# Accepted Manuscript


RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites

Christoph Lippuner, Chandra Ramakrishnan, Walter U. Basso, Marc W. Schmid, Michal Okoniewski, Nicholas C. Smith, Michael Hässig, Peter Deplazes, Adrian B. Hehl

PII: S0020-7519(18)30018-3
DOI: https://doi.org/10.1016/j.ijpara.2017.10.007
Reference: PARA 4028

To appear in: *International Journal for Parasitology*

Received Date: 25 July 2017
Revised Date: 6 October 2017
Accepted Date: 21 October 2017

Please cite this article as: Lippuner, C., Ramakrishnan, C., Basso, W.U., Schmid, M.W., Okoniewski, M., Smith, N.C., Hässig, M., Deplazes, P., Hehl, A.B., RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites, *International Journal for Parasitology* (2018), doi: https://doi.org/10.1016/j.ijpara.2017.10.007

# RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites

Christoph Lippuner[a,e,1,4,*], Chandra Ramakrishnan[a,4,*], Walter U. Basso[a,2], Marc W. Schmid[b], Michal Okoniewski[c,3], Nicholas C. Smith[d], Michael Hässig[e], Peter Deplazes[a], Adrian B. Hehl[a]

[a] *Institute of Parasitology, University of Zurich, Winterthurerstrasse 266a, 8057, Zurich, Switzerland*

[b] *Service and Support for Science IT, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

[c] *Functional Genomics Center Zurich, Winterthurerstrasse 190, 8057, Zürich, Switzerland*

[d] *Research School of Biology, Australian National University, Canberra, ACT, 2601, Australia*

[e] *Department of Farm Animals, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, 8057 Zürich, Switzerland*

* Corresponding authors: chandra.ramakrishnan@uzh.ch or christoph.lippuner@dbmr.unibe.ch

[1]Current address: Department of Anaesthesiology and Pain Medicine, Inselspital, University of Bern, Bern, Switzerland; Department for BioMedical Research, University of Bern, Bern, Switzerland

[2]Current address: Institute of Parasitology, Vetsuisse Faculty and Faculty of Medicine, University of Bern, Länggass-Strasse 122, 3012 Bern, Switzerland

[3]Current address: ID Scientific IT Services, ETHZ, Zurich, Switzerland

[4]These authors contributed equally.

Supplementary data associated with this article.

**Abstract**

*Cryptosporidium parvum* is a major cause of diarrhoea in humans and animals. There are no vaccines and few drugs available to control *C. parvum*. In this study, we used RNA-Seq to compare gene expression in sporozoites and intracellular stages of *C. parvum* to identify genes likely to be important for successful completion of the parasite's life cycle and, thereby, possible targets for drugs or vaccines. We identified 3774 protein-encoding transcripts in *C. parvum.* Applying a stringent cut-off of eight fold for determination of differential expression, we identified 173 genes (26 coding for predicted secreted proteins) upregulated in sporozoites. On the other hand, expression of 1259 genes was upregulated in intestinal stages (merozoites/gamonts) with a gene ontology enrichment for 63 biological processes and upregulation of 117 genes in 23 metabolic pathways. There was no clear stage specificity of expression of AP2-domain containing transcription factors, although sporozoites had a relatively small repertoire of these important regulators. Our RNA-Seq analysis revealed a new calcium-dependent protein kinase, bringing the total number of known Calcium-dependent protein kinases (CDPKs) in *C. parvum* to 11. One of these, CDPK1, was expressed in all stages, strengthening the notion that it is a valid drug target. By comparing parasites grown in vivo (which produce bona fide thick-walled oocysts) and in vitro (which are arrested in sexual development prior to oocyst generation) we were able to confirm that genes for oocyst wall proteins are expressed in gametocytes and the proteins stockpiled rather than generated de novo in zygotes. RNA-Seq analysis of *C. parvum* revealed genes that are expressed in a stage-specific manner and others whose expression is required at all stages of development. The functional significance of these can now be addressed through recent advances in transgenics for *C. parvum*, and may lead to the identification of viable drug and vaccine targets.

*Keywords*: *Cryptosporidium parvum;* Apicomplexa; Calf; Transcriptome; Sporozoites; Intestinal stages; in vitro stages; Intervention targets

## 1. Introduction

*Cryptosporidium* is a genus of apicomplexan parasites recognised as one of the four main causes of diarrhoea in young children and, hence, a major contributor to early childhood mortality worldwide (Kotloff et al., 2013; Striepen, 2013; Checkley et al., 2015; Ryan and Hijjawi, 2015; Certad et al., 2017. Furthermore, Cryptosporidia are significant causes of potentially fatal diarrhoea in AIDS patients (Morgan et al., 2000). They are also extremely problematic in livestock, causing profuse diarrhoea and considerable economic losses in young calves and lambs in industrialised agriculture (Joachim et al., 2003; O'Handley and Olson, 2006). Although nearly 20 species and genotypes of *Cryptosporidium* have been recorded in humans (Ryan and Hijjawi, 2015), the vast majority of cases associated with diarrhoea in humans are caused by *Cryptosporidium hominis* or *Cryptosporidium parvum* (Certad et al., 2017) Whilst the former appears to be largely human-specific, the latter is also a significant cause of disease in cattle and, thus, a parasite of substantial zoonotic concern (Abeywardena et al., 2015).

Numerous developmental and molecular peculiarities indicate that *Cryptosporidium* spp. are more closely related to a primitive group of apicomplexans, the gregarines, than to coccidian parasites such as *Toxoplasma gondii* or *Eimeria* spp. (reviewed recently (Ryan and Hijjawi, 2015)). However, the developmental life cycles of, for example, *C. hominis* and *C. parvum,* are superficially similar to classic enteric coccidian development. Thus, infection follows a faecal-oral route and begins with the ingestion of sporulated oocysts. Sporozoites are released in the small intestine shortly after ingestion, invade enterocytes and form intracellular parasitophorous vacuoles at the apical surface of host cells, but remain extracytoplasmic. Within these, type I meronts develop, forming eight merozoites. Merozoites released from type I meronts either repeat this proliferative step or progress to type II meronts, containing four merozoites. The latter transform into the sexual stages, i.e., macrogametocytes, whose prominent wall-forming bodies can be readily detected in thin section transmission electron micrographs, or microgametocytes containing bullet-shaped microgametes, which are more difficult to identify. Fertilisation leads to the formation of oocysts that sporulate in the intestine (endogenous sporulation) without forming sporocysts and are immediately infectious upon shedding. There is an unresolved controversy about the ability of *Cryptosporidium* to replicate outside of a host organism (Clode et al., 2015).

There are no vaccines or chemoprophylactics to prevent infection with *Cryptosporidium* and very few chemotherapeutic options for its treatment (Checkley et al., 2015). New opportunities to uncover targets for drugs are inherent in the availability of genome sequence databases for *C. parvum* (Abrahamsen et al., 2004) and *C. hominis* (Abrahamsen et al., 2004) and the development of technologies for the genetic manipulation of the former parasite (Vinayak et al., 2015). The *Cryptosporidium* genome has several unique features: it is markedly reduced (at around 9 million

bases and <4,000 genes) compared with other enteric apicomplexans (e.g., ~52 million bases and >9,000 genes for *Eimeria tenella,* ~63 million bases and >8,000 genes for *T. gondii*), displaying streamlined metabolic pathways, minimal modes of energy production and an inability to synthesise essential building blocks for amino acid, purine, and fatty acid synthesis. It lacks an apicoplast, due to secondary reduction, and has a reduced mitochondrion (mitosome) without a role in energy metabolism (Keeling, 2004). Thus, *C. parvum* relies heavily on nutrient acquisition from the host (Abrahamsen et al., 2004; Xu et al., 2010) and, arguably, its commensal bacterial flora for intracellular survival. All these features of *Cryptosporidium* result in many of the unique anti-apicomplexan drug targets being absent. Moreover, the complexity and limitations of the available, reliably reproducible and highly productive in vitro cultivation protocols, as well as the lack of robust small animal models for reproducing the entire *Cryptosporidium* life cycle, make new target identification and validation somewhat challenging.

Comparative RNA-Seq analysis can be exploited to uncover critical molecules and pathways in enteric parasite development; we have recently used such approaches to highlight unique expression profiles in the development of *T. gondii* in the cat small intestine (Walker et al., 2015) and *E. tenella* in the caeca of chickens (Walker et al., 2015). In this study, we use similar strategies to compare *C. parvum* stages prepared from the small intestinal epithelium of experimentally infected calves with parasites propagated in vitro. The latter proliferate efficiently as merozoites but they apparently lack the ability to differentiate to oocysts in culture, and can be considered arrested in sexual development. Hence, we used comparative analysis of gene expression as a means to identify factors likely to be relevant for successful sexual development and oocyst formation. In contrast to most intestinal apicomplexans, sporulation ensues without delay in *Cryptosporidium*, resulting in excretion of infectious oocysts containing fully developed sporozoites, which can be considered a stable developmental end point in terms of their mRNA complement until they invade a new host cell to develop into first generation merozoites. Thus, RNA-Seq data of sporozoites from purified oocysts completes the full characterisation of developmental gene expression during the *Cryptosporidium* life cycle.

## 2. Materials and methods

### *2.1. Preparation of parasites and RNA*

Oocysts of a field strain of *C. parvum*, IPZ:CH-Crypto_K6769, were purified from faeces using a protocol published previously (Meloni and Thompson, 1996). The purity and identification of these oocysts as *C. parvum* "cattle genotype" was confirmed using PCR and direct sequencing of a fragment of the 18S rDNA using published methods (Ward et al., 2002). Oocysts of *C. parvum* IPZ:CH-Crypto_K6769 were stored in PBS with antibiotic/antimycotic solution (Sigma-Aldrich,

Switzerland) and passaged regularly through calves purchased from farms in the Canton of Zurich (Switzerland) at 1-3 days of age prior to experimental infections; calves were confirmed to be *Cryptosporidium*-free prior to experimental inoculation by modified Ziehl-Neelsen stain and by coproantigen detection (Bio-X Diagnostics®). The calves were housed in an experimental animal facility of the Institute of Parasitology of the University of Zürich, Switzerland. Individual calves were housed in pens of ~9 $m^2$ on a litter of straw and sawdust over a rubber mat on a concrete base. The pens were cleaned daily. Calves were fed ad libitum with a commercial milk substitute (UFA200) supplied by an automatic dispenser. Water and hay were also provided ad libitum.

