745 predicted proteins, only 63% of them are present in all other annotated *Cryptosporidium* genomes such as *C. parvum* IOWA II, *C. meleagridis* UKMEL1, *C. baileyi* TAMU-09Q1 and *C. muris* RN66. The study also revealed that 110 predicted protein-coding genes were present in the three newly sequenced genomes but not observed in the *C. parvum* predicted proteome. Moreover, the researchers identified a total of 10,526 SNPs in the genome TU502 2012, while 4,394 were identified in *C. hominis* UKH1, mapped to the reference *C. hominis*. These findings suggest that there is still a need for further optimization of the assemblies of *Cryptosporidium* genomes, as well as additional analyses to support genome annotations and to identify and characterize differences among these genomes that can explain the differing phenotypes observed in these parasites, such as host range, infectivity, and pathogenesis.

Single-cell genomics is a novel approach to genome sequencing that has been particularly useful in understanding the cellular composition of complex tissues, discovering new microbial species, and advancing diagnostics. This technique is especially useful for analyzing uncultivable unicellular organisms, such as *Cryptosporidium*. Troell et al. [14] conducted an important study in which they sequenced 10 single-cell genomes of *Cryptosporidium*-infected hosts from clinical samples, using a robust procedure for preparing single cells. Fluorescence-activated cell sorting (FACS) was initially used to sort cells by using fluorescent antibodies specific to oocyst surface antigens. The oocysts were then screened using 18SrRNA. The single genomes were sequenced using Illumina MiSeq, with coverage ranging from 67% to 95% of the reference genome (*Cryptosporidium parvum* Iowa II). The combined data from the 10 genomes resulted in a high coverage of 99.8% at 20x. The *Cryptosporidium* sequences were distributed across the 10 single-cell genomes, with an average of 1.3 Gbp per sample (range 0.3-1.9 Gbp). Interestingly, the metagenome covered 97.8% of the genome, and comparison between the metagenome and single-amplified genomes revealed a small number of SNPs ranging from 13 to 51. These results suggest that high-quality genome data can be obtained from unicellular parasites with good precision and high coverage, making this approach promising for future studies. Furthermore, the exceptional advantage of a single-cell sequencing approach is the ability to examine mixed infections and test for the presence of multiple species/subtypes.

Another interesting study was conducted in Bangladesh, where unexpected high diversity within *C. hominis* was confirmed using whole-genome sequencing with long-reads (Pacific Biosciences) on 63 samples from infants over a 2-year period [15]. Of the 63 samples, 32 were derived from *C. hominis* and had >80% genome coverage at 10x. The authors used an improved WTSI reference genome (*Cryptosporidium parvum* Tyzzer ATCC PRA-67D) assembly for *C. parvum* to map the sequences. A total of 36,780 SNPs was found to vary between the Bangladesh *C. hominis* isolates,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

but only 1,582 (4.3%) occurred with a frequency >20 in the 32 whole genomes sequenced. Seven regions of particularly high nucleotide diversity were identified, including the gp60 gene, the ortholog of the *C. parvum* protein COPS, and the sequences encoding the insulinase-like peptidase. The authors also found that recombination in the pathogen was proven by the decay of linkage disequilibrium in the genome over <300 bp. The discovery of highly polymorphic regions in the parasite genome implies that these regions are under strong selective pressure to diversify and potentially underlie host-parasite interactions. The polymorphic regions contain open reading frames for membrane and secreted proteins, suggesting that they play a critical role in the parasite's adaptation to the host or immune evasion.

In 2018, Nash et al. [16] published an announcement with results from the whole genome sequencing of *C. parvum* UKP1, which was isolated from a patient suffering from cryptosporidiosis. The *C. parvum* IIc gp60 UKP1 type IIaA17G1R1 is considered a human-adapted subtype, as it has only been identified in humans and, to a lesser extent, European hedgehogs. The authors aimed to identify markers that could be used to distinguish transmission routes and potential virulence traits. The researchers purified and extracted DNA from fecal samples, and performed whole-genome sequencing using both 454 GS FLX Titanium and Illumina HiSeq 2500 sequencing technologies. The analysis resulted in 1.6 million reads, which were mapped against a reference *C. parvum* isolate, Iowa II. The eight chromosomes of the parasite were assembled into 14 contigs, with a total genome size of 8,881,956 bp, a G+C content of 30.20%, an N50 value of 1,092,230 bp, and a largest contig length of 1,333,759 bp. To provide a clearer perspective on the sequencing and results generated in *Cryptosporidium* genomics research, the following table (Table 1) summarizes the key aspects of these studies.

Table 1. Genomic sequencing data across multiple platforms for various *Cryptosporidium* species.

Species	Genomes ID	Platform	Genome size (Mbp)	# of Contigs	# of Reads	Contig N50 (bp)	Average Coverage
C. hominis	TU502 237895	Sanger Dideoxy Sequencing	8.70	1,422	-	48,000	12
	TU502 2012	Illumina MiSeq	9.10	119	1,810,060	238,509	96
	30976	Ilumina Genome Analyzer IIx 100bp paired end	9.05	53	35,360,353	470,636	511
	37999	Ilumina Genome Analyzer IIx 100bp paired end	9.05	78	16,569,87	406,678	367.4
	33537	454 GS-FLX Titanium	9.60	1,464	1,157,140	27,749	31
	30974	454 GS-FLX Titanium	8.84	443	1,048,412	78,110	43
	SWEH2	Ion Torrent	8.81	1629	1,791,829	9,465	35.2
	SWEH5	Ion Torrent	8.82	1342	2,058,197	14,514	42.4
	UdeA01	Illumina MiSeq	9.04	8	1,080,44	1,103,974	53.4
	UKH1	Illumina MiSeq	9.14	156	3,798,205	179,408	197.3
	UKH3	Illumina MiSeq	9.07	179	1,238,762	167,737	35.8
	UKH4	Illumina HiSeq	9.39	2164	11,895,367	48,766	321.5
	UKH5	Illumina HiSeq	9.06	526	12,649,912	81,885	362.2
C. parvum	Iowa II 5807	Sanger Dideoxy Sequencing	9.10	18	-	1,014,526	13
	UKP1	Illumina HiSeq	8.88	14	26,000,000	1,092,230	600
	31727	Illumina Genome Analyzer IIx 100 bp paired-end	9.08	337	13,074,496	76,396	116.0
	34902	Illumina Genome Analyzer IIx 100 bp paired-end	9.11	1,076	18,907,631	21,594	168.7

