

# ***Cryptosporidium* uses multiple distinct secretory organelles to interact with and modify its host cell**

## **Authors**

Amandine Guérin<sup>1</sup>, Katherine M. Strelau<sup>1\*</sup>, Konstantin Barylyuk<sup>2\*</sup>, Bethan A. Wallbank<sup>1</sup>, Laurence Berry<sup>3</sup>, Oliver M. Crook<sup>4</sup>, Kathryn S. Lilley<sup>2</sup>, Ross F. Waller<sup>2</sup>, and Boris Striepen<sup>1,†</sup>

## **Affiliations**

<sup>1</sup>Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>2</sup>Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK

<sup>3</sup>LPHI, CNRS, Université de Montpellier, Montpellier 34095, France

<sup>4</sup> Department of Statistics, University of Oxford, Oxford OX1 3LB, UK

\*These authors contributed equally.

† Lead contact: Boris Striepen Tel.: 1-215-573-9167; Fax: 1-215-746-2295; e-mail: striepen@upenn.edu

**SHORT TITLE:** *Cryptosporidium* secreted effectors

## **SUMMARY**

*Cryptosporidium* is a leading cause of diarrheal disease in children and an important contributor to early childhood mortality. The parasite invades and extensively remodels intestinal epithelial cells, building an elaborate interface structure. How this occurs at the molecular level and the contributing parasite factors are largely unknown. Here, we generated a whole-cell spatial proteome of the *Cryptosporidium* sporozoite, and used genetic and cell biological experimentation to discover the *Cryptosporidium* secreted effector proteome. These findings reveal multiple organelles, including an original secretory organelle, and generate numerous compartment markers by tagging native gene loci. We show that secreted proteins are delivered to the parasite-host interface where they assemble into different structures including a ring that

anchors the parasite into its unique epicellular niche. *Cryptosporidium* thus uses a complex set of secretion systems during and following invasion that act in concert to subjugate its host cell.

**KEYWORDS:** *Cryptosporidium*, Apicomplexa, microneme, rhoptry, dense granules, small granules, secretion, effectors, spatial proteomics

## **INTRODUCTION**

Diarrheal disease resulting from infections with enteric pathogens is a leading cause of death in children. Among them, the apicomplexan parasite *Cryptosporidium* has been found to be highly prevalent and associated with severe disease in children under the age of 2<sup>1</sup>. Malnourished children and those with compromised immunity, suffer particularly severe disease while the infection is self-limited in immunocompetent adults. Vaccines and fully efficient drugs to prevent or treat the disease are lacking<sup>2-4</sup>. Infection occurs by ingestion of oocysts, spore-like stages that contain four invasive sporozoites<sup>5</sup>. The life cycle unfolds within the epithelium of the small intestine and includes meronts, which replicate asexually, as well as male and female gametes<sup>6-8</sup>. Parasite sex, followed by meiosis, leads to a new generation of oocysts. Parasite replication occurs within enterocytes and is restricted to a vacuole at the apical face of the infected cell<sup>5</sup>. Electron microscopy of infected intestines revealed significant host cell remodeling induced by the parasite, including shortening of the microvilli, destabilization of the epithelial junctions, and establishment of a highly complex structure at the host-parasite interface<sup>9,10</sup>. This includes a pedestal of host actin, two highly ordered bands of proteins of unknown provenance, and a ring-like structure resembling a tight junction, that are hypothesized to work together to confine the parasite to the apex of the host cell. The anabolic capabilities of *Cryptosporidium* are dramatically reduced<sup>11,12</sup>, and the parasite thus relies heavily on the import of metabolites from the host cell<sup>13</sup>, which likely occurs through the feeder organelle, an elaborate membranous component of the host-parasite interface<sup>5</sup>. The molecular composition of the feeder organelle, and the mechanisms by which it is assembled are unknown.