Biological material for RNA extraction and RNA-Seq analysis was generated using in vivo and in vitro stages as well as freshly excysted sporozoites. For the former, a calf was inoculated orally with 2.5 x $10^9$ oocysts and euthanised 47 h p.i. (day 2). Subsequently, another calf was inoculated orally with 1.2 x $10^9$ oocysts and sacrificed 93 h p.i. (day 4). The small intestine was immediately removed and chilled on ice. All subsequent processing of tissue was done using pre-chilled solutions in a 4 °C cold room. Each small intestine was divided into five segments of equal length. In step 1, all segments were rinsed with PBS and then filled with HBSS (w/o $Mg^{2+}$/$Ca^{2+}$) / 2 mM DTT / 5 mM EDTA, ligated and incubated for 15-30 min. The HBSS wash fraction was then collected and stored on ice for parasite enrichment. In step 2, the five segments were each divided into two parts. Part 1 was filled with HBSS (w/o $Mg^{2+}$/$Ca^{2+}$) / 5 mM EDTA, ligated and incubated for another 15-20 min. The liquid was collected and added to the HBSS wash collected in Step 1. Part 2 was filled with 80 ml of trypsin-EDTA solution (Sigma, 3924, diluted 1:1 with PBS) and incubated for approximately 25 min. Trypsin was inactivated by adding 40 ml of FCS and this liquid was also added to the HBSS wash collected in Step 1. The combined material was pelleted and then mixed with RLT buffer (guanidine thiocyanate lysis buffer) (supplemented with 10 µl of 2-mercaptoethanol/ml buffer, RNeasy® Mini- or Midikit, QIAGEN). The resulting lysate was passed eight times through a 20-gauge needle fitted to an RNase-free syringe. The homogenised solution was then shock frozen in liquid nitrogen and stored at –80°C until RNA preparation.

In preparation for in vitro infections with *C. parvum,* HCT-8 cells (human ileocecal adenocarcinoma cells, ATCC, CCL-244) were cultured in a humidified incubator at 37°C and 5% $CO_2$ as described previously (Upton et al., 1995). Culture medium (RPMI 1640 with L- glutamine, 5% heat inactivated FCS, 5% Optimem+Glutamax, 1 mM sodium pyruvate, 50 U/ml of penicillin G, 50 U/ml of streptomycin, 0.25 µg/ml of amphotericin B, and 50 µg/ml of gentamycin) was changed every 48-72 h as required. Confluent (minimum 95%) T75 tissue culture flasks were used for in vitro culture. The monolayer was washed twice with pre-warmed PBS and cells were covered

with 5 ml of infection medium (RPMI 1640 with L-glutamine, 5% Optimem+Glutamax, 1 mM sodium pyruvate, 50 U/ml of penicillin G, 50 U/ml of streptomycin, 0.25 µg/ml of amphotericin B, and 50 µg/ml of gentamycin, 500 µg/ml of taurocholic acid sodium salt hydrate, 15 mM HEPES). To infect the HCT-8 cells, $4.6x10^6$ oocysts were re-suspended in 10 mM hydrochloric acid, incubated for 10 min at 37 °C and pelleted thereafter. The pellet was re-suspended with culture medium and was added to the HCT-8 cells. After 2 h of incubation, cells were washed twice with 10 ml of pre-warmed PBS to remove unexcysted oocysts and free sporozoites, and 15 ml of pre-warmed long-term infection medium (infection medium supplemented with 5 ml of heat inactivated FCS, 1 µg/ml of folic acid, 4 µg/ml of 4-aminobenzoic acid, 2 µg/ml of D-pantothenic acid hemicalcium, 0.1 U/ml of Caninsulin (Intervet), 35 µg/µl of L-ascorbic acid and, 50 mM glucose) were added. The medium was exchanged every 24 h, as described previously(Upton et al., 1995). After 48 h and 96 h of in vitro cultivation, the medium was removed and cells were washed twice with ice-cold PBS to remove free parasites, cell debris and floating cells. Another 10 ml of ice-cold PBS were added and infected cells were scraped away and pelleted at 4 °C. Pelleted material was mixed with RLT buffer (supplemented with 10 µl of 2-mercaptoethanol/ml buffer, RNeasy® Mini-Kit, QIAGEN, Switzerland). The cells were lysed by centrifugation using a QiaShredder (QIAGEN) and processed immediately for RNA preparation.

To isolate sporozoites, oocysts were centrifuged at 6000 *g* for 2 min at room temperature. The resulting pellet was re-suspended in 10 mM HCl and incubated for 10 min at 37°C. Parasites were centrifuged again and re-suspended in 2 mM sodium taurocholate (Sigma)/PBS and incubated at 37°C for 16 min. Excystation rates of over 95% were reached. Pelleted parasites were mixed with RLT buffer (supplemented with 10 µl of 2-mercaptoethanol/ml of buffer, RNeasy® Mini-Kit, QIAGEN). The lysate was centrifuged through a QiaShredder (QIAGEN) and immediately processed for RNA preparation.

Total RNA was isolated from the various samples using a RNeasy® mini- or midikit (QIAGEN) following the manufacturer's instructions including an on-column DNase I digestion (RNase free DNase-Set, QIAGEN). Total RNA was enriched for mRNA in two rounds using the MicroPolyA Purist Kit (Thermo Fisher Scientific, Switzerland). The RNA was quantified using a Qubit fluorometer (Thermo Fisher Scientific, Switzerland) and its quality was analysed with a Bioanalyzer 2100 (Agilent, Switzerland) and the Agilent 6000 Pico kit (Supplementary Fig. S1).

*2.2. cDNA library construction and RNA-Seq*

Genome-wide transcriptome libraries were produced from duplicate biological samples for parasites harvested from calf intestines 2 and 4 days p.i. with *C. parvum* oocysts, parasites

harvested from in vitro cultures 2 and 4 days post-inoculation, and from sporozoites purified from oocysts, using protocols previously described by us (Schnyder et al., 2009). The transcriptome libraries (0.5 pM) were used for emulsion PCR. The barcoded libraries were pooled prior to performing emulsion PCR and the resulting beads were loaded onto either full SOLiD 4 slides (Applied Biosystems, Switzerland), for the in vivo samples, or on to SOLiD 5500xl chips (Applied Biosystems) for in vitro and sporozoite samples, following the manufacturer's instructions. SOLiD ToP Sequencing chemistry was used to produce pair end (50 bp + 35 bp for SOLiD 4, 75 bp + 35 bp for SOLiD 5500xl) sequencing reads.

Short reads generated in this study were deposited at NCBI Sequence Read Archive (SRA, www.ncbi.nlm.nih.gov/sra) and are accessible through the accession number **SRP069013**. To remove potential contamination with host RNA, reads of the in vitro and the in vivo/sporozoite samples were first aligned to the human or bovine genomes, respectively (ensembl82) with Subread (version 1.4.6-p4, options -B 10 -m 5; (Liao et al., 2013)). Reads without an alignment were then aligned to the *C. parvum* Iowa II genome (CryptoDB 8.0) with Subread allowing up to 10 alignments per read (options: -H -B 10 -m 5, only forward F5 reads). Count tables were generated with the annotation from CryptoDB (release 8.0, http://cryptodb.org/common/downloads/release-8.0/CparvumIowaII/gff/data/CryptoDB-8.0_CparvumIowaII.gff) and with Rcount (Schmid and Grossniklaus, 2015) with an allocation distance of 50 bp for calculation of the weights of the reads with multiple alignments, consideration of the strand information, and a minimal number of five hits. Technical replicates (two each for sporozoites, in vivo 2 days p.i. and in vivo 4 days p.i.) were summed. Read counts of < 5 were considered not expressed and were set to 0. Variation in gene expression was analysed with a general linear model in R with the package DESeq2 (version 1.10.0, (Love et al., 2014)) according to a factorial design with the two explanatory factors ORIG (sporozoites, in vitro, and in vivo) and TIME (2/4 days p.i.) combined into a single factor. Specific conditions were compared with linear contrasts and *P* values were adjusted for multiple testing (Benjamini-Hochberg). Genes with an adjusted *P* value below 0.05 were considered to be differentially expressed.

### *2.3. BLAST, Blast2GO and alignments*

Protein domains were identified using a Simple Modular Architecture Research Tool (SMART (Schultz et al., 1998). A protein BLAST search was performed using the NCBI webtool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with subsequent SmartBLAST analysis. Protein sequences retrieved from CryptoDB (Puiu et al., 2004) release 26, www.cryptodb.org) were used for a BLAST search using Blast2GO (version 3.1.3). Alignments were performed using the MUSCLE tool on http://www.ebi.ac.uk/Tools/msa/muscle/.

*2.4. Ethics statement*

Housing of, and experimentation on, calves was conducted with approval and under the supervision of the Cantonal Veterinary Office and the Ethics Committee of the Canton of Zurich (Kantonales Veterinaeramt Zurich, Obstgartenstrasse 21, 8090, Zurich, Switzerland) and carried out according to Swiss Law and Guidelines on Animal Welfare, and the specific regulations of the Canton of Zurich under Permit Number 161/2009.

## 3. Results and Discussion

*3.1. Isolation of parasites and RNA-Seq analysis*

A completed and annotated genome sequence for *C. parvum* (CryptoDB, release 8.0 and Blast2GO, Supplementary Table S1) provided the basis for analysis of gene expression during the life cycle (Abrahamsen et al., 2004). Previous studies of *C. parvum* gene expression considered mRNA levels of specific sets of genes and/or developmental stages (Mullapudi et al., 2007; Zhang et al., 2012), or used proteomics to analyse a specific stage (i.e. sporozoites) (Sanderson et al., 2008; Siddiki, 2013). Although the latter gave valuable qualitative information on expression of approximately one-third of the ~3700 gene models in CryptoDB, the information was neither comprehensive nor suitable for comparative analysis. Here, we used quantitative RNA-Seq analysis to explore gene expression at a genome-wide level and in different developmental stages isolated form experimentally infected animals and cell cultures.