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

	35090	Illumina Genome Analyzer IIx 100 bp paired-end	9.04	3,256	14,188,762	4,248	3.256
C. baileyi	TAMU-09Q1	gDNA Illumina library fragment size (bp) 654	8.43	145	6,240,960	203,018	70.06
C. muris	5808	4.5x Sanger and 10x 454	9.25	97		520,347	10
***C. chipmunk* genotype I**	1280935	Illumina Genome Analyzer IIx 100 bp paired-end	9.05	50	9,509,783	117,886	200
C. bovis	310047	Illumina HiSeq 250 bp paired-end	9.11	59	7,080,000	444,382	196
C. ryanae	515981	Illumina HiSeq 250 bp paired-end	9.06	100	5,130,000	231,122	142.5
C. meleagridis	UKMEL1	Illumina MiSeq	8.9	57	11,431,022	322,908	110.4

2.2. Unraveling Cryptosporidium's Secrets through Comparative Genomics

The genomic sequencing of *C. parvum* conducted by Abrahamsen et al. revealed crucial insights into the intraspecies differences and host specificity of *Cryptosporidium*. In continuation and increasing this direction, Widmer et al. [17] employed a whole genome sequencing approach that utilized single nucleotide polymorphisms (SNPs) to investigate the anthroponotic *C. parvum* isolate TU114 (subtype IIcA5G3b) in detailed comparison to the reference *C. parvum* IOWA (subtype IIaA15G2R1) and *C. hominis* TU502 (subtype IaA25R3). The analyses identified 12,000 SNPs of *C. parvum* TU114 which were significantly more similar to *C. hominis* than to *C. parvum* IOWA, despite the two *C. parvum* genome sequences being tenfold more similar to each other than either was to *C. hominis* overall. This suggests that host preference might be associated with some divergent genetic loci. Highly divergent regions were found near the ends of many chromosomes, containing genes that were almost twice as large as the genome-wide average. These genes were mainly transporters, particularly ABC transporters, and proteins with signal peptides. Furthermore, the identification of three genes (cgd1_650, cgd3_3370, and cgd6_5260) exhibiting significant allelic similarity between TU114 and *C. hominis* suggests that their evolution may be attributed to the parasite's adaptation to distinct host species.

Guo et al. [18] conducted comparative studies on *Cryptosporidium chipmunk* genotype I, a newly emerged zoonotic pathogen in some industrialized nations. The authors utilized WGS to investigate the transmission of the parasite from rodents to humans. Their aim was to develop new tools to compare the genetic similarity among *Cryptosporidium chipmunk* genotype I isolates from humans, wildlife, and water. To achieve this goal, the authors used the Illumina TruSeq (v3) library protocol to sequence DNA and identified the gp60 gene and nucleotide sequence encoding another mucin protein, the ortholog of cgd1_470 in *C. parvum*, by mapping to the reference genome of *C. parvum*. These molecular markers are useful for subtyping within the species due to their high sequence polymorphism. In their study, the authors characterized chipmunk genotype I in 25 human samples received from U.S. states and Sweden, one specimen each from an eastern gray squirrel, a chipmunk, and a deer mouse, and four water samples from New York. Sequencing data revealed the presence of 9,509,783 bp in 853 contigs, resulting in a 98.98% coverage of the genome after mapping. Interestingly, the remaining contigs were mostly small and from the *C. hominis* IbA10G2 and *C. parvum* subtype, likely the result of mixed infections. Based on the WGS results, genotyping succeeded in generating 15 subtypes among the 30 isolates analyzed from humans, wildlife, and storm water. At the mucin locus, two subtypes were obtained: MCI and MCII, with differences in the number of a 30-bp minisatellite repeat. The authors' findings provide valuable information for understanding the transmission dynamics of *Cryptosporidium chipmunk* genotype I and may aid in the development of effective control strategies.

C. hominis subtypes IbA10G2 and IaA28R4 are known to be highly virulent and have a worldwide distribution. Guo et al. [19] conducted a study in which the genomes of two field specimens of these subtypes were sequenced using both 454 and Illumina technologies and

 doi:10.20944/preprints202307.0870.v1

compared. The reference genome used for the analysis was *Cryptosporidium parvum* IOWA. The results revealed that 8.59-9.05 Mb of *Cryptosporidium* sequences in 45–767 assembled contigs were generated from four DNA isolates, with a coverage of 94.36-99.47%. The genomes showed almost 97% nucleotide sequence identity with the genome of *C. parvum*. Major insertions and deletions (InDels) between *C. hominis* and *C. parvum* genomes were found in the telomeric regions. Comparative analysis of the four genomes of *C. hominis* and *C. parvum* revealed differences in the 5′ and 3′ ends of chromosome 6 and the gp60 region; such differences were largely the result of genetic recombination.

In 2017, Feng et al. [20] expanded our understanding of the genetic factors involved in host adaptation by *C. parvum*, specifically those related to proteins encoded by certain genes. The genomes of three *C. parvum* isolates from China and Egypt were sequenced using the Illumina Genome Analyzer IIx. The authors used the *C. parvum* Iowa II genome as a reference and were able to sequence over 99.3% of the genomes of the *C. parvum* isolates. Significant genomic differences were found between the sequenced genomes and the reference IOWA genome with the 5,191-5,766 single nucleotide variants and the majority occurring in subtelomeric regions including those encoding SKSR secretory proteins, the MEDLE family of secretory proteins, and insulinase-like proteases of chromosomes 1, 4, and 6. To infer the role of subtelomeric gene duplications in host adaptation by *C. parvum* subtype families, it is necessary to conduct a direct comparative genomic analysis of specimens from regions with known distinct distributions of these two subtype families among calves and lambs. Mucin proteins and other families of secretory proteins encoding invasion were found to be the most polymorphic between *C. parvum* isolates. These proteins play a role in sporozoite invasion and are highly immunogenic, making genes encoding them naturally polymorphic.

In a further comparative study of *Cryptosporidium* genomes in 2020, the authors used a comprehensive phylogenomic approach to analyze *C. hominis*, *C. parvum*, and *C. meleagridis* genomes that occur in humans [21]. They assembled de novo and compared raw whole genome sequencing data from 23 publicly available genomes of *C. hominis*, *C. parvum*, and *C. meleagridis*. The results of the analysis revealed that most genomes have a size of 9.0 Mb. Only *C. parvum* from humans (UKP14) and two *C. hominis* isolates (SWEH2 and SWEH5) have a genomic size less than 9.0 Mb. *C. parvum* genome identity within the species, in comparison to the *C. parvum* Iowa type II isolate used as a reference genome, was found to be 99.51-99.93%. The genetic similarities between *C. hominis* genomes were around 96.8%, whereas *C. meleagridis* had a lower global identity of 91.5%. *C. hominis* and *C. meleagridis* had the highest number of single nucleotide variants in coding regions, with more than 150,000 and 400,000, respectively. Single nucleotide variants detection also revealed that *C. hominis* and *C. meleagridis* had the highest number of these, with *C. meleagridis* accounting for 50% of all identified variants. Additionally, the highest number of deletions was found in *C. meleagridis*, followed by *C. hominis*. Although some indel genes have undergone partial annotation, the majority of them still encode uncharacterized proteins, posing a challenge for future research.