Work on bacterial and protist pathogens has identified numerous secreted and/or injected effector proteins. Their functions include facilitating pathogen attachment and invasion,

establishment of the vacuole in which replication will occur, immune evasion through transcriptional repression or inactivation of restriction factors through enzymatic modification or direct binding<sup>14-17</sup>. Interestingly, apicomplexan parasites have compartmentalized their secretory proteins into three distinct organelles that allow discharge at different times and into different cellular environments. Microneme proteins are secreted onto the surface of the extracellular parasite to serve as adhesins for attachment and motility<sup>18</sup>. Rhoptries secrete directly into the host cytoplasm during invasion, providing a function analogous to the bacterial type-three secretion system and play essential roles in invasion and immune evasion<sup>14</sup>. Finally, dense granules are secreted inside the forming parasitophorous vacuole (PV) during and likely following invasion<sup>19</sup>. It has been speculated that there may be different subpopulations of dense granules<sup>20</sup>, however, this remains to be formally demonstrated. Once secreted, some dense granule proteins remain in the PV where they play important roles in molecular transport, while others are translocated further to act in immune evasion<sup>16</sup>. The invasive forms of apicomplexan parasites are thus prepacked with multiple organelles containing effector proteins. However, between different apicomplexan species very few of these effectors are conserved, likely reflecting adaptation to their different hosts.

In recent studies, we used transcriptional patterns and polymorphism to identify *Cryptosporidium* proteins injected during invasion<sup>6</sup> or translocated from the growing intracellular parasite<sup>21</sup>. Here, we define a much more comprehensive set of *Cryptosporidium* effectors applying hyperplexed localisation of organelle proteins by isotope tagging (hyperLOPIT)<sup>22</sup> to the infectious sporozoite stage. We identified 154 secretory proteins and assigned them to four parasite organelles, dramatically extending our knowledge of the content of the micronemes and rhoptries, and found 79 dense granule proteins. Through our analyses we also discovered a fourth secretory compartment, not observed in other apicomplexans, that is distinct in content, timing of biogenesis, and delivery. We reporter-tagged numerous proteins and used high resolution expansion microscopy to establish their localization prior to and after delivery to the host cell and the host parasite interface. Overall, this study allowed predicting the localization of more than a thousand proteins in 14 subcellular compartments and provides a critical road map

to dissect *Cryptosporidium* host-parasite interaction and the role that parasite effectors play in shaping it.

## **RESULTS**

### **Comprehensive spatial proteome mapping of *Cryptosporidium* sporozoites reveals their subcellular compartments**

To define the content of *Cryptosporidium* secretory organelles and identify effector proteins, we conducted a comprehensive survey of the spatial proteome by hyperLOPIT. This technique exploits similarities of protein abundance distribution profiles upon detergent-free lysis and fractionation of cells to identify the protein cohorts that represent subcellular compartments (figure 1A). We chose sporozoites, as they are the infectious stage and obtainable in the numbers and purity required for hyperLOPIT, and conducted two independent experiments (see STAR Methods and table S1 for details). Purified oocysts were triggered to release sporozoites by treatment with bleach and taurodeoxycholate<sup>23</sup> and then subjected to three sequential rounds of nitrogen cavitation to break the cells while aiming to maintain the integrity of subcellular membranes. The homogenates were separated by density gradient ultracentrifugation into a supernatant enriched in cytosolic and soluble proteins, and a membrane fraction at the interface between the gradient layers. The later was further resolved into nine fractions by equilibrium-density centrifugation in a linear density gradient. We applied a quantitative shotgun proteomics approach labeling the peptides derived from each of the ten fractions with a distinct isobaric TMT10plex tag. The TMT10plex-tagged peptides were then pooled and fractionated by two-dimensional liquid chromatography followed by quantitative MS/MS analysis.

In two experiments, we identified 1810 and 2321 parasite proteins at a protein FDR level of 5% and measured their relative abundance distribution across ten subcellular fractions. Overall, 2389 proteins were found in at least one dataset, and 1742 proteins were shared (figure 1B) and we concatenated the relative quantification data for those shared proteins to increase the power of the analysis<sup>24</sup>. To visualize the resulting 20-dimensional dataset, we used dimensionality reduction by t-distributed stochastic neighbor embedding<sup>25</sup> (t-SNE, figure 1C). The