RNA was isolated from the in vitro cultivated *C. parvum*, infected calf intestines (Supplementary Fig. S2A) and excysted sporozoites (Supplementary Fig. S2B), enriched for mRNA and sequenced on a SOLiD platform. For read mapping, we used the *C. parvum* Iowa II genome sequence.

To identify differentially expressed genes, we used the genome annotation from CryptoDB (release 8.0) and Blast2GO (Supplementary Table S1). Correlations in gene expression between the samples were plotted (Supplementary Fig. S3); due to the distinct gene expression patterns in the sporozoite samples, the comparative analyses of the proliferation stages isolated from calves or cell culture were performed separately. As expected, the highest number of reads mapping to *C. parvum* genes was obtained for sporozoites due to low contamination of the sample with host material (Supplementary Table S2). Overall, we found evidence for expression of 3763 genes (3802 including the sporozoite data), from which transcripts not encoding proteins (e.g. rRNAs, small nucleolar RNAs (snoRNAs), tRNA, etc.) were removed to obtain 3744 (3774 with sporozoites) protein-encoding transcripts in total (Supplementary Table S3 and S4). Normalisation and differential analyses were performed using DEseq2 (Love et al., 2014). Pair-wise, hierarchical

clustering using all genes in all samples showed that the sporozoite sample formed a separate clade whilst the in vivo and the in vitro samples formed two distinctly separate clades (Fig. 1A). The four in vitro samples as well as the two in vivo samples appear highly correlated and all six samples are more similar to each other than to the sporozoite datasets. For a more in-depth analysis of the in vivo and in vitro samples, the data were subjected to correlation analysis excluding the sporozoite samples (Fig. 1B). All in vitro samples formed one clade with the two replicates at each time point forming subclades. However, there was a high overall correlation between the samples. Only minor differences in gene expression were observed between the time points in the in vitro and in vivo samples, probably due to a mixture of various developmental stages. This highlights the difficulties in sampling distinct developmental stages for *C. parvum* whose life cycle is very rapid. We therefore combined the two time points for the in vitro and in vivo samples for differential gene expression analysis. One thousand, four hundred and forty-one differentially expressed genes (i.e., >4-fold difference in normalised reads) were identified with more than half of these genes being down-regulated in sporozoites (Fig. 1C). More detailed analyses of these samples, excluding the sporozoite dataset, showed that 496 genes are differentially expressed (i.e., >4-fold difference in normalised reads) between both in vitro samples and in vivo samples, with approximately half of the genes down-regulated in in vitro samples and half upregulated (Fig. 1D). Differential expression between sporozoites and the in vivo or in vitro stages, as well as between the latter two stages, was analysed (using data including and excluding the sporozoites; non-protein coding, snoRNA, rRNA and tRNA removed; Supplementary Tables S5-S10). Comparison of the up- and down-regulated genes between the samples showed that more than 1000 genes were more highly expressed in sporozoites than in the in vivo or in vitro stages, respectively (Fig. 1E), which corresponds to 30.9% and 27.6% of the total protein coding genes. In contrast, expression of 559 and 605 genes was lower in sporozoites compared with the in vivo or in vitro stages, respectively, representing 14.7% and 15.9% of all protein-encoding genes. Three hundred and twenty-eight genes were more highly expressed in the in vivo stages compared with the in vitro sample, whilst expression of 220 genes was lower, corresponding to 8.6% and 5.8%, respectively, of all genes (Fig. 1E).

*3.2. Analysis of differential expression in vivo and in vitro*

Propagation of *C. parvum* in continuous culture in hollow fiber systems (Morada et al., 2016) has yielded bona fide thick-walled oocysts whilst less complex culture systems on cell monolayers (Upton et al., 1994a; Upton et al., 1994b; Castellanos-Gonzalez et al., 2013) led to production of thin walled oocysts only, if any (Arrowood, 2002). We attempted to explore this apparent arrest in development of bona fide oocysts to identify differences in expression profiles between unproductive sexual development in in vitro culture and infection in vivo. For in vitro

cultivation, we essentially followed the protocol for infection of HCT-8 cell monolayers (Upton et al., 1995), which resulted in robust amplification of the parasite population but did not yield any observable oocysts within the 4 day time frame used in these experiments. Cultured parasites were harvested at the same time points as those in experimental infections (i.e., 2 days and 4 days p.i.) for RNA extraction and analysis of gene expression. Application of a higher cut-off for differential expression, i.e., at least four-fold difference in mRNA levels (normalised reads/kb), was used to compare in vivo and in vitro samples. Due to the minor differences (Fig. 1B), biological replicate samples (in vitro samples only) were averaged, as were sampling time points, allowing for a global comparison of in vitro and in vivo gene expression (Supplementary Table S5-S10). Combining the day 2 and 4 samples for both the in vitro and in vivo sample sets is further supported by the observation that the sexual phase of development, exemplified here by expression of *Cryptosporidium* oocyst wall protein (COWP) genes (Fig. 2), occurs as early as day 2 p.i. in vivo, as indicated in the existing literature (O'Donoghue, 1995). Using these criteria, we found higher expression of 55 and 253 genes in vitro and in vivo, respectively. However, parsing of those samples using gene ontology (Blast2Go) revealed no basis for a robust biological interpretation of the developmental block of in vitro cultured parasites. Nevertheless, the analysis showed that 59 of the 253 significantly more highly expressed genes in the combined in vivo dataset encode for predicted secreted proteins. This group also contains eight genes coding for proteins larger than 100 kDa. Among those is CpCOWP1, one of the three most abundant oocyst wall proteins (Sanderson et al., 2008) (see also Section 3.5). Whether the seven other very large secreted proteins in this group are as yet unidentified structural components of the oocyst remains to be investigated. Conversely, in the group of genes with higher expression in vitro, parsing according to gene ontology terms showed a surprisingly large representation of genes involved in biological processes such as DNA/chromatin organisation, DNA metabolism, and nucleosome organisation (43% of all hits), whilst only six genes coding for secreted proteins were identified. The significant expression of major and minor COWPs in vitro points to an arrest late in sexual development, which may explain the paucity of biological information gained from a comparison with in vivo stages.

### *3.3. Differential gene expression in sporozoites*

Correlation analysis revealed major differences in gene expression in sporozoites compared with parasites replicating intracellularly (Fig. 1A). Using stringent cut-offs for calling regulated genes in a comparison of sporozoites and in vivo samples, we identified only 161 genes (25 coding for predicted secreted proteins) with at least eight-fold higher normalised reads in sporozoites. On the other hand, 1095 genes were at least eight-fold more highly expressed in intestinal stages (merozoites/gamonts, Supplementary Table S5-S10) with a gene ontology enrichment for 63

biological processes and significant upregulation of 117 genes in 23 metabolic pathways. General upregulation of metabolism features most prominently in this comparative analysis after parsing the subset of intestinal stage-regulated genes according to biological processes or molecular function. While the mechanisms of gene expression regulation in *Cryptosporidium* are not well explored (Mullapudi et al., 2007; Oberstaller et al., 2013; Oberstaller et al., 2014), systematic analysis of genes with similar expression profiles from cultured parasites has produced direct evidence for co-regulation by ApiAP2 and other transcription factors (Oberstaller et al., 2013; Oberstaller et al., 2014). Using the datasets produced by Oberstaller et al. (2013) we could not find clustered distribution for any of the 200 co-regulated sets of genes. Similarly, we found that genes expressed at higher levels in sporozoites or the intestinal stage did not cluster more frequently on the eight *C. parvum* chromosomes than random samples of equal size (data not shown).

*3.4. Expression profiles of the C. parvum ApiAP2 transcription factors*

The apicomplexan Apetala2 transcription factors (apiAP2) have been shown to regulate gene expression before or during stage transition (Yuda et al., 2009; Yuda et al., 2010; Iwanaga et al., 2012; Radke et al., 2013; Walker et al., 2013a; Yuda et al., 2015) as well as be involved in virulence and host cell invasion (Walker et al., 2013b). A comprehensive analysis of the AP2-domain containing transcription factors in *C. parvum* has been described by Oberstaller et al. (Oberstaller et al., 2014). Using their list of identified transcription factors and our Blast2GO analysis (Supplementary Table S1), we analysed their expression in the different life cycle stages of *C. parvum*. We found that two putative genes encoding AP2-domain containing transcription factors, cgd8_3230 (predicted to be an ApiAP2XI-3) and cgd6_2600 are >4-fold upregulated in sporozoites compared with the in vitro and >20-fold compared to the in vivo stages (Fig. 3). Interestingly, sporozoites seem to have a small repertoire of ApiAP2. The only gene upregulated in vitro is the predicted ApiAP2IV-2 gene (cgd5_4250) showing >2-fold more reads than the in vivo sample and >12-fold more than the sporozoites. More than four-fold and two-fold upregulation in vivo versus in vitro can be seen for cgd5_2570 and cgd6_1140 (predicted as apiAP2VIII-5 and ApiAP2X-11), respectively. However, none of these regulated transcription factors are exclusively stage-specific. This may be linked to the observed redundancy in the *C. parvum* ApiAP2-binding motifs (Oberstaller et al., 2014). Further analyses of the downstream targets and stage specificity can now be addressed through recent developments in transgenesis (Vinayak et al., 2015).

*3.5. Macrogamete-associated genes*

Components of the oocyst walls in *Cryptosporidium* spp. are secreted from wall forming bodies (Spano et al., 1997; Aldeyarbi and Karanis, 2016) in the macrogametocyte that contain inter

alia COWP proteins. As expression of COWP genes is specific to macrogametocytes, they are good indicators for sexual stage development (Smith et al., 2005). Consistent with previous observations (Upton et al., 1995), we found considerable expression of the more abundant COWPs in the culture-derived samples at both time-points but not as high as in the samples collected from experimentally infected calves (Fig. 2). Expression of all COWPs in sporozoites was silenced.