The three common intestinal *Cryptosporidium* species in cattle exhibit significant differences in host range, pathogenicity, and public health implications. To gain insight into the genetic determinants of biological differences among *C. bovis* and *C. ryanae*, Xu et al. [22] sequenced their genomes and mapped them to reference *C. parvum* IOWA. Interestingly, the authors proved that *C. bovis* and *C. ryanae* have gene organization and metabolic pathways similar to *C. parvum*. However, in their genomes lacks of invasion-associated mucin glycoproteins, insulinase-like proteases, MEDLE secretory proteins, and other SPDs which suggest the narrower host range of these two species. Furthermore, 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* which should be further investigated. Comparative genomic analysis of these three intestinal species reveals reductions in secreted pathogenesis determinants in bovine-specific and non-pathogenic *Cryptosporidium* species.

2.3. Overcoming Challenges in Isolating Cryptosporidium DNA from Clinical Samples

Cryptosporidium is a challenging parasite to study due to the lack of effective in vitro cultivation methods for completing its life cycle. As a result, extracting high purity DNA from oocysts obtained

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

from fecal specimens remains difficult. In this review, we discuss recent advances in methods for isolating and enriching *Cryptosporidium* genomic DNA from clinical specimens for whole-genome sequencing.

Guo et al. [23] developed a method that combined sucrose and cesium chloride density gradient separation with immunomagnetic separation (IMS) for purifying oocysts from fecal samples of humans and animals with six *Cryptosporidium* species or genotypes. WGS was used to verify the genome coverage and contamination. The results showed that the generated sequence data covered 94.5% to 99.7% of *Cryptosporidium* genomes, with minor contamination from bacterial, fungal, and host DNA. Twenty WGA products with low CT values were submitted to whole-genome sequencing, generating sequence data covering 94.5% to 99.7% of *Cryptosporidium* genomes with mostly minor contamination from bacterial, fungal, and host DNA. These results suggest that this strategy can effectively isolate and enrich *Cryptosporidium* DNA from fecal specimens for whole-genome sequencing.

Anderson et al. [24] performed a simplified version of the above method using IMS combined with PCR to sequence a positive *C. hominis* sample (subtype IbA9G3). The results showed a 91.25% target DNA coverage, with an improvement in parameters compared with the previously published *C. hominis* reference strain TU502.

Similarly, Hadfield et al. [25] developed a preparatory method based on a combination of salt flotation, IMS, and surface sterilization of oocysts prior to DNA isolation directly from human fecal samples. The IMS method was the most efficient and was applied to the sequencing of 17 fecal samples. The results showed eight new whole genome sequences, including two *C. hominis* and six *C. parvum* subtypes. The authors concluded that their method could be useful for clinical sample analysis. To provide whole perspective of the genomic studies, the following table (Table 2) summarizes the key aspects of the research during the 20 years of *Cryptosporidium* spp. genomics.

Table 2. Timeline and progression of genomic research in *Cryptosporidium* species.

Year	The greatest milestone	Genomic approach	Outcome of study	Reference
1999	Initial genomic exploration into *C. parvum* Iowa strain	Random sequence analysis	✓ Sequence survey comprising only 2.5 % of the *C. parvum* genome ✓ Disadvantage of method is limited availability of gene sequences in the public databases at the time	[8]
2000	First cDNA sequence survey of *C. parvum* Iowa oocysts/sporozoites	Random sequence analysis with GSS approach to gene discovery	✓ First broad-based molecular views into the basic biology and cellular metabolism of this experimentally intractable apicomplexan parasite	[9]
2004	Complete genome sequencing of *C. parvum* Iowa type II strain	Whole-genome with shotgun Sanger sequencing	✓ Whole genome was sequenced from a plasmid insert library ✓ "HAPPY" map was used to create scaffold and ordering the sequence contigs	[10]
2004	Complete genome sequencing *C. hominis* TU502	Whole-genome with shotgun Sanger sequencing	✓ Whole genome was sequenced from a plasmid insert library ✓ "HAPPY" map was not available, the authors constructed large (~7–8-fold coverage), bacterial artificial chromosome (BAC) libraries for this species and used the *C. parvum* HAPPY map to guide assembly	[11]
2012	Comparative genome analysis of two *C. parvum* isolates (TU114 and *C. parvum* IOWA)	Whole-genome sequencing	✓ Small number of highly diverged genes ✓ Transporter genes' over-representation indicates species-specific infection ability	[17]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

2015	Sequencing of genomes *C. chipmunk* genotype I	Whole genome sequencing	✓ subtyping tool based on two markers for genetic characterization of *C. chipmunk* genotype I has been developed ✓ *C. chipmunk* genotype I isolates from humans and wildlife are genetically similar	[18]
2015	Comparative genome analysis of *C. hominis* and *C. parvum*	Whole genome sequencing	✓ Occurrence of genetic recombination in virulent ✓ *C. hominis* subtypes and telomeric gene duplications in *C. parvum* ✓ Sequence similarity and recombination in gp60 region indicate potential role in emergence of highly transmissible *C. hominis* subtypes	[19]
2016	Sequencing of genomes: *C. meleagridis* UKMEL1, *C. baileyi* TAMU-09Q1 and *C. hominis* TU502_2012 and UKH1	Draft genome sequencing	✓ The genome assembly of *C. hominis* is significantly more complete and less fragmented than previous version ✓ The first versions of genome sequence assemblies and annotations for each isolate	[13]
2016	Genome sequencing of *Cryptosporidium* spp. in clinical samples	Single cell sequencing	✓ Workflow for whole genome sequencing of single cells of parasite ✓ Combining sequence data from all single cell genomes, almost the entire reference genome (99.7 %) was accounted for and most of this (98.8 %) was described at > 20× coverage	[14]
2017	Sequencing of the genomes of two specimens of *C. parvum* form China and Egypt	Whole genome sequencing	✓ Differences in subtelomeric gene families, such as SKSR, MEDLE proteins, and insulinase-like proteases, found between sequenced and reference genomes ✓ Most polymorphic genes between genomes mainly encode invasion-related mucin proteins and other secretory protein families	[20]
2018	Analysis of genetic diversity of *C. hominis* infections in slum-dwelling infants in Bangladesh	Long-read resequencing	✓ High rates of sexual recombination and regions of the genome that were highly polymorphic, suggesting areas under selection	[15]
2018	Analysis of a zoonotic isolate of *C. parvum* UKP1 isolated from a person with cryptosporidiosis	Draft genome sequencing	✓ Sequences needed to identify markers important in distinguishing routes of transmission and potential virulence traits for better epidemiological analysis and risk assessment	[16]
2020	Comparative analysis of *Cryptosporidium* species that infect humans	Whole-genome sequencing	✓ Synonymous single nucleotide variants were the most common in *C. hominis* and *C. meleagridis*, while in *C. parvum*, they accounted for around 50% of the SNV observed	[21]
2020	Sequencing of the genomes of *C. bovis* and *C. ryanae*	Whole-genome sequencing	✓ Genome of *C. bovis* has a gene content and organization more similar to *C. ryanae* than to other *Cryptosporidium* species sequenced to date	[22]