t-SNE plot revealed non-random distribution suggesting protein clustering based on fractionation profile and, hence, subcellular colocation. Using unsupervised clustering of the untransformed abundance distribution profiles by hierarchical density-based spatial clustering of applications with noise (HDBSCAN), we discovered twelve protein clusters (color in figure 1C). To identify secretory organelles, we explored the distribution of proteins predicted by SignalP<sup>26</sup> to have an N-terminal signal peptide, and mapped those onto the t-SNE plot (figure 1D). We further annotated the data through compiling a manually curated list of 293 marker proteins experimentally shown to reside in specific compartments in *Cryptosporidium*, as well as proteins with high-confidence assignments as orthologs of proteins with confirmed and conserved localization across Apicomplexa (figure 1E and table S2). Based on the co-enrichment of predicted secretory proteins and prioritized markers, we could identify clusters for the micronemes, rhoptries, the crystalloid body, and the oocyst wall. Conspicuously, two clusters of secreted proteins lacked known markers (figure 1E dashed circles with no color). Overall, the combination of unsupervised machine-learning analysis, protein sequence features, and manual annotation allowed us to recognize multiple *Cryptosporidium* subcellular niches. Our curated marker list was not exhaustive, and some protein clusters were populated with proteins of unknown function.

Next, we applied a novelty-detecting Bayesian classification method based on T-augmented Gaussian mixture models (Novelty-TAGM)<sup>27</sup>. This semi-supervised algorithm uses a training reference list to identify additional proteins within a cluster of established subcellular identity and allows for the *de novo* identification of clusters that lack established markers. This identified 11 clusters in addition to the 11 established using known markers (figure 1F and table S3). GoTerm enrichment analysis showed that six of the uncharacterized clusters correspond to a combination of cytosolic and/or nuclear proteins, but no clear border could be drawn between the two compartments (purple gradient in figure 1F and table S4). In addition, we found strongly supported clusters for the ribosome (40 and 60S), 20S proteasome, inner membrane complex (IMC) and plasma membrane, endoplasmic reticulum (ER), apical cap, and the highly reduced *Cryptosporidium* mitochondrion, the mitosome. Importantly, the two original signal peptide-enriched clusters were verified by Novelty-TAGM as distinct and yet uncharacterized secretory niches in *Cryptosporidium*.

## Crystalloid body and oocyst wall proteins are specifically expressed in the female gamete

One of the best resolved clusters enriched in secretory proteins contained the *Cryptosporidium* Oocyst Wall Protein 1 (COWP1), a known component of the wall that encases and protects sporozoites during transmission<sup>8,28</sup>. Computational studies had predicted 8 additional members of the COWP family<sup>29</sup>, and all of them co-occurred in this cluster (table S3). A further 35 proteins were assigned to this cluster by Novelty-TAGM (figure 2A). They included multiple apple, EGF, Sushi, and chitin binding domain proteins that likely share structural functions with COWPs and participate in assembling the multi-layered oocyst shell<sup>30</sup>. Numerous enzymes were also found in this oocyst wall cluster (table S3). Three of them had a glucose-methanol-choline oxidoreductase fold, one was a glycosyltransferase, another a polysaccharide pyruvyl transferase, and many were annotated as peptidases and hydrolases. The transcripts for the proteins that make up this cluster are highly expressed in female gametes but generally absent in asexual stages (figure 2B). This is consistent with the observation that oocyst biogenesis is exclusive to female parasites, which assemble precursors in wall forming bodies<sup>8,28</sup>.

We identified a second female exclusive secretory cluster (figure 2B). This cluster contained homologs of *Plasmodium* crystalloid body proteins, which are expressed in the mosquito phase of the lifecycle (ookinetes and young oocysts), several being necessary for insect to mammal transmission<sup>31,32</sup> (figure 2A and table S3). In *Cryptosporidium*, the crystalloid body is a honey-combed multivesicular structure localized to the basal end of the sporozoite<sup>33</sup>, but no proteins had been previously identified from it. To experimentally test our assignment, we tagged the endogenous locus of one protein, the Pleckstrin Homology (PH) domain protein cgd8\_4310, a homolog of *Plasmodium berghei* protein PBANKA\_0704800 localized in the crystalloid body of ookinetes<sup>34</sup>. We verified insertion into the locus (figure S1) and assessed the localization of the protein in extracellular sporozoites. HA-labeling showed accumulation at the end of the parasite, basal to the nucleus, consistent with the crystalloid body form and localization (figure 2C). Among the proteins found in this cluster, CpCCP1, also named Cpa135 (cgd7\_1730), had previously been suggested to be a microneme protein based on its polar labeling using an antibody<sup>35</sup>. To resolve this discrepancy, we reporter-tagged the CpCCP1 locus. We saw polar labeling but, using the