*3.6. Mucins*

Mucins are heavily glycosylated proteins that are involved in *Cryptosporidium* host cell attachment and/or invasion (Cevallos et al., 2000; O'Connor et al., 2009) and tethering of sporozoites to the oocyst wall (Chatterjee et al., 2010), and are bound by the lectin *Maclura pomifera* agglutinin (MPA). Our data revealed that gp900 (cgd7_4020), previously shown to be on the sporozoite surface and involved in mediating host cell attachment (Cevallos et al., 2000), is expressed mainly in in vivo stages and to a lesser extent in sporozoites (Fig. 4). However, we detected mucin gp40/15 (cgd6_1080) in in vivo and in vitro stages, but not in sporozoites (Fig. 4) in contrast to previous studies where the protein was detected in merozoites and sporozoites (O'Connor et al., 2007) and shown to play a role in host cell attachment (Cevallos et al., 2000). CpMuc5 (cgd2_430) has been shown to be expressed in all intestinal stages and has been localized to merozoites and sporozoites (O'Connor et al., 2009); our data confirm expression in all stages with upregulation in sporozoites (Fig. 4). Only another mucin gene, cgd4_3550, present in all stages, is up-regulated in sporozoites (Fig. 4). Taken together, we find mucins mainly expressed in the intestinal stages with very few expressed additionally in sporozoites.

Based on this analysis, we conclude that macrogamete development including expression of oocyst wall genes occurs readily in vitro. Further research is required to address the question of whether successful development to thick-walled, environmentally resistant and infectious oocysts is dependent on additional factors conferred by the cell culture system (e.g., a more complex cellular ecosystem) or on fertilisation by microgametes.

Chatterjee et al. (2010) (Chatterjee et al., 2010) used MS analysis of excysted or sonicated oocyst walls to determine the abundance of oocyst wall proteins (COWP1-9). Their findings correlate exactly with our RNA-Seq analysis, i.e., COWPs 1, 6 and 8 being the most abundant, followed by significantly lower levels of COWPs 2, 3 and 4. COWPs 5, 7, and 9 are present in very low amounts. However, whilst Chatterjee et al. (2010) concluded that CpMPA4 is an oocyst wall-associated protein, our analyses reveal that it is, in fact, most highly expressed in sporozoites.

*3.7. Microgamete-associated genes*

In contrast to biflagellate microgametes from Coccidia, *Cryptosporidium* male gametes are bullet shaped with a large, elongated nucleus surrounded by microtubules and an apical adhesive zone (Vetterling et al., 1971; Bird and Smith, 1980; Goebel and Braendler, 1982; Landsberg and Paperna, 1986; Fayer et al., 1990; Ostrovska and Paperna, 1990; Cheadle et al., 1999; Alvarez-Pellitero and Sitja-Bobadilla, 2002; Akiyoshi et al., 2003). To date, no exclusively flagellar components have been identified (Bird and Smith, 1980) and it is not known if microgametes move actively to fertilise the macrogamete. Using the annotation from CryptoDB, Blast2GO and known flagellar components from *Plasmodium* (Sinden et al., 2010), we were able to detect only a few genes in our transcriptomes that could potentially be associated with microgametes. Due to the non-flagellar structure of *Cryptosporidium* microgametes, it is not surprising that intraflagellar transport proteins and radial spoke components are absent from the genome. Few dyneins can be detected in the genome, of which a putative heavy chain is barely expressed and the light chain, Tctex1 – one of the three potential light chains – is absent in sporozoites (Supplementary Fig. S4A). We cannot make firm conclusions about the stage specificity of these genes. Interestingly, out of three genes with a gene ontology (GO) term for motile cilium, a gene annotated as a flagellar outer arm protein, as well as a hypothetical protein (cgd5_1550 and cgd8_1690, Supplementary Table S1), are exclusively expressed in in vivo stages (Supplementary Fig. S4B). The basal body component, SAS-6, that in *Plasmodium* is only present in male gametocytes and plays a role in flagellum assembly (Marques et al., 2015), is also present in *C. parvum* albeit the protein is very divergent (Marques et al., 2015). It is expressed in all stages (Supplementary Fig. S4C) and, similar to the *T. gondii* orthologue, seems not to be sexual stage-specific (de Leon et al., 2013). We anticipated that *C. parvum* might possess a HAP2-like protein. HAP2 was first shown to be a male gamete fusion factor in *Arabidopsis* (Liu et al., 2008) and later in *Chlamydomonas reinhardtii* and *Plasmodium berghei* (Liu et al., 2008), and has also been detected in the sexual stages of *E. tenella* (Walker et al., 2015). Using the amino acid sequence from *P. berghei* HAP2 for a BLAST search in CryptoDB, we identified a protein (CryptoDB ID cgd8_2220) that Blast2GO annotated as a male gamete fusion factor (Supplementary Table S1). Using the sequence of the *Cryptosporidium reinhardtii* H/G domain (Liu et al., 2015) to define the boundaries, we aligned putative *C. parvum* HAP2 to *C. reinhardtii*, *P. berghei* and *T. gondii* HAP2 amino acid sequences and found a corresponding domain (Supplementary Fig. S4D). However, cgd8_2220 is expressed in all stages of *C. parvum* and at equal levels and, therefore, can be ruled out as a marker for microgametes. On the other hand, annotated centrins, which are calcium-binding proteins from the centrosome and have been shown in *P. falciparum* to be upregulated in gametocytes (Mahajan et al., 2008), seem to be upregulated in vivo and absent in sporozoites, indicating a potential role in sexual reproduction (Supplementary Fig. S4E). Similarly, α- and β-tubulin are upregulated in in vivo stages and not

present in sporozoites (Supplementary Fig. S4F). Interestingly, α-tubulin seems to be an orthologue of the putative *T. gondii* α-tubulin I that is upregulated in the feline enteroepithelial stages (Ramakrishnan et al., unpublished data). Taken together, we have identified a few in vivo stage-specific genes that may be associated specifically with microgametes. This stage remains elusive for further study without any reproducible in vitro culturing system for sexual stages or specific isolation procedures from calf intestines in place.

*3.8. Meiosis-specific genes*

During meiosis, programmed double strand breaks (DSBs) occur that have to be repaired eventually by homologous recombination, single strand annealing or non-homologous end-joining (reviewed in (Norbury and Hickson, 2001)). Using CryptoDB and Blast2GO annotations, we identified Rad51, which is involved in homology recognition and strand exchange, and two of its homologues as well as a Dmc1-like protein and a Rad54-like protein; Dmc1 is a recombinase whilst Rad54 is helicase. All of these genes are expressed at very low levels in the sporozoite but, except for cgd4_2050, none of them seem exclusive to in vivo stages (Supplementary Fig. S5A). This is especially interesting for the DMC1-like protein, which is described as meiosis-specific in yeast (Bishop et al., 1992). Similarly, out of the nucleases, the meiosis-specific *spo11* gene is expressed in the in vivo and in vitro stages. We identified a putative homologue of Meu13 (cgd2_510) using the blastp tool in CryptoDB and the protein sequence from fission yeast. This gene is also expressed in the in vivo and in vitro stages, but its putative association partner *mcp7* was only detected in the early in vivo stage (Supplementary Fig. S5A). Taken together, our results may indicate that parasites in in vitro cultures are partially prepared to differentiate to sexual stages on a transcriptional level.

*3.9. Calcium-dependent protein kinases*

Calcium-dependent protein kinases (CDPKs) are found in plants and protists but not in fungi or animals (reviewed in (Harper and Harmon, 2005), rendering them attractive drug targets. Calcium signaling in apicomplexans regulates secretion, motility, invasion and egress (reviewed in (Billker et al., 2009)). Interestingly, *C. parvum* and *T. gondii* CDPK1 can be selectively and potently inhibited by a class of kinase inhibitors, which results in a block in host cell invasion (Murphy et al., 2010). Etzold et al. (2014) identified seven CDPKs in *C. parvum* whilst eight genes and two CDPK-like proteins had been described previously (Artz et al., 2011). Using Blast2GO, we identified an additional CDPK (cgd7_1260, Supplementary Fig. S6A). We used blastp and Smart BLAST to identify the closest homologues in other apicomplexans and *Arabidopsis thaliana,* and aligned the predicted serine/threonine protein kinase domains (Supplementary Fig. S6B). Although

the catalytic domain can be clearly identified, the kinase function is probably abolished in cgd7_1260 due to the lack of a critical aspartate. The *C. parvum* CDPK1 gene is expressed in all stages (Supplementary Fig. S6A). An effective drug against cryptosporidiosis needs to be effective against the merozoites/schizonts as well as sexual stages and, thus, this expression profile strengthens the notion that CDPK1 is a valid drug target. CDPK6 also is expressed in all stages at comparable levels but all the other kinase genes show lower transcript abundance in the merozoites/schizonts and sexual stages. CDPK2, 2A and cgd5_2270 are upregulated in sporozoites (Supplementary Fig. S6A) and may play an important role in sporozoite motility and invasion.

## 4. Conclusions

RNA-Seq analysis of sporozoites and intracellular stages of *C. parvum* grown in vivo or in vitro has identified some genes that are expressed in a stage-specific manner and others whose expression is required at all stages of development. These new data sets provide valuable information for functional genomic studies that can now be addressed through recent advances in transgenics for *C. parvum* (Vinayak et al., 2015), and may lead to the identification of new, viable drug and vaccine targets.

## Acknowledgements

This work was supported by research fellowships from the University of Zurich, Switzerland, to CL and WUB. We thank Hans-Peter Müller for valuable assistance in animal care.

**References**

Abeywardena, H., Jex, A.R., Gasser, R.B., 2015. A perspective on *Cryptosporidium* and *Giardia*, with an emphasis on bovines and recent epidemiological findings. Adv Parasitol 88, 243-301.

Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear, P.H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., Kapur, V., 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science 304, 441-445.

Akiyoshi, D.E., Dilo, J., Pearson, C., Chapman, S., Tumwine, J., Tzipori, S., 2003. Characterization of *Cryptosporidium meleagridis* of human origin passaged through different host species. Infect Immun 71, 1828-1832.

Aldeyarbi, H.M., Karanis, P., 2016. The fine structure of sexual stage development and sporogony of *Cryptosporidium parvum* in cell-free culture. Parasitology, 1-13.

Alvarez-Pellitero, P., Sitja-Bobadilla, A., 2002. *Cryptosporidium molnari* n. sp. (Apicomplexa: Cryptosporidiidae) infecting two marine fish species, *Sparus aurata* L. and *Dicentrarchus labrax* L. Int J Parasitol 32, 1007-1021.

Arrowood, M.J., 2002. In vitro cultivation of *Cryptosporidium* species. Clin Microbiol Rev 15, 390-400.