3. Proteome

The first publications about the *Cryptosporidium* genome presented the possibility of investigating the proteins expressed by this parasite; its proteome can provide comple-mentary information about the biology of this complex organism. However, despite the development of proteomic techniques, one of the significant challenges in studying the proteome of *Cryptosporidium* is the difficulty in cultivating the parasite, resulting in inad-equate material for analysis. Accordingly, only a few proteomic analyses have been con-ducted on *Cryptosporidium*, emphasizing the need for more research in this area.

Magnuson et al. [26] conducted the first such study before the genomic era of *Cryptosporidium* investigations; in their study they implemented Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) for proteomic analysis of the parasite. The important effect of these investigations was that they achieved obtainment of reproducible patterns of spectral markers with increasing sensitivity after lysing the oocysts with a freeze-thaw procedure. However, because of the lack of genomic data, their analysis focused mainly on differentiating between *C. parvum* and *C. muris,* based on the spectral peaks patterns specific to the genus. Furthermore, the authors proved that the disinfection of the oocysts resulted in the reduction in and/or elimination of the patterns of spectral markers. Using the same MALDI-TOF MS approach, Glassmeyer et al. [27] conducted a study seven years later and received spectral peaks specific to the oocyst and as well as sporozoites. In this study, the authors focused mainly on sample preparation, including mixing and standing for 45 minutes before spotting. As a result, the mass spectra of intact oocysts showed similar peaks to those of sporozoites, indicating that the laser ablates both the oocyst wall and its internal constituents.

The availability of the *Cryptosporidium parvum* genome sequence has allowed for large-scale global analyses of its expressed proteome in oocysts and sporozoites. Additionally, the widespread adoption of liquid chromatography coupled with mass spectrometry technology (LC-MS/MS) has facilitated more in-depth protein analysis, opening new avenues for future studies on the biology of *Cryptosporidium*.

In the first such study, Snelling et al. [28] aimed at the known proteome of *C. parvum* in the non-excysted (transmissive) and excysted (infective) forms. The authors used LC-MS/MS with a stable isotope N-terminal labeling strategy to identify and quantify proteins in soluble fractions of both non-excysted and excysted sporozoites. The authors confirmed the expression of many previously hypothesized proteins, as well as the presence of several secreted or surface proteins in sporozoites that are unique to either the Apicomplexa or the *Cryptosporidium* genus. The analysis identified a total of 303 *C. parvum* proteins, of which 26 proteins were found to be significantly upregulated during excystation. These findings suggest that future vaccine development strategies could potentially focus on selecting antigens that block the recognition and attachment of the parasite to host cells, thereby preventing infection. Of particular interest, analysis revealed the presence of seven proteins that are unique to the genus *Cryptosporidium* and/or the phylum Apicomplexa. Notably, five of these proteins contain a signal peptide at their N-terminus, indicating that they are secreted proteins. These proteins are likely to play a specialized role in the invasion machinery of the parasite, and further investigation of their functions could provide insights into the pathogenicity of the parasite. In addition, the parasite is likely protected against stress and apoptosis by the abundant presence of heat shock proteins (HSP) and ribosomal proteins, which act as protective mechanisms. Three Apicomplexa-specific proteins and five *Cryptosporidium*-specific proteins were found to be augmented in excysted invasive sporozoites. The authors also discovered eight promising targets for developing vaccines or chemotherapies that could prevent parasite entry into host cells e.g. glycosylinositol phospholipids and glycosylphosphatidylinositol protein anchors which are abundant in the surface membranes and are increasingly recognized as important modulators of the immune function during infection. Moreover, this study provides further evidence of the expression of antigenic proteins, namely Cpa135 (N64) and GP900 (N88). These proteins not only induce the production of specific antibodies but also elicit a cellular Th1 response. Furthermore, the outcome of the studies was the identification of proteins that represent around 8% of the proteome, predicted based on the whole genome sequencing by Anderson et al. and Xu et al.

Sanderson's paper [29] presents in-depth analysis of the expressed protein repertoire of *Cryptosporidium parvum*. The authors used three independent proteome platforms (2-DE LCMS/MS; 1-DE LC-MS/MS; and multi-dimensional protein identification technology (MudPIT analysis)) to identify more than 4800 individual proteins representing 1237 nonredundant proteins. The peptide data were mapped to the respective locations on the

C. parvum genome, and a publicly accessible interface was developed for data-mining and visualization in CryptoDB. The data provide a valuable resource for improved annotation of the

 doi:10.20944/preprints202307.0870.v1

genome, verification of predicted hypothetical proteins, and identification of proteins not predicted by current gene models. The expressed proteome indicates the expression of proteins important for invasion and the intracellular establishment of the parasite, including surface proteins, constituents of the remnant mitochondrion, and apical organelles. Proteomic analysis of the *C. parvum* genome revealed that between 25% and 40% of the potential gene products were classified as "hypothetical proteins" due to a lack of sufficient similarity to known proteins. However, the study provides proteomic evidence for 482 of these hypothetical proteins in *C. parvum*, which represents 39% of the total proteome detected. It is estimated that 69%-74% of the coding region of the *C. parvum* genome contains genes based on imperfect algorithmic prediction programs. Furthermore, the study suggests that an additional 72 regions of the genome scaffold may contain coding exons that were not previously recognized by current gene prediction models. In this study, the dataset of transporter-like proteins associated with the oocyst/sporozoite life- cycle stages has been expanded to include 64 proteins. The proportion of the total proteome that these proteins represent is estimated to be 9%, which is consistent with the prediction based on genome analysis. It is worth noting that, despite using such advanced techniques, identified proteins corresponded to 30% of the predicted proteome which highlight the current limitations in understanding *Cryptosporidium* biology.