nucleus as a reference point, we determined this label to be basal and not apical, and therefore consistent with the crystalloid body (figure 2D).

To achieve higher resolution determination of protein locations, we applied ultrastructure expansion microscopy (U-ExM) in combination with the N-hydroxysuccinimide ester dye (NHS-ester). The NHS-ester enabled visualization of the conoid (c), the single rhoptry (r), abundant tightly packed granules (g) apical to the nucleus and showed the overall outline of the parasite (figure 2E). We applied U-ExM to the two crystalloid body tagged lines. CpCCP1 labeling was lost during the procedure (this protein putatively binds to lipids and membrane associated proteins are poorly preserved in this approach, figure S2A). However, the PH domain-containing protein highlighted the body basal to the nucleus with a granular staining pattern consistent with the multivesicular nature of this structure (figure 2F). We, therefore, named this protein crystalloid body protein 1 (CB1). It is unclear whether the function of crystalloid body proteins involves their secretion, however, we found no evidence for discharge during or after invasion (figure 2G).

#### **Identification of an expanded set of *Cryptosporidium* microneme and rhoptry proteins**

Two additional clusters enriched for secretory proteins included markers for the microneme and rhoptry (figure 1F). We investigated whether the uncharacterized proteins assigned in these clusters were indeed microneme or rhoptry proteins, respectively. We HA-tagged the C-terminus of two candidates and investigated their localization in the sporozoite (figure S1). Labeling for the putative microneme protein cgd7\_1960 showed accumulation at the apical end of extracellular sporozoites and colocalized with the micronemal marker protein gp40/gp900<sup>36</sup> (figure 3A and S2B). We also tagged the rhoptry-assigned protein cgd5\_20 and found it to be localized adjacent and basal to the rhoptry neck protein CpPRP1<sup>37</sup> (figure 3B). This localization closely resembles our previous observations for ROP1 to 6<sup>6</sup>, suggesting a localization in the bulb of the rhoptry. Using U-ExM, we visualized cgd7\_1960-HA associated with numerous small uniform vesicles that filled the apical half of the sporozoites, a pattern consistent with micronemes observed by electron microscopy<sup>33</sup> (figure 3C). The cgd5\_20-HA staining overlapped with the single rhoptry demarcated by NHS-ester staining of sporozoites but was restricted to

the bulbous part (figure 3D). Following the established nomenclature<sup>6,38</sup>, the proteins were named CpMIC2 and ROP7, but we note that they are not orthologous to *T. gondii* MIC2 and ROP7.

We also examined the patterns of transcription for these two clusters. We noted transcript enrichment in asexual meronts and fertilized female gametes *in vivo* for both (figure S2C). These are the two stages that produce invasive merozoites and sporozoites, respectively, and therefore require micronemes and rhoptries. Note that *in vitro* female gametes are not fertilized and thus do not form sporozoites. Consistent with this, the two clusters lack expression in this stage. Transcription peaks at the end of the asexual replication cycle (12h) which is also consistent with the timing of biogenesis of invasion related organelles (figure 3E)<sup>6,39</sup>. The microneme cluster contains the five previously known micronemal proteins, cp23, gp60 (gp40/15), gp900, MIC1 (TSP8) and p30, and a homolog of *Toxoplasma* Toxolysin 4 (cgd2\_4270). Among the remaining proteins, 3 have an uncharacterized homolog in *Toxoplasma* (TGME49\_216820, TGME49\_265790, and TGME49\_205680), all of which are assigned as micronemal proteins by *Toxoplasma* hyperLOPIT<sup>40</sup> while 15 are unique to *Cryptosporidium* (figure 3F and table S3). The rhoptry cluster contained all 7 previously identified proteins (ROP1 to 6 and cpPRP1), 8 proteins with homologs in *Toxoplasma* including TgARO, TgDHHC7, TgRON6, TgRON9, TgRON10 and TgRON11, and 2 uncharacterized putative rhoptry proteins (TGME49\_279420 and TGME49\_254070)<sup>40</sup> (figure 3F and table S3). Nine proteins are unique to *Cryptosporidium*, among them an ankyrin domain protein, one with a proline rich extension, and one with a formin homology domain two. To assess whether the tagged proteins are secreted, parasites were observed during and following invasion into host cells. Immunofluorescence assays showed that CpMIC2 relocated to the surface of the parasite prior to invasion (figure 3G and S2D), while ROP7 accumulated in the parasitophorous vacuole space following invasion (figure 3H).