Artz, J.D., Wernimont, A.K., Allali-Hassani, A., Zhao, Y., Amani, M., Lin, Y.H., Senisterra, G., Wasney, G.A., Fedorov, O., King, O., Roos, A., Lunin, V.V., Qiu, W., Finerty, P., Jr., Hutchinson, A., Chau, I., von Delft, F., MacKenzie, F., Lew, J., Kozieradzki, I., Vedadi, M., Schapira, M., Zhang, C., Shokat, K., Heightman, T., Hui, R., 2011. The *Cryptosporidium parvum* kinome. BMC Genomics 12, 478.

Billker, O., Lourido, S., Sibley, L.D., 2009. Calcium-dependent signaling and kinases in apicomplexan parasites. Cell Host Microbe 5, 612-622.

Bird, R.G., Smith, M.D., 1980. Cryptosporidiosis in man: parasite life cycle and fine structural pathology. J Pathol 132, 217-233.

Bishop, D.K., Park, D., Xu, L., Kleckner, N., 1992. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell 69, 439-456.

Castellanos-Gonzalez, A., Cabada, M.M., Nichols, J., Gomez, G., White, A.C., Jr., 2013. Human primary intestinal epithelial cells as an improved in vitro model for *Cryptosporidium parvum* infection. Infect Immun 81, 1996-2001.

Certad, G., Viscogliosi, E., Chabe, M., Caccio, S.M., 2017. Pathogenic Mechanisms of *Cryptosporidium* and *Giardia*. Trends Parasitol 33, 561-576.

Cevallos, A.M., Bhat, N., Verdon, R., Hamer, D.H., Stein, B., Tzipori, S., Pereira, M.E., Keusch, G.T., Ward, H.D., 2000. Mediation of *Cryptosporidium parvum* infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody. Infect Immun 68, 5167-5175.

Chatterjee, A., Banerjee, S., Steffen, M., O'Connor, R.M., Ward, H.D., Robbins, P.W., Samuelson, J., 2010. Evidence for mucin-like glycoproteins that tether sporozoites of *Cryptosporidium parvum* to the inner surface of the oocyst wall. Eukaryot Cell 9, 84-96.

Cheadle, M.A., Toivio-Kinnucan, M., Blagburn, B.L., 1999. The ultrastructure of gametogenesis of *Cryptosporidium baileyi* (eimeriorina; cryptosporidiidae) in the respiratory tract of broiler chickens (*Gallus domesticus*). J Parasitol 85, 609-615.

Checkley, W., White, A.C., Jr., Jaganath, D., Arrowood, M.J., Chalmers, R.M., Chen, X.M., Fayer, R., Griffiths, J.K., Guerrant, R.L., Hedstrom, L., Huston, C.D., Kotloff, K.L., Kang, G., Mead, J.R., Miller, M., Petri, W.A., Jr., Priest, J.W., Roos, D.S., Striepen, B., Thompson, R.C., Ward, H.D., Van Voorhis, W.A., Xiao, L., Zhu, G., Houpt, E.R., 2015. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. Lancet Infect Dis 15, 85-94.

Clode, P.L., Koh, W.H., Thompson, R.C., 2015. Life without a Host Cell: What is *Cryptosporidium*? Trends Parasitol 31, 614-624.

de Leon, J.C., Scheumann, N., Beatty, W., Beck, J.R., Tran, J.Q., Yau, C., Bradley, P.J., Gull, K., Wickstead, B., Morrissette, N.S., 2013. A SAS-6-like protein suggests that the *Toxoplasma conoid* complex evolved from flagellar components. Eukaryot Cell 12, 1009-1019.

Etzold, M., Lendner, M., Daugschies, A., Dyachenko, V., 2014. CDPKs of *Cryptosporidium parvum*--stage-specific expression in vitro. Parasitol Res 113, 2525-2533.

Fayer, R., Speer, C.A., Dubey, J.P., 1990. Cryptosporidiosis in Man and Animals. CRC Press, Boca Raton.

Goebel, E., Braendler, U., 1982. Ultrastructure of microgametogenesis, microgametes and gametogamy of *Crytosporidium* sp. in the small intestine of mice. Protistologica 18, 14.

Harper, J.F., Harmon, A., 2005. Plants, symbiosis and parasites: a calcium signalling connection. Nat Rev Mol Cell Biol 6, 555-566.

Iwanaga, S., Kaneko, I., Kato, T., Yuda, M., 2012. Identification of an AP2-family protein that is critical for malaria liver stage development. PLoS One 7, e47557.

Joachim, A., Krull, T., Schwarzkopf, J., Daugschies, A., 2003. Prevalence and control of bovine cryptosporidiosis in German dairy herds. Vet Parasitol 112, 277-288.

Keeling, P.J., 2004. Reduction and compaction in the genome of the apicomplexan parasite *Cryptosporidium parvum*. Dev Cell 6, 614-616.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., Wu, Y., Sow, S.O., Sur, D., Breiman, R.F., Faruque, A.S., Zaidi, A.K., Saha, D., Alonso, P.L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J.B., Omore, R., Oundo, J.O., Hossain, A., Das, S.K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R.A., Antonio, M., Hossain, M.J., Akinsola, A., Mandomando, I., Nhampossa, T., Acacio, S., Biswas, K., O'Reilly, C.E., Mintz, E.D., Berkeley, L.Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R.M., Levine, M.M., 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382, 209-222.

Landsberg, J.H., Paperna, I., 1986. Ultrastructural study of the coccidian *Cryptosporidium* sp. from stomachs of juvenile cichlid fish. Dis. Aquat.Organ. 2, 8.

Liao, Y., Smyth, G.K., Shi, W., 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41, e108.

Liu, Y., Pei, J., Grishin, N., Snell, W.J., 2015. The cytoplasmic domain of the gamete membrane fusion protein HAP2 targets the protein to the fusion site in *Chlamydomonas* and regulates the fusion reaction. Development 142, 962-971.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., Billker, O., 2008. The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. Genes Dev 22, 1051-1068.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

Mahajan, B., Selvapandiyan, A., Gerald, N.J., Majam, V., Zheng, H., Wickramarachchi, T., Tiwari, J., Fujioka, H., Moch, J.K., Kumar, N., Aravind, L., Nakhasi, H.L., Kumar, S., 2008. Centrins, cell cycle regulation proteins in human malaria parasite *Plasmodium falciparum*. J Biol Chem 283, 31871-31883.

Marques, S.R., Ramakrishnan, C., Carzaniga, R., Blagborough, A.M., Delves, M.J., Talman, A.M., Sinden, R.E., 2015. An essential role of the basal body protein SAS-6 in *Plasmodium* male gamete development and malaria transmission. Cell Microbiol 17, 191-206.

Meloni, B.P., Thompson, R.C., 1996. Simplified methods for obtaining purified oocysts from mice and for growing *Cryptosporidium parvum* in vitro. J Parasitol 82, 757-762.

Morada, M., Lee, S., Gunther-Cummins, L., Weiss, L.M., Widmer, G., Tzipori, S., Yarlett, N., 2016. Continuous culture of *Cryptosporidium parvum* using hollow fiber technology. Int J Parasitol 46, 21-29.

Morgan, U., Weber, R., Xiao, L., Sulaiman, I., Thompson, R.C., Ndiritu, W., Lal, A., Moore, A., Deplazes, P., 2000. Molecular characterization of *Cryptosporidium* isolates obtained from human immunodeficiency virus-infected individuals living in Switzerland, Kenya, and the United States. J Clin Microbiol 38, 1180-1183.

Mullapudi, N., Lancto, C.A., Abrahamsen, M.S., Kissinger, J.C., 2007. Identification of putative cis-regulatory elements in *Cryptosporidium parvum* by de novo pattern finding. BMC Genomics 8, 13.

Murphy, R.C., Ojo, K.K., Larson, E.T., Castellanos-Gonzalez, A., Perera, B.G., Keyloun, K.R., Kim, J.E., Bhandari, J.G., Muller, N.R., Verlinde, C.L., White, A.C., Jr., Merritt, E.A., Van Voorhis, W.C., Maly, D.J., 2010. Discovery of Potent and Selective Inhibitors of Calcium-Dependent Protein Kinase 1 (CDPK1) from *C. parvum* and *T. gondii*. ACS Med Chem Lett 1, 331-335.

Norbury, C.J., Hickson, I.D., 2001. Cellular responses to DNA damage. Annu Rev Pharmacol Toxicol 41, 367-401.

O'Connor, R.M., Burns, P.B., Ha-Ngoc, T., Scarpato, K., Khan, W., Kang, G., Ward, H., 2009. Polymorphic mucin antigens CpMuc4 and CpMuc5 are integral to *Cryptosporidium parvum* infection in vitro. Eukaryot Cell 8, 461-469.

O'Donoghue, P.J., 1995. *Cryptosporidium* and cryptosporidiosis in man and animals. Int J Parasitol 25, 139-195.

O'Handley, R.M., Olson, M.E., 2006. Giardiasis and cryptosporidiosis in ruminants. Vet Clin North Am Food Anim Pract 22, 623-643.

O'Connor, R.M., Wanyiri, J.W., Cevallos, A.M., Priest, J.W., Ward, H.D., 2007. *Cryptosporidium parvum* glycoprotein gp40 localizes to the sporozoite surface by association with gp15. Mol Biochem Parasitol 156, 80-83.

Oberstaller, J., Joseph, S.J., Kissinger, J.C., 2013. Genome-wide upstream motif analysis of *Cryptosporidium parvum* genes clustered by expression profile. BMC Genomics 14, 516.

Oberstaller, J., Pumpalova, Y., Schieler, A., Llinas, M., Kissinger, J.C., 2014. The *Cryptosporidium parvum* ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems. Nucleic Acids Res 42, 8271-8284.

Ostrovska, K., Paperna, I., 1990. *Cryptosporidium* sp. of the starred lizard Agama stellio: Ultrastructure and life cycle. Parasitol Res 76, 712-720.

Puiu, D., Enomoto, S., Buck, G.A., Abrahamsen, M.S., Kissinger, J.C., 2004. CryptoDB: the *Cryptosporidium* genome resource. Nucleic Acids Res 32, D329-331.

Radke, J.B., Lucas, O., De Silva, E.K., Ma, Y., Sullivan, W.J., Jr., Weiss, L.M., Llinas, M., White, M.W., 2013. ApiAP2 transcription factor restricts development of the *Toxoplasma* tissue cyst. Proc Natl Acad Sci U S A 110, 6871-6876.