Another example of a proteomic study aimed at exploring the stage-specific proteome of Cryptosporidium parvum is [30]. The authors paid particular attention to the presence of C. parvum hydrophobic and membrane proteins. Therefore, the study aimed to reduce sample complexity by protein fractionation, enabling the detection of proteins present in lower abundance in a complex protein mixture. In this study, therefore, they used SDS–PAGE to fractionate the sporozoite protein of C. parvum, followed by an LC–MS/MS approach. A total of 135 protein hits were recorded from 20 gel slices, with many hits occurring in more than one band. Excluding non-Cryptosporidium entries and proteins with multiple hits, 33 separate C. parvum entries were identified. These included structural, metabolic and hypothetical proteins covering a wide range of pHs and molecular weights. Several metabolic enzymes, including the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase and protein disulphide isomerase, were identified in this study. These findings support the hypothesis that glycolysis is the primary energy source in C. parvum and that the parasite relies mainly on the anaerobic oxidation of glucose for energy production. The presence of these enzymes in the proteome is notable and provides further evidence of the reliance of the parasite on glycolysis.

With the development of mass-spectrometry as a protein discovery tool, studies on the *Cryptosporidium* proteome have become more specific and focused on characterizing specific fractions of the parasite. The first such study, using mass spectrometry, aimed at identifying the components of the oocyst wall and affinity purification to isolate glycoproteins from the sporozoites, in order to investigate how sporozoites are tethered to the oocyst wall[31]. The investigation revealed that COWPs comprise 75% of oocyst wall proteins, with COWP1 being the most abundant. COWP8 and COWP6 were also present. Affinity-purified glycoproteins included known mucin-like glycoproteins, a GalNAc-binding lectin, and novel glycoproteins, suggesting their involvement in tethering sporozoites to the oocyst wall. These results suggest that mucin-like glycoproteins may contribute to the fibrils and/or globules that tether sporozoites to the inner surface of oocyst walls.

Another such examination focused on organelles called rhoptries and their contents which are crucial in establishing productive infection during invasion [32]. In order to identify rhoptry proteins and explore the parasite's invasion pathway and pathogenic mechanisms subcellular fractionation and mass spectrometry analysis were used.

The authors detected 22 potential novel rhoptry proteins by fraction analysis, using LC–MS/MS and online software. It is worth noting that these novel candidates could serve as targets for future research on the *C. parvum* invasion pathway and the function of rhoptry proteins.

In the recent years, studies have mainly focused on identifying and characterizing proteins involved in the host–parasite interaction, elucidating the parasite's virulence mechanisms, and discovering potential drug targets for the development of effective therapeutic agents. Li et al. [33] investigated the excystation process in *Cryptosporidium andersoni* by collecting and purifying oocysts

 doi:10.20944/preprints202307.0870.v1

from naturally infected adult cows. Using both immunological and molecular methods, this study identified several proteins as putative virulence factors. Their study yielded interesting data: they identified 1586 proteins in
C. andersoni, and found 17 differentially expressed proteins (DEPs) after excystation, including 10 upregulated and seven downregulated proteins. This is a larger number of proteins than previously reported for *C. parvum*, suggesting that *C. andersoni* may have a more complex proteome or different proteins involved in excystation. This study identified several proteins as putative virulence factors using both immunological and molecular methods. For instance, serine protease and aminopeptidase are associated with excystation, while P23 and P30 are involved in adhesion. TRAP and thrombospondin-related anonymous protein are implicated in parasite gliding and cell penetration. Cp2, Cap135, and secreted phospholipase are associated with invasion, while HSP70 and HSP90 are involved in stress protection. These proteins play key roles in various stages of the *C. parvum* life cycle and are potential targets for future drug and vaccine development. These findings provide new insights into the protein composition of *C. andersoni* oocysts and potential excystation factors, which could be useful in identifying genes for diagnosis, vaccine development, and immunotherapy for *Cryptosporidium*.

In the field of *Cryptosporidium* research, limited peptide and protein data are available for various isolates of *C. parvum* and other *Cryptosporidium* species, such as *C. hominis,* due to the lack of easily accessible animal models. As a result, proteomic studies have made their way into the analysis of clinical samples.

Gathercole et al. [34] developed a refined protocol for purifying oocysts from clinical fecal samples of human patients using salt flotation and potassium bromide density centrifugation. The purified oocysts were then analyzed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify specific spectral markers unique to *Cryptosporidium hominis* and *C. parvum*. This study successfully demonstrated the potential utility of the purification method in clinical samples, providing high-quality reference spectra for further peptide and protein discovery and biomarker identification. The results also showed the applicability of MALDI-TOF MS in analyzing *Cryptosporidium* spp. from natural infections, broadening the range of microorganisms that can be studied using mass spectrometry techniques.

In 2021, Kacar et al. [35] conducted a study to investigate the effectiveness of proteome method analysis on the sera of Holstein calves naturally infected with *Cryptosporidium* spp., after being treated with high-quality colostrum and paromomycin. The study analyzed sera samples collected at the 0th and 3rd days post-treatment across three different groups of animals: the first group received only paromomycin (PC); the second group received paromomycin and colostrum (PCOL); and the third group received paromomycin, colostrum, and sodium bicarbonate (PBCOL). The study's results indicated that colostrum treatment had the most positive effect, as significant changes in several proteins were observed. These proteins played important roles in binding/transporter, catalytic activity, regulation of molecular functions, and regulation of structural–molecular activity. Additionally, the study revealed down-regulation of many inflammatory mechanisms and processes in the colostrum-treated groups, which provides insights into the mechanism of action of colostrum in treating the infection.

Unfortunately, there is currently no fully effective drug or vaccine available to treat cryptosporidiosis, despite the reemergence of *Cryptosporidium* infection. A large-scale label-free proteomics approach was employed to understand the detailed interaction between the host and *Cryptosporidium parvum* [36]. Among the 4406 proteins identified, 121 proteins were differentially abundant (>1.5-fold cutoff, $P < 0.05$) in *C. parvum*-infected HCT-8 cells compared with uninfected cells. Specifically, 67 proteins were upregulated, and 54 proteins were downregulated 36 hours post-infection. Analysis of the differentially abundant proteins revealed that the host cells mounted an interferon-centered immune response against *C. parvum* infection and that the parasite extensively inhibited metabolism-related enzymes in the host cells. Several proteins were further verified using quantitative real-time reverse transcription polymerase chain reaction and Western blotting. This systematic analysis of the proteomics of *C. parvum*-infected HCT-8 cells identified a range of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

functional proteins that participate in host anti-parasite immunity or act as potential targets during infection, providing new insights into the molecular mechanism of *C. parvum* infection.