## **79 proteins make up the proteome of the dense granule compartment**

The 'cluster four' identified by Novelty-TAGM was well resolved by HDBSCAN and is highly enriched in proteins containing a signal peptide. However, none of the proteins it contained had been studied before or had putative functions assigned. We reporter-tagged four proteins of this cluster, cgd3\_600, cgd5\_1440, cgd1\_590 and cgd3\_1690 to investigate their cellular localization



(figure S1). All shared a similar localization in the sporozoite: apical of the nucleus, but basal to micronemes and rhoptry (figure 4A-B and S3A-B). U-ExM revealed their presence associated with heavily NHS-ester-stained granules, notably the more apical population of these granules (figure 4C-D, S3C-D and S4A). We noted subtle differences in sub-compartmentalization of our candidates: cgd3\_600, cgd3\_1690 and cgd5\_1440 were located within the NHS-ester-stained granule, while cgd1\_590 appeared more peripheral (figure 4E), however, further experimentation is needed to understand the extent and relevance of these differences. Overall, the labeling suggested a localization to organelles described as dense granules in *Cryptosporidium* and *Toxoplasma* by electron microscopy<sup>33,41</sup>. We performed immuno-electron microscopy using the cgd5\_1440-HA-tagged line and Figure 4F shows accumulation of gold-beads in single membrane-bound organelles that match the location, shape, and size of canonical dense granules (mean =  $202 \pm 22.7\text{nm}$  (figure S3E))<sup>19</sup>. We thus named them DG1 (cgd3\_600), DG2 (cgd5\_1440), DG3 (cgd1\_590) and DG4 (cgd3\_1690) for dense granule proteins 1 to 4.

Inspection of the 79 dense granule proteins identified only three with orthologs in other apicomplexans. Cgd8\_2160 has an uncharacterized ortholog in *Toxoplasma* (TGME49\_280380) predicted to be a dense granule protein by *Toxoplasma* hyperLOPIT<sup>40</sup>, and both share a poly (ADP-ribose) glycohydrolase domain. The second is the Serine/Threonine phosphatase cgd7\_4420 with an uncharacterized ortholog in *Toxoplasma* (TGME49\_297650), there predicted to reside in the Golgi. The third, cgd7\_2620, is an ortholog of *Plasmodium* ClpB1 and ClpB2/HSP101 (table S3). ClpB1 is an apicoplast protein while ClpB2/HSP101 is secreted into the PV space and involved in protein translocation across the PV membrane<sup>42,43</sup>. We also observed members of paralogous families of secreted proteins unique to *Cryptosporidium* including the CpLSP and HCD family, as well as proteins characterized by the presence of kinase-like, kinesin-like, and cysteine-rich domains (table S3).