Ryan, U., Hijjawi, N., 2015. New developments in *Cryptosporidium* research. Int J Parasitol 45, 367-373.

Sanderson, S.J., Xia, D., Prieto, H., Yates, J., Heiges, M., Kissinger, J.C., Bromley, E., Lal, K., Sinden, R.E., Tomley, F., Wastling, J.M., 2008. Determining the protein repertoire of *Cryptosporidium parvum* sporozoites. Proteomics 8, 1398-1414.

Schmid, M.W., Grossniklaus, U., 2015. Rcount: simple and flexible RNA-Seq read counting. Bioinformatics 31, 436-437.

Schnyder, M., Kohler, L., Hemphill, A., Deplazes, P., 2009. Prophylactic and therapeutic efficacy of nitazoxanide against *Cryptosporidium parvum* in experimentally challenged neonatal calves. Vet Parasitol 160, 149-154.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proc Nat Acad Sci 95, 5857-5864.

Siddiki, A.Z., 2013. Sporozoite proteome analysis of *Cryptosporidium parvum* by one-dimensional SDS-PAGE and liquid chromatography tandem mass spectrometry. J Vet Sci 14, 107-114.

Sinden, R.E., Talman, A., Marques, S.R., Wass, M.N., Sternberg, M.J., 2010. The flagellum in malarial parasites. Curr Opin Microbiol 13, 491-500.

Smith, H.V., Nichols, R.A., Grimason, A.M., 2005. *Cryptosporidium* excystation and invasion: getting to the guts of the matter. Trends Parasitol 21, 133-142.

Spano, F., Puri, C., Ranucci, L., Putignani, L., Crisanti, A., 1997. Cloning of the entire COWP gene of *Cryptosporidium parvum* and ultrastructural localization of the protein during sexual parasite development. Parasitology 114 ( Pt 5), 427-437.

Striepen, B., 2013. Parasitic infections: Time to tackle cryptosporidiosis. Nature 503, 189-191.

Upton, S.J., Tilley, M., Brillhart, D.B., 1994a. Comparative development of *Cryptosporidium parvum* (Apicomplexa) in 11 continuous host cell lines. FEMS Microbiol Lett 118, 233-236.

Upton, S.J., Tilley, M., Brillhart, D.B., 1995. Effects of select medium supplements on in vitro development of *Cryptosporidium parvum* in HCT-8 cells. J Clin Microbiol 33, 371-375.

Upton, S.J., Tilley, M., Nesterenko, M.V., Brillhart, D.B., 1994b. A simple and reliable method of producing in vitro infections of *Cryptosporidium parvum* (Apicomplexa). FEMS Microbiol Lett 118, 45-49.

Vetterling, J.M., Takeuchi, A., Madden, P.A., 1971. Ultrastructure of *Cryptosporidium wrairi* from the guinea pig. J Protozool 18, 248-260.

Vinayak, S., Pawlowic, M.C., Sateriale, A., Brooks, C.F., Studstill, C.J., Bar-Peled, Y., Cipriano, M.J., Striepen, B., 2015. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. Nature 523, 477-480.

Walker, R., Gissot, M., Croken, M.M., Huot, L., Hot, D., Kim, K., Tomavo, S., 2013a. The *Toxoplasma* nuclear factor TgAP2XI-4 controls bradyzoite gene expression and cyst formation. Mol Microbiol 87, 641-655.

Walker, R., Gissot, M., Huot, L., Alayi, T.D., Hot, D., Marot, G., Schaeffer-Reiss, C., Van Dorsselaer, A., Kim, K., Tomavo, S., 2013b. *Toxoplasma* transcription factor TgAP2XI-5 regulates the expression of genes involved in parasite virulence and host invasion. J Biol Chem 288, 31127-31138.

Walker, R.A., Sharman, P.A., Miller, C.M., Lippuner, C., Okoniewski, M., Eichenberger, R.M., Ramakrishnan, C., Brossier, F., Deplazes, P., Hehl, A.B., Smith, N.C., 2015. RNA Seq analysis of the *Eimeria tenella* gametocyte transcriptome reveals clues about the molecular basis for sexual reproduction and oocyst biogenesis. BMC Genomics 16, 94.

Ward, P.I., Deplazes, P., Regli, W., Rinder, H., Mathis, A., 2002. Detection of eight *Cryptosporidium* genotypes in surface and waste waters in Europe. Parasitology 124, 359-368.

Xu, T., Ping, J., Yu, Y., Yu, F., Yu, Y., Hao, P., Li, X., 2010. Revealing parasite influence in metabolic pathways in Apicomplexa infected patients. BMC Bioinformatics 11, S13.

Yuda, M., Iwanaga, S., Kaneko, I., Kato, T., 2015. Global transcriptional repression: An initial and essential step for *Plasmodium* sexual development. Proc Natl Acad Sci U S A 112, 12824-12829.

Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T., Kaneko, I., 2010. Transcription factor AP2-Sp and its target genes in malarial sporozoites. Mol Microbiol 75, 854-863.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G.R., Janse, C.J., Waters, A.P., Kato, T., Kaneko, I., 2009. Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. Mol Microbiol 71, 1402-1414.

Zhang, H., Guo, F., Zhou, H., Zhu, G., 2012. Transcriptome analysis reveals unique metabolic features in the *Cryptosporidium parvum* oocysts associated with environmental survival and stresses. BMC Genomics 13, 647.

**Figure legends**

Fig. 1. Expression analysis of *Cryptosporidium parvum* genes in different life cycle stages. (A) Sample correlation and pairwise comparison based on DESeq2 normalised (Love et al., 2014) gene expression values of all genes with at least five reads in at least one of the samples. Samples were clustered using Pearson correlation and hierarchical agglomerative clustering (complete linkage). (B) Same as in (A), but excluding sporozoite data. (C) Expression values of genes differentially expressed between sporozoites and the populations grown in vivo and in vitro. Out of 3775 genes tested, 1612 genes were found to be differentially expressed (false discovery rate (FDR) < 0.01, absolute log2 fold change > 2). (D) Expression values of genes differentially expressed between populations grown in vivo and populations grown in vitro (517 genes with FDR < 0.01, absolute log2 fold change > 1). (C, D) Expression values correspond to genewise Z-scores of scaled, log2-transformed and DESeq-normalised data. Samples/genes were clustered using euclidean distance and hierarchical agglomerative clustering (complete linkage). (E) Number of up- and downregulated genes between sporozoites and in vivo/in vitro stages as well as between in vivo and in vitro stages. For this comparison, a FDR of < 0.01 and a minimal regulation of 2-fold ($log_2FC$ > 1) was used. sporo, sporozoites; dpi, days p.i.

Fig. 2. Expression analysis of *Cryptosporidium parvum* COWP genes. Comparative depiction of nine COWP mRNA levels (linear scale, normalised reads/kb); inset: expression levels of the six low abundance COWP genes (scale: log2 (normalised reads/kb)). dpi, days p.i.

Fig. 3. Expression profiles of the *Cryptosporidium parvum* ApiAP2 transcription factors. Comparative depiction of ApiAP2 mRNA levels (linear scale, normalised reads/kb). (A) Comparison of all detected ApiAP2 genes. (B) Comparison of ApiAp2 genes with expression levels below 500 normalised reads/kb. dpi, days p.i.

Fig. 4. Expression profiles of the *Cryptosporidium parvum* mucin and mucin-like gene (Smith et al., 2005) mRNA levels (normalised reads/kb). dpi, days p.i.

**Supplementary figure legends**

Supplementary Fig. S1. Quality control of RNA using an Agilent Bioanalyzer and Picochip. Fluorescence intensity (FU) is plotted against migration time (s) or size (nucleotides (nt)). Parasite-specific 26S rRNA is indicated by arrows. (A) In vivo RNA samples. Sample 1 of the 4 days p.i. sample is a pool of three different samples (fractions). (B) In vitro RNA samples. (C) Sporozoite RNA samples. dpi, days p.i.

Supplementary Fig. S2. Microscopy images of *Cryptosporidium parvum*. (A) H & E staining of intestinal section from an infected calf. Parasites are visible at the epithelium (arrowheads). Scale bar = 50 µm. (B) Differential interference contrast image of excysted sporozoites (arrowheads).

Supplementary Fig. S3. Pairwise comparison of normalised and log2(x+1)-transformed gene expression values of *Cryptosporidium parvum*. The data was normalised with DESeq2 (with default parameters (Love et al., 2014) using either all samples ("including sporozoites") or only the in vivo and in vitro samples ("without sporozoites"). Colours indicate the point densities from high (red) to low (blue). Correlation plots are given including or excluding the sporozoite samples. CorP, Pearson correlation; CorS, Spearman correlation; n, number of genes plotted. (A) Comparisons of sporozoite data with in vivo or in vitro data. (B) Comparisons of in vivo 2 days p.i. data with in vivo 4 days p.i. and in vitro data (analysis includes sporozoites). (C) Comparisons of in vivo 4 days p.i. data with in vitro data (analysis includes sporozoites). (D) Comparisons of in vitro 2 days p.i. data with in vitro 4 days p.i. data (analysis includes sporozoites). (E) Comparisons of in vivo 2 days p.i. data with in vivo 4 days p.i. and in vitro data (analysis excludes sporozoites). (F) Comparisons of in vivo 4 days p.i. data with in vitro data (analysis excludes sporozoites). (G) Comparisons of in vitro 2 days p.i. data with in vitro 4 days p.i. data (analysis excludes sporozoites). dpi, days p.i.

**Reference**

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

Supplementary Fig. S4. Differential expression of *Cryptosporidium parvum* microgamete-associated genes. (A) Protein sequence alignment of the H/G domain in Hap2 homologues of *Chlamydomonas reinhardtii, Plasmodium berghei, Toxoplasma gondii*, and *C. parvum*. Bar graph showing expression levels of cgd8_2220, the putative *C. parvum hap2* (normalised reads). (B-F)

Bar graphs showing expression levels of putative microgamete-associated genes (normalised reads or normalised reads/kb). dpi, days p.i.