Karpe et al. [37] investigated the biochemical interactions underlying *Cryptosporidium parvum* infection in C57BL/6J mice. To do so, they collected various specimens from the mice 10 days post-infection, including fecal samples, blood, liver tissues, and luminal contents. They then analyzed the proteomes and metabolomes of these specimens using high-resolution liquid chromatography and low-resolution gas chromatography coupled with mass spectrometry. Their aim was to shed light on the biochemical mechanisms of cryptosporidiosis, which could have implications for clinical diagnosis. Univariate and multivariate statistical analysis revealed altered host and microbial energy pathways during infection, including depleted glycolysis/citrate cycle metabolites, increased short-chain fatty acids and D-amino acids, an abundance of bacteria associated with a stressed gut environment, and upregulated host proteins involved in energy pathways and *Lactobacillus* glyceraldehyde-3-phosphate dehydrogenase. Liver oxalate also increased during infection. They observed that microbiome-parasite relationships were more influential than the host-parasite association in mediating major biochemical changes in the mouse gut during cryptosporidiosis. Defining this parasite-microbiome interaction is the first step towards building a comprehensive cryptosporidiosis model to discover biomarkers and develop rapid and accurate diagnostics.

The advent of genomic and proteomic techniques has enabled a deeper analysis of the genomes and proteomes of the major *Cryptosporidium* species, leading to improved insights into their biology and pathogenicity. It is worth noting that these surveys address a wide range of questions, spanning basic biology, drug target identification, taxonomy, and population biology. Tables 3 and 4 offers a detailed overview of *Cryptosporidium* proteomics studies, encapsulating critical information about the specimens, applied techniques, search engines utilized, and reference genomes, providing a comprehensive context to the review.

However, further work is needed to fully explore the genomes and proteomes of *Cryptosporidium*, and challenges remain in studying this pathogen due to its intracellular nature, genetic diversity, and difficulties in sample collection and preparation. Nonetheless, ongoing research efforts and advances in technology are likely to lead to new strategies for controlling and preventing *Cryptosporidium* infections.

Table 3. Chronological Advancements in Proteomic Analysis of *Cryptosporidium* Species.

Year	The greatest milestone	Outcome of Study	Reference
2000	Initial proteomic study of whole and freeze-thawed *C. parvum* oocysts and freeze-thawed *C. muris*	✓ Identify spectral peaks that can distinguish *Cryptosporidium* at the genus level, as well as specific peaks that enable differentiation between *C. parvum* and *C. muris* oocysts ✓ The pioneering utilization of MALDI-TOF peptide mass fingerprinting (PMF) for *Cryptosporidium* oocysts analysis	[26]
2007	Proteomic analysis of *C. parvum* oocysts	✓ Improvements of the sample preparation before analysis with using MALDI-TOF peptide mass fingerprinting (PMF) ✓ Spectral peaks that were specific to the oocyst and sporozoites	[27]
2007	Large scale global proteomic analysis of nonexcyted and excyted *C. parvum* sporozoites	✓ Identification of around 200 proteins, representing about 6 % of the predicted proteome ✓ 26 proteins were found to have significantly higher expression levels post-excystation, relative to unexcysted oocyst	[28]
2008	In depth analysis of the expressed protein repertoire of *C. parvum*	✓ A total of 642. 282 and 1154 non-redundant proteins were identified from the 1-DE, 2-DE, and MudPIT analyses, respectively ✓ 1237 non-redundant proteins were identified from excysted oocysts and sporozoites	[29]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2023 doi:10.20944/preprints202307.0870.v1

2010	Proteome analysis for identifying the key components of the *C. parvum* oocyst wall	✓ COWPs constitute about 75 % of the proteins identified in the oocyst walls ✓ COWP1 is the dominant oocyst wall protein	[30]
2013	Proteome analysis of *C. parvum* sporozoites	✓ In total, 135 hits were recorded from analysis of all 20 gel slices ✓ 41% of which were unique hits for *Cryptosporidium*	[31]
2015	Proteomic analysis of rhoptry-enriched fractions from *C. parvum*	✓ 22 potential novel rhoptry proteins were detected ✓ Novel candidate proteins may be considered targets for researching the invasion pathway of *C. parvum* and the pathogenic mechanisms of rhoptry proteins	[32]
2021	Proteomic analysis of *C. andersoni* oocysts before and after excystation	✓ Total of 1586 proteins were identified ✓ 17 of 1586 were differentially expressed proteins (DEPs) upon excystation and had multiple biological functions associated with control of gene expression at the level of transcription and biosynthetic and metabolic processes	[33]
2021	Proteomic analysis of *Cryptosporidium* spp. from clinical samples	✓ Utility of the purification method for oocysts from clinical stool samples ✓ Implementation of MALDI-TOF MS for clinical sample analysis	[34]
2021	Assessing the effectiveness of cow colostrum for treating cryptosporidiosis in calves and its impact on serum proteomes	✓ The use of colostrum in the treatment of cryptosporidiosis affects the serum proteomes of calves ✓ Serum amyloid A was the most altered proteome in the sera of calves with colostral treatment	[35]
2021	Characterize the changes to the proteome induced by *C. parvum* infection	✓ Among 4406 proteins identified, 121 proteins were identified as differentially abundant in *C. parvum* infected HCT-8 cells compared with uninfected cells ✓ Wide range of functional proteins that participate in host anti-parasite immunity or act as potential targets during infection, provides new insights into the molecular mechanism of *C. parvum* infection	[36]
2021	Investigation the underlying biochemical interaction in C57BL/6J mice infected with *C. parvum*	✓ Glycolysis and glutaminolysis were significantly impacted in the jejunum and ileum during cryptosporidiosis ✓ Gut microbiome response to cryptosporidiosis was detected via increased levels of D-amino acids and SCFAs ✓ Ability of multi-omics to contribute a robust understanding of gut infections and demonstrates the previously unreported infection interactomics as the parasite passes through the gut	[37]

Table 4. Comprehensive Overview of Techniques and Resources in *Cryptosporidium* Proteomics Research.