### ***Cryptosporidium* dense granules are secreted during invasion to multiple locations**

The identification of the protein content of *Cryptosporidium* dense granules provided the opportunity to define the compartment these secreted effectors target upon dense granule discharge. We found that during the early stages of invasion, characterized by a C-shaped

sporozoite morphology<sup>6</sup>, all four dense granules proteins were secreted, and labeled a structure at the invasion site underlying the apical tip of the sporozoite (figure 5A-B and S3F-G). At 2 hours post invasion, DG 1, 2 and 3 were observed at the interface between parasite and host cell, while DG4 was found associated with the PV (figure 5C-D, S3H-I and S4B-D). U-ExM on intracellular parasites revealed that DG2 forms a ring at the interface that is also apparent by NHS-ester-staining (figure 5E and video S1). In later stages of parasite intracellular development (4 nuclei), DG2 was seen both at the interface (highlighted by arrowhead) as well as in the nascent intracellular dense granules (figure 5F and video S1). We also investigated protein localization *in vivo* by performing immunofluorescence on sectioned mouse intestine infected with DG3-HA. Figure 5G shows accumulation of DG3 at the infection site (green), right beneath the intracellular parasite's cytosol labeled with an antibody to lactate dehydrogenase (LDH)<sup>44</sup> shown in red. Again, we observed a ring-like structure at the interface between the host and the parasite (highlighted by labeling for host actin with phalloidin in figure 5H).

#### ***Cryptosporidium* possesses a secretory organelle not described in other apicomplexans**

Novelty-TAGM analysis identified two further protein clusters, '9' and '11', that were enriched in previously uncharacterized secretory proteins (figure 1F). We tagged the loci of one protein from each cluster (cgd1\_3810 for cluster 9 and cgd4\_3050 for cluster 11, figure S1). Both tagged proteins were observed just apical of the nucleus and in a pattern distinct from all other *Cryptosporidium* organelles observed in our studies thus far (figure 6A-B). U-ExM revealed numerous small vesicles around the nucleus that were distinct from dense granules (figure 6C-D). We named this organelle small granules (SG) owing to their apparent smaller size compared to dense granules, and two of its proteins SG1 (cgd1\_3810) and SG2 (cgd4\_3050). Combining both clusters yielded 23 proteins which include two proteins (cgd7\_200 and cgd7\_3390) with uncharacterized homologs in *Toxoplasma* (TGME49\_268690 and TGME49\_232600) assigned by *Toxoplasma* hyperLOPIT to the ER and apicoplast<sup>40</sup>. The small granule clusters also contain multiple proteins belonging to the WYLE and insulinase multigene families of putative effectors<sup>11</sup>. Immunofluorescence assays conducted on sporozoites at different time points following infection showed that small granules are secreted during invasion (figure S5A-B). Two hours after

invasion, SG1 was found to localize to a ring at the host/parasite interface (figure 6E) while SG2 was found within the VVL staining suggesting a localization in the PV (figure 6F). The localization of SG1 *in vivo* matches the *in vitro* results, again revealing a ring at the interface of host and parasite (figure 6G-H).

There are, therefore, parallels in the post-invasion fates of dense and small granule proteins, which prompted the question of why these two classes of secretory organelles might exist. We thus examined their expression programs. Dense granule protein transcription is skewed towards the end of the 12-hour intracellular replication cycle while small granule transcripts are already evident at the two- and six-hour time points (figure S5C). Moreover, small granule transcripts are most abundant in sporozoites, and show low expression in *in vitro* female gametes, whereas dense granule proteins show a markedly different life-stage expression profile (figure S5D). Collectively, these data suggest different timing of organelle biogenesis that might also result in different secretion patterns. To test this, we conducted time course experiments, imaging DG3-HA and SG1-HA expressing parasites every two hours over the first asexual cycle. We observed multiple SG1 positive granules in parasites with a single nucleus as early as four hours. In contrast, newly formed dense granules were not detected prior to the eight-hour timepoint (figure 6I and S5E-G show representative images and their quantification). We thus conclude that small granules are distinct from dense granules in size, localization, protein content and timing of biogenesis and therefore constitute a distinctive secretory organelle.

## **DISCUSSION**

*Cryptosporidium* occupies a central position in our understanding of the evolution of intracellular parasitism. Recent whole genome phylogenies place *Cryptosporidium* at the base of the apicomplexan tree, within or close to the gregarines, which parasitize invertebrates<sup>45,46</sup>. Common with several gregarines, *Cryptosporidium* shows a profound reduction of the mitochondrion and loss of the relict plastid<sup>47</sup>. Our dataset experimentally validates the presence of a highly reduced mitochondrion in *Cryptosporidium* (see ‘mitosome’ in figure 1). The mitochondrial genome, and much of the mitochondrial electron transport chain has been lost in *Cryptosporidium*<sup>48</sup>, but we demonstrate that multiple components of the mitochondrial iron-

sulfur cluster assembly machinery including cysteine desulfurase, ISC scaffold protein, ferredoxin NADPH oxidoreductase, chaperone hsp70, nucleotide exchange factor grpE, ISC transporter, and P-loop NTPase localize to the same organelle as the beta subunit of MPP and the alternative ubiquinol oxidase (table S3).