Supplementary Fig. S5. Differential expression of *Cryptosporidium parvum* meiosis-associated genes. (A) Expression values in normalised reads/kb. (B) Life cycle of *Cryptosporidium* Sexual development takes place after schizogony. Gametocytes release gametes that fuse to a zygote in which meiosis takes place. Further development of oocysts containing four sporozoites renders the parasite infectious again. Reprinted from Trends in Parasitology, 21(3), Smith, H.V., Nichols, R.A.B., Grimason, A.M., *Cryptosporidium* excystation and invasion: getting to the guts of the matter, pp. 133–142, (2005), with permission from Elsevier. dpi, days p.i.

Supplementary Fig. S6. Differential expression of calcium-dependent protein kinases in *Cryptosporidium parvum*. (A) Expression values in normalised reads/kb. (B) Alignment of serine/threonine protein kinase domains of cgd7_1260 with the domains of CDPK6 from *Plasmodium falciparum* and *Toxoplasma gondii* as well as of CDPK21 of *Arabidopsis thaliana*. Catalytic residues are highlighted in yellow. In cgd7_1260, a critical aspartate is missing (highlighted in red). dpi, days p.i.

**Highlights:**

- RNA-Seq was used to compare gene expression of sporozoites versus intracellular *Cryptosporidium parvum.*
- RNA-Seq revealed *C. parvum* genes that are expressed in a stage-specific manner and others needed at all stages of development.
- 173 genes (26 for predicted secreted proteins) were upregulated in sporozoites.
- 1259 genes were upregulated in intracellular stages, across 63 biological processes and 23 metabolic pathways.
- Our results allow identification of viable drug and vaccine targets against *C. parvum*.

Figure 1

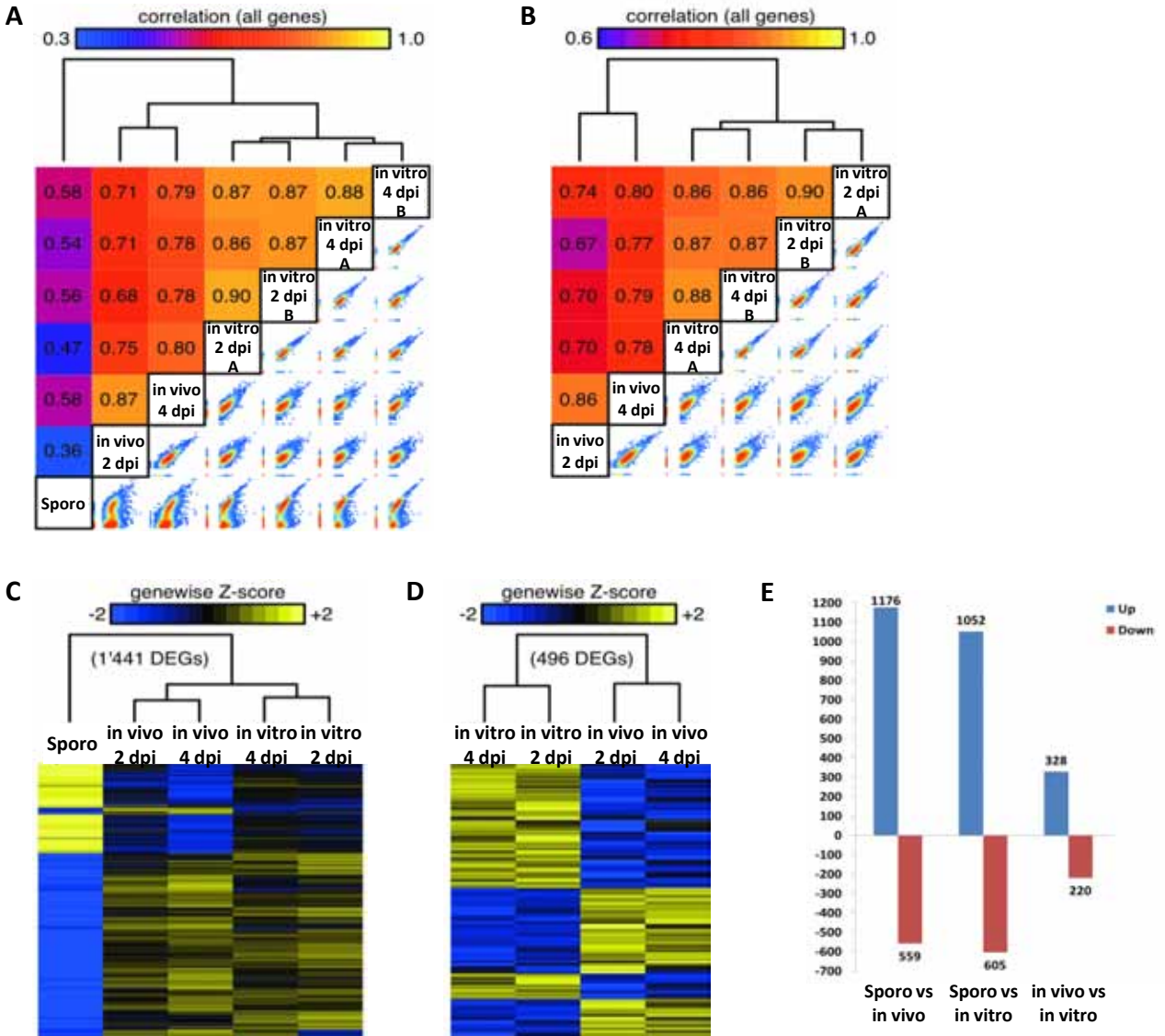

A
correlation (all genes)
0.3
1.0
0.58 0.71 0.79 0.87 0.87 0.88
in vitro 4 dpi B
0.54 0.71 0.78 0.86 0.87
in vitro 4 dpi A
0.56 0.68 0.78 0.90
in vitro 2 dpi B
0.47 0.75 0.80
in vitro 2 dpi A
0.58 0.87
in vivo 4 dpi
0.36
in vivo 2 dpi
Sporo
B
correlation (all genes)
0.6
1.0
0.74 0.80 0.86 0.86 0.90
in vitro 2 dpi A
0.67 0.77 0.87 0.87
in vitro 2 dpi B
0.70 0.79 0.88
in vitro 4 dpi B
0.70 0.78
in vitro 4 dpi A
0.86
in vivo 4 dpi
in vivo 2 dpi
C
genewise Z-score
-2
+2
(1'441 DEGs)
Sporo
in vivo 2 dpi
in vivo 4 dpi
in vitro 4 dpi
in vitro 2 dpi
D
genewise Z-score
-2
+2
(496 DEGs)
in vitro 4 dpi
in vitro 2 dpi
in vivo 2 dpi
in vivo 4 dpi
E
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
1176
559
1052
605
328
220
Up
Down
Sporo vs in vivo
Sporo vs in vitro
in vivo vs in vitro

**Figure 2**

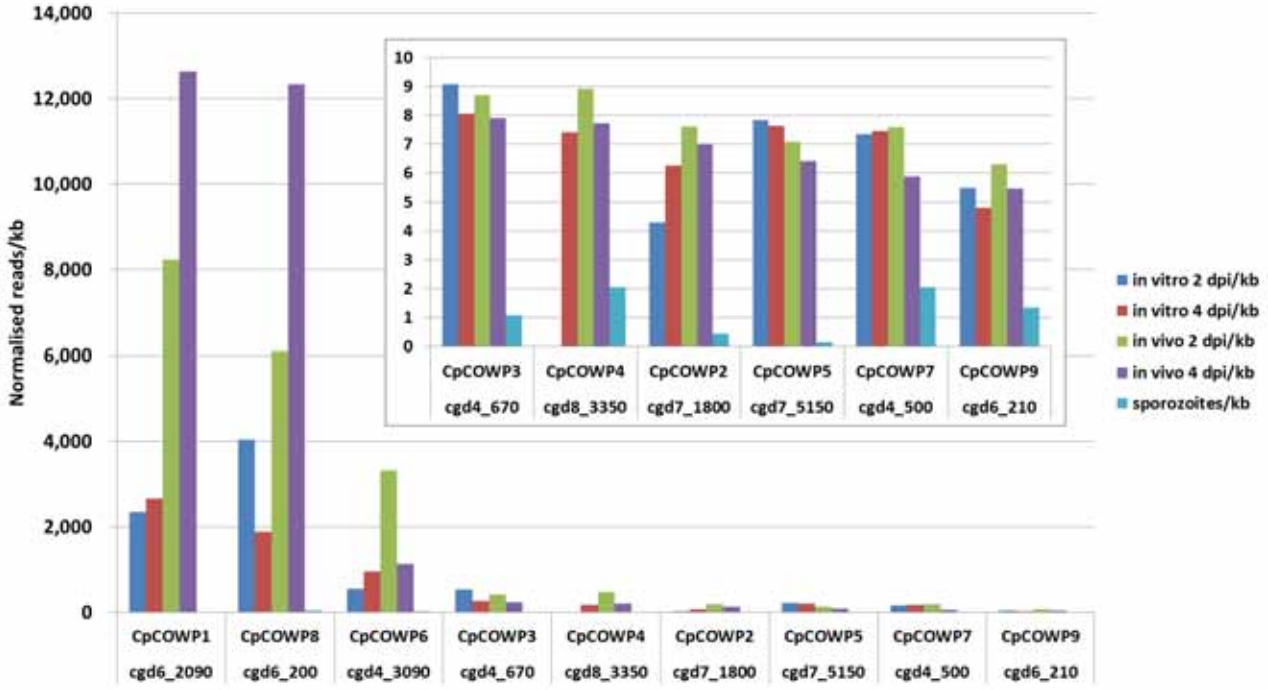
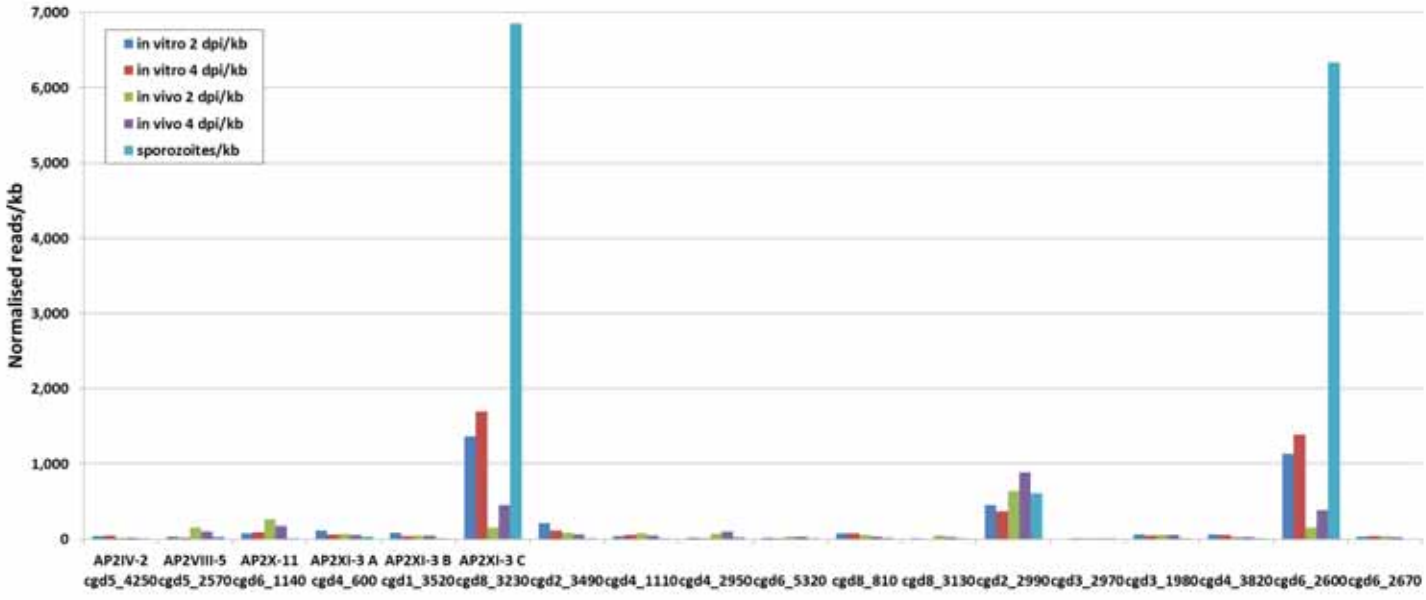