Specimens	Strain	Technique	Search Engines	Reference genome	Protein coding genes in reference genome
C. parvum and *C.muris* oocysts	Iowa and RN66	MALDI-TOF peptide mass fingerprinting (PMF)	-	-	-
C. parvum sporozoites	Iowa	MALDI-TOF MS	-	-	-
C. parvum sporozoites non-excysted and excysted	ISSC162	combination of LC-MS/MS and iTRAQ isobaric labelling	ProQUANT software 1.1 (Applied Biosystems)	*C. parvum* Iowa type II	3941
C. parvum excysted oocyst/sporozoite	Iowa	three independent platforms: 1-DE LC-MS/MS, 2-DE LC-MS/MS and MudPIT	MASCOT search tool, SEQUEST algorithm version 27	*C. parvum* Iowa type II	3941

 doi:10.20944/preprints202307.0870.v1

C. parvum sporozoites	Iowa	SDS-PAGE and LC-MS/MS	MASCOT search tool	*C. parvum Iowa type II* *C. hominis TU502*	3941 3886
C. parvum oocysts	Iowa	LC-MS/MS	SEQUEST search tool, NR database at the NCBI	*C. parvum Iowa type II* *C. hominis TU502*	3941 3886
C. parvum Isolate	Iowa	SDS-PAGE and LC-MS/MS	MASCOT in the NCBI, CryptoDB v5.0, EupathDB v2.16 databases	*C. parvum* Iowa type II	3941
C. andersoni oocysts	-[1]	SDS-PAGE and LC-MS/MS	MaxQuant search engine (v.1.5.2.8), UniProt database	*C. andersoni* 30847	3876
Cryptosporidium spp.	-	MALDI-TOF MS	flexControl software on a microflex LT/SH MALDI-TOF	*C. parvum* Iowa type II *C. hominis* TU502	3941 3886
Cryptosporidium spp.	-	label-free proteomic quantification techniques and LC-MS/MS	Maxquant search engine (v.1.5.2.8, Max Planck Institute of Biochemistry, Munich, German)	*C. parvum* Iowa type II *C. hominis* TU502	3941 3886
C. parvum	-	multi-omics approach: GC-MS and LC-HR-MS	The Protein Discoverer 2.2 (Thermo Scientific) and Sequest HT search engines	*C. parvum* Iowa type II *C. hominis* TU502	3941 3886

Unknown strains from clinical samples.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, JD; formal analysis, JS; resources, JD; data curation, JD; writing—original draft preparation, JD; writing—review and editing JD; supervision, JS, TC. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

References

1. Kváč, M.; McEvoy, J.; Stenger, B.; Clark, M. Cryptosporidiosis in other vertebrates. *Cryptosporidium: parasite and disease* **2014**, 237-323.
2. Pumipuntu, N.; Piratae, S. Cryptosporidiosis: A zoonotic disease concern. *Veterinary world* **2018**, *11*, 681.
3. Kotková, M.; Nemejc, K.; Hanzal, B.S.V.; McEvoy, J.; Kvác, M. Cryptosporidium ubiquitum, C. muris and Cryptosporidium deer genotype in wild cervids and caprines in the Czech Republic. *Folia Parasitologica* **2016**, *63*, 1.
4. Jirku, M.; Valigurová, A.; Koudela, B.; Krízek, J.; Modrý, D.; Slapeta, J. New species of Cryptosporidium Tyzzer, 1907 (Apicomplexa) from amphibian host: morphology, biology and phylogeny. *Folia parasitologica* **2008**, *55*, 81.
5. Wang, T.; Wei, Z.; Zhang, Y.; Zhang, Q.; Zhang, L.; Yu, F.; Qi, M.; Zhao, W. Cryptosporidium in asymptomatic children in Southern Xinjiang, China and the potential of zoonotic transmission. **2021**.
6. Checkley, W.; White, A.C.; Jaganath, D.; Arrowood, M.J.; Chalmers, R.M.; Chen, X.-M.; Fayer, R.; Griffiths, J.K.; Guerrant, R.L.; Hedstrom, L. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *The Lancet Infectious Diseases* **2015**, *15*, 85-94.
7. Mead, J.R.; Arrowood, M.J. Treatment of cryptosporidiosis. In *Cryptosporidium: parasite and disease*, Springer: 2013; pp. 455-486.
8. Liu, C.; Vigdorovich, V.; Kapur, V.; Abrahamsen, M.S. A random survey of the Cryptosporidium parvum genome. *Infection and immunity* **1999**, *67*, 3960-3969.
9. Strong, W.B.; Gut, J.; Nelson, R.G. Cloning and sequence analysis of a highly polymorphic Cryptosporidium parvum gene encoding a 60-kilodalton glycoprotein and characterization of its 15-and 45-kilodalton zoite surface antigen products. *Infection and immunity* **2000**, *68*, 4117-4134.