Gregarines are epiparasitic, a peripheral cellular interaction, where only the apex of the parasite is submerged within the host cell. *Cryptosporidium* may be viewed as the next step in the progression towards full parasite internalization<sup>49</sup>. Conservation of ultrastructural features of the host-parasite interface between gregarines and *Cryptosporidium* has been noted<sup>50</sup>, but very little is known about these structures at the molecular level. This study assigned organellar localization to 1107 *Cryptosporidium* proteins. We paid particular attention to secretory organelles as secreted effectors are central to the ability of intracellular pathogens to manipulate their host cell<sup>14,17</sup>. We assigned 154 secretory proteins to four compartments. The red queen hypothesis predicts that the adaptation of parasites to specific hosts, and the constant arms race between host and pathogen, will result in highly specialized, and fast-evolving protein cohorts amongst the secreted effectors<sup>51</sup>. In contrast, housekeeping functions underlie purifying selection driving conservation. While we found a high degree of conservation for the proteasome, ribosome, ER, and the combined nucleus and cytosol compartment within the Apicomplexa (using *Toxoplasma gondii* and *Plasmodium falciparum* as reference), secretory organelles, in particular dense and small granules, were populated by proteins largely unique to *Cryptosporidium* (figure 7A). Rhoptries, dense granules, micronemes and small granules also stood out with the highest rates of non-synonymous to synonymous single nucleotide polymorphism among 118 genomes of different species and strains of *Cryptosporidium* underscoring their role in host adaptation (figure 7B).

Apicomplexan host cell invasion as established in *Toxoplasma* and *Plasmodium* relies on a structure called the moving junction, built from components secreted by the micronemes (AMA1) and the rhoptries (RON2/4/5)<sup>52</sup>. However, these proteins are not present in *Cryptosporidium*. Rhoptries and micronemes are likely to hold the key to understanding *Cryptosporidium* invasion, and the hyperLOPIT dataset greatly expanded our knowledge on their content. One particularly interesting rhoptry candidate (cgd3\_1330) possesses a formin

homology domain, that might trigger the actin polymerization observed during *Cryptosporidium* invasion<sup>6,53,54</sup>. We also identified numerous additional microneme proteins featuring likely adhesin domains, and a CLAMP protein homolog (cgd4\_2350)<sup>55</sup>. Two proteins ascribed to the micronemes using antibody reagents (cgd1\_3550 and cgd2\_1590) show different hyperLOPIT assignment and may require further study<sup>56</sup>. The microneme proteome includes five thrombospondin (TSP) domain proteins (CpTSP7-11) out of the 12 previously predicted<sup>57</sup>. Interestingly, TSP proteins are also present in the rhoptry (CpTSP12), and three are part of the yet uncharacterized hyperLOPIT cluster 8 (TRAPC1, CpTSP3 and 6). Work in *Toxoplasma* suggested two subpopulations of micronemes<sup>58,59</sup>. We sought to test this by tagging both TRAPC1 (cgd1\_3500) and CpTSP6 (cgd6\_2310) but were unable to in multiple attempts.