Normalised reads/kb
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
CpCOWP1 cgd6_2090
CpCOWP8 cgd6_200
CpCOWP6 cgd4_3090
CpCOWP3 cgd4_670
CpCOWP4 cgd8_3350
CpCOWP2 cgd7_1800
CpCOWP5 cgd7_5150
CpCOWP7 cgd4_500
CpCOWP9 cgd6_210
10
9
8
7
6
5
4
3
2
1
0
CpCOWP3 cgd4_670
CpCOWP4 cgd8_3350
CpCOWP2 cgd7_1800
CpCOWP5 cgd7_5150
CpCOWP7 cgd4_500
CpCOWP9 cgd6_210
in vitro 2 dpi/kb
in vitro 4 dpi/kb
in vivo 2 dpi/kb
in vivo 4 dpi/kb
sporozoites/kb

Figure 3

A


in vitro 2 dpi/kb
in vitro 4 dpi/kb
in vivo 2 dpi/kb
in vivo 4 dpi/kb
sporozoites/kb
Normalised reads/kb
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
AP2IV-2 cgd5_4250
AP2VIII-5 cgd5_2570
AP2X-11 cgd6_1140
AP2XI-3 A cgd4_600
AP2XI-3 B cgd1_3520
AP2XI-3 C cgd8_3230
cgd2_3490
cgd4_1110
cgd4_2950
cgd6_5320
cgd8_810
cgd8_3130
cgd2_2990
cgd3_2970
cgd3_1980
cgd4_3820
cgd6_2600
cgd6_2670

B

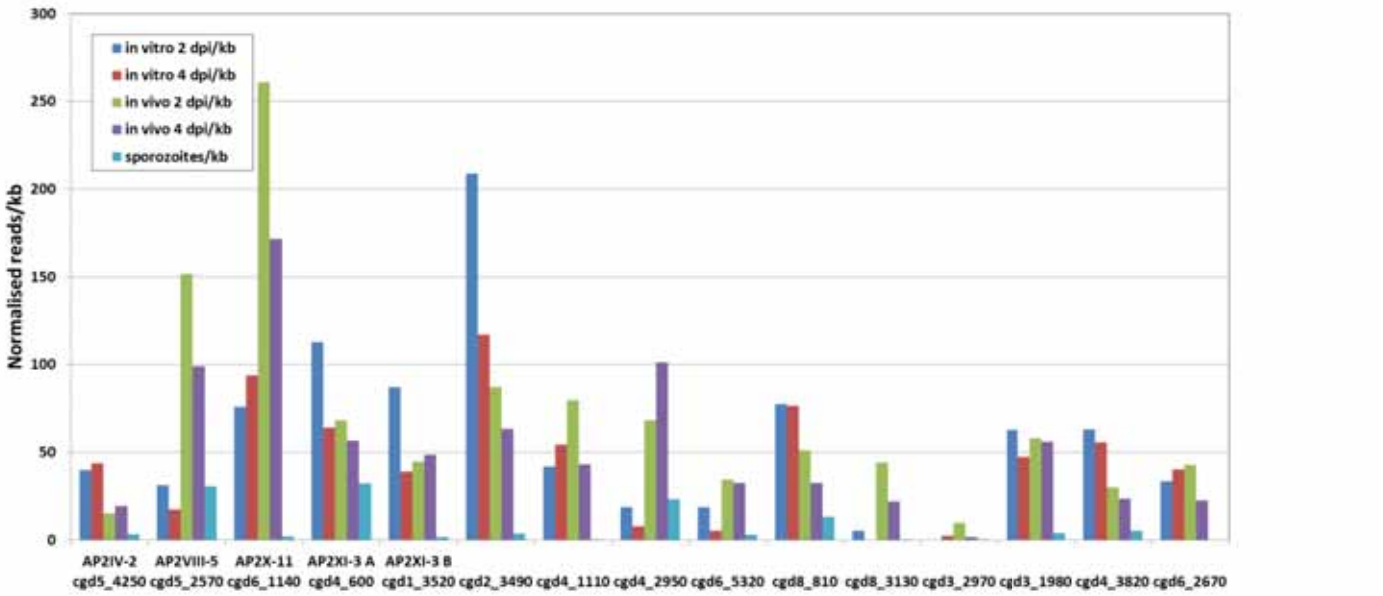
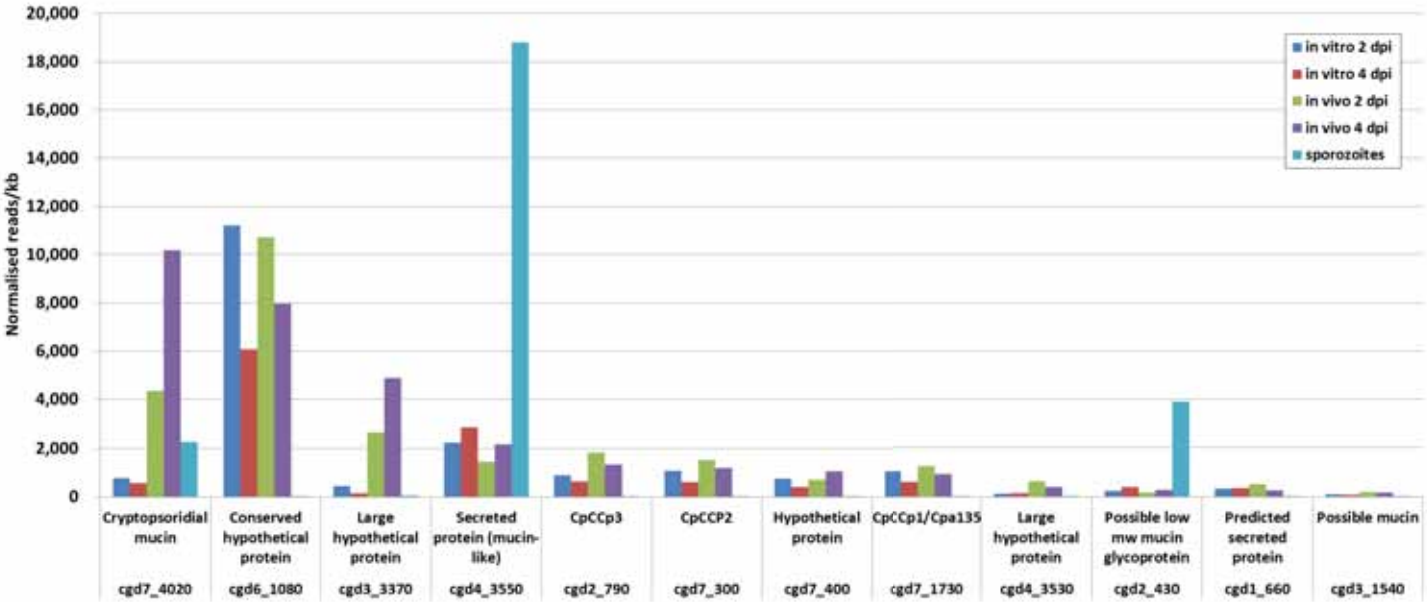

in vitro 2 dpi/kb
in vitro 4 dpi/kb
in vivo 2 dpi/kb
in vivo 4 dpi/kb
sporozoites/kb
Normalised reads/kb
300
250
200
150
100
50
0
AP2IV-2 cgd5_4250
AP2VIII-5 cgd5_2570
AP2X-11 cgd6_1140
AP2XI-3 A cgd4_600
AP2XI-3 B cgd1_3520
cgd2_3490
cgd4_1110
cgd4_2950
cgd6_5320
cgd8_810
cgd8_3130
cgd3_2970
cgd3_1980
cgd4_3820
cgd6_2670

**Figure 4**


Normalised reads/kb
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
in vitro 2 dpi
in vitro 4 dpi
in vivo 2 dpi
in vivo 4 dpi
sporozoites
Cryptopsoridial mucin
cgd7_4020
Conserved hypothetical protein
cgd6_1080
Large hypothetical protein
cgd3_3370
Secreted protein (mucin-like)
cgd4_3550
CpCCp3
cgd2_790
CpCCP2
cgd7_300
Hypothetical protein
cgd7_400
CpCCp1/Cpa135
cgd7_1730
Large hypothetical protein
cgd4_3530
Possible low mw mucin glycoprotein
cgd2_430
Predicted secreted protein
cgd1_660
Possible mucin
cgd3_1540

*Graphical Abstract (for review