10. Abrahamsen, M.S.; Templeton, T.J.; Enomoto, S.; Abrahante, J.E.; Zhu, G.; Lancto, C.A.; Deng, M.; Liu, C.; Widmer, G.; Tzipori, S. Complete genome sequence of the apicomplexan, Cryptosporidium parvum. *Science* **2004**, *304*, 441-445.
11. Xu, P.; Widmer, G.; Wang, Y.; Ozaki, L.S.; Alves, J.M.; Serrano, M.G.; Puiu, D.; Manque, P.; Akiyoshi, D.; Mackey, A.J. The genome of Cryptosporidium hominis. *Nature* **2004**, *431*, 1107-1112.
12. Widmer, G.; Sullivan, S. Genomics and population biology of Cryptosporidium species. *Parasite immunology* **2012**, *34*, 61-71.
13. Ifeonu, O.O.; Chibucos, M.C.; Orvis, J.; Su, Q.; Elwin, K.; Guo, F.; Zhang, H.; Xiao, L.; Sun, M.; Chalmers, R.M. 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. *FEMS Pathogens and Disease* **2016**, *74*, ftw080.
14. Troell, K.; Hallström, B.; Divne, A.-M.; Alsmark, C.; Arrighi, R.; Huss, M.; Beser, J.; Bertilsson, S. Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes. *BMC genomics* **2016**, *17*, 1-12.
15. Gilchrist, C.A.; Cotton, J.A.; Burkey, C.; Arju, T.; Gilmartin, A.; Lin, Y.; Ahmed, E.; Steiner, K.; Alam, M.; Ahmed, S. Genetic diversity of Cryptosporidium hominis in a Bangladeshi community as revealed by whole-genome sequencing. *The Journal of infectious diseases* **2018**, *218*, 259-264.
16. Nash, J.H.; Robertson, J.; Elwin, K.; Chalmers, R.A.; Kropinski, A.M.; Guy, R.A. Draft genome assembly of a potentially zoonotic Cryptosporidium parvum isolate, UKP1. *Microbiology Resource Announcements* **2018**, *7*, e01291-01218.
17. Widmer, G.; Lee, Y.; Hunt, P.; Martinelli, A.; Tolkoff, M.; Bodi, K. Comparative genome analysis of two Cryptosporidium parvum isolates with different host range. *Infection, Genetics and Evolution* **2012**, *12*, 1213-1221.
18. Guo, Y.; Cebelinski, E.; Matusevich, C.; Alderisio, K.A.; Lebbad, M.; McEvoy, J.; Roellig, D.M.; Yang, C.; Feng, Y.; Xiao, L. Subtyping novel zoonotic pathogen Cryptosporidium chipmunk genotype I. *Journal of clinical microbiology* **2015**, *53*, 1648-1654.
19. Guo, Y.; Tang, K.; Rowe, L.A.; Li, N.; Roellig, D.M.; Knipe, K.; Frace, M.; Yang, C.; Feng, Y.; Xiao, L. 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.
20. Feng, Y.; Li, N.; Roellig, D.M.; Kelley, A.; Liu, G.; Amer, S.; Tang, K.; Zhang, L.; Xiao, L. Comparative genomic analysis of the IId subtype family of Cryptosporidium parvum. *International Journal for Parasitology* **2017**, *47*, 281-290.
21. Arias-Agudelo, L.M.; Garcia-Montoya, G.; Cabarcas, F.; Galvan-Diaz, A.L.; Alzate, J.F. Comparative genomic analysis of the principal Cryptosporidium species that infect humans. *PeerJ* **2020**, *8*, e10478.
22. Xu, Z.; Li, N.; Guo, Y.; Feng, Y.; Xiao, L. Comparative genomic analysis of three intestinal species reveals reductions in secreted pathogenesis determinants in bovine-specific and non-pathogenic Cryptosporidium species. *Microbial Genomics* **2020**, *6*.
23. Guo, Y.; Li, N.; Lysén, C.; Frace, M.; Tang, K.; Sammons, S.; Roellig, D.M.; Feng, Y.; Xiao, L. Isolation and enrichment of Cryptosporidium DNA and verification of DNA purity for whole-genome sequencing. *Journal of clinical microbiology* **2015**, *53*, 641-647.
24. Andersson, S.; Sikora, P.; Karlberg, M.L.; Winiecka-Krusnell, J.; Alm, E.; Beser, J.; Arrighi, R.B. It's a dirty job—a robust method for the purification and de novo genome assembly of Cryptosporidium from clinical material. *Journal of microbiological methods* **2015**, *113*, 10-12.
25. Hadfield, S.J.; Pachebat, J.A.; Swain, M.T.; Robinson, G.; Cameron, S.J.; Alexander, J.; Hegarty, M.J.; Elwin, K.; Chalmers, R.M. Generation of whole genome sequences of new Cryptosporidium hominis and Cryptosporidium parvum isolates directly from stool samples. *BMC genomics* **2015**, *16*, 1-12.
26. Magnuson, M.L.; Owens, J.H.; Kelty, C.A. Characterization of Cryptosporidium parvum by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Applied and environmental microbiology* **2000**, *66*, 4720-4724.
27. Glassmeyer, S.T.; Ware, M.W.; SCHAEFER III, F.W.; Shoemaker, J.A.; Kryak, D.D. An improved method for the analysis of Cryptosporidium parvum oocysts by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Journal of Eukaryotic Microbiology* **2007**, *54*, 479-481.
28. Snelling, W.J.; Lin, Q.; Moore, J.E.; Millar, B.C.; Tosini, F.; Pozio, E.; Dooley, J.S.; Lowery, C.J. Proteomics Analysis and Protein Expression during Sporozoite Excystation of Cryptosporidium parvum (Coccidia, Apicomplexa)* S. *Molecular & cellular proteomics* **2007**, *6*, 346-355.
29. Sanderson, S.J.; Xia, D.; Prieto, H.; Yates, J.; Heiges, M.; Kissinger, J.C.; Bromley, E.; Lal, K.; Sinden, R.E.; Tomley, F. Determining the protein repertoire of Cryptosporidium parvum sporozoites. *Proteomics* **2008**, *8*, 1398-1414.
30. Siddiki, A.Z. Sporozoite proteome analysis of Cryptosporidium parvum by one-dimensional SDS-PAGE and liquid chromatography tandem mass spectrometry. *Journal of Veterinary Science* **2013**, *14*, 107-114.

31. Chatterjee, A.; Banerjee, S.; Steffen, M.; O'Connor, R.M.; Ward, H.D.; Robbins, P.W.; Samuelson, J. Evidence for mucin-like glycoproteins that tether sporozoites of Cryptosporidium parvum to the inner surface of the oocyst wall. *Eukaryotic cell* **2010**, *9*, 84-96.
32. Huang, Y.; Rongsheng, M.; Wei, C.; Peng, Z.; Kai, S.; Xiaojiao, Y.; Xiaojuan, W.; Xiangpei, W.; Zhaoguo, C. Isolation and Proteomic Analysis of Rhoptry-Enriched Fractions from Cryptosporidium parvum. *Iranian Journal of Public Health* **2015**, *44*, 1187.
33. Li, D.; Cui, Z.; Wang, L.; Zhang, K.; Cao, L.; Zheng, S.; Zhang, L. TMT-based Proteomic Analysis and Protein Expression During Excystation of Cryptosporidium Anderson. **2021**.
34. Gathercole, R.; Tranfield, E.; Xia, D.; Perez-Cordon, G.; Robinson, G.; Timofte, D.; Zendri, F.; Chalmers, R.M. Analysis of Cryptosporidium spp. from clinical samples by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Journal of applied microbiology* **2021**, *131*, 1840-1847.
35. Kaçar, Y.; Baykal, A.T.; Aydin, L.; Batmaz, H. Evaluation of the efficacy of cow colostrum in the treatment and its effect on serum proteomes in calves with cryptosporidiosis. *Veterinary Immunology and Immunopathology* **2022**, *248*, 110429.
36. Li, T.; Liu, H.; Jiang, N.; Wang, Y.; Wang, Y.; Zhang, J.; Shen, Y.; Cao, J. Comparative proteomics reveals Cryptosporidium parvum manipulation of the host cell molecular expression and immune response. *PLoS Neglected Tropical Diseases* **2021**, *15*, e0009949.
37. Karpe, A.V.; Hutton, M.L.; Mileto, S.J.; James, M.L.; Evans, C.; Shah, R.M.; Ghodke, A.B.; Hillyer, K.E.; Metcalfe, S.S.; Liu, J.-W. Cryptosporidiosis modulates the gut microbiome and metabolism in a murine infection model. *Metabolites* **2021**, *11*, 380.