*Cryptosporidium* elaborates an interface featuring multiple rings, bands, tight junctions, and a membranous metabolite exchanger named the feeder organelle. In this study we define the protein content of the dense granules (79 proteins) and demonstrate that following secretion these proteins are found at this interface and are presumably involved in its formation and function. INS-16 (cgd3\_4270) was recently assigned to dense granule-like organelles using an antibody but was not observed in our dataset<sup>60</sup>. We examined four of these proteins and found them to be secreted from the apical region of the sporozoite into the nascent vacuole during invasion, where three remained associated with the interface (figure 5 and S3). These structures match the previously described dense band and ring-shaped junction<sup>5,10</sup>. Interestingly, ROP5 and 6 similarly form a ring, and we note that these rhoptry proteins feature a transmembrane domain while the dense granule proteins identified do not<sup>6</sup>. While the true nature of the ring remains to be revealed, electron microscopy shows a tight apposition of host and parasite membranes akin to a tight junction. Multiple dense granule proteins in *Toxoplasma* have been shown to be translocated into the host cell cytoplasm and nucleus, where they suppress host immunity<sup>17</sup>. The four *Cryptosporidium* proteins tested thus far do not appear to have such tropism, however, we note HSP101 in the dense granule proteome, which is associated with translocation in *Plasmodium*<sup>61</sup>, and could function similarly in *Cryptosporidium*.

Using the hyperLOPIT dataset we discovered a previously uncharacterized apicomplexan secretory organelle, the small granules. Initially we suspected them to be a precursor of one of

the other secretory compartments, however, further analysis showed them to be morphologically distinct, independently secreted, to have no overlap in protein content with rhoptries, micronemes, or dense granules, and to follow a unique temporal pattern of biogenesis. We studied two of these proteins and found one to be delivered to the host-parasite interface, as are proteins secreted from dense granules and rhoptries (figure 7C). Protein targeting thus does not provide an immediate rationale for differential compartmentalization, however, timing might. Dense granule, rhoptry, and microneme proteins are assembled at the end of schizogony to be loaded into new invasive zoites, but their mRNA is no longer found in extracellular sporozoites. Small granule proteins are transcribed and translated much earlier during intracellular development. This suggests the continuous delivery of effectors during intracellular development in contrast to deployment once during invasion (figure 7D). This is reminiscent of ‘non-canonical’ dense granules in *Toxoplasma*, which have been associated with the sporozoite stage, and may be secreted constitutively<sup>20,62</sup>. We note that the *Cryptosporidium* sporozoite appears to be loaded, not only with small granule proteins as shown here, but also with mRNA encoding them<sup>8</sup>. Stored but silenced mRNA plays an important role in *Plasmodium* gametogenesis<sup>63</sup> to jumpstart development, but has not been studied in asexual stages.

This study also defines the molecular components of two organelles critical to transmission: oocyst wall and crystalloid body. The oocyst wall protects the parasite from the environment and is crucial to the resistance of *Cryptosporidium* to water disinfection. We note that the oocyst wall proteome contained two patched family homologs. Patched family proteins transport cholesterol and other lipids<sup>64</sup> and are thus prime candidates for the delivery of the lipid component of the oocyst wall believed to confer its resistance to chemical attack<sup>30</sup>. The crystalloid body has been shown to be critical for transmission in *Plasmodium spp.*, however its function remains enigmatic. Interestingly, crystalloid body proteins are, to date, the only proteins found to be present in the sporozoite but absent in the merozoite. This suggests a role important for the development and survival of the sporozoite and recent studies targeting a transcription factor controlling their expression lend support to this idea<sup>65</sup>.

The sheer number of proteins discovered to be secreted and or translocated in this study underscores the complexity of the *Cryptosporidium* host-parasite interface and the role of the

406 parasite in shaping it. Building and maintaining the structure requires multiple organelles, and  
407 future studies will reveal how delivery is organized over space and time. The *Cryptosporidium*  
408 secreted effector proteome provides an important resource to dissect this process. Overall, these  
409 data significantly elevate the opportunities to study *Cryptosporidium* and related parasites by  
410 populating major parts of the cell with proteins.

## **Acknowledgments**

We thank Fabio Tosini and Honorine Ward for sharing antibodies, and Mike Deery for proteomic data acquisition. This work was funded in part by grants from the National Institutes of Health (R01AI112427 and R01AI148249) and the Pennsylvania Department of Health to B.S., the Wellcome Trust (214298/Z/18/Z) to RFW, an EMBO postdoctoral fellowship to A.G. (ALTF 58-2018), a Leverhulme Early Career Fellowship co-funded by the Isaac Newton Trust (ECF-2015-562) to K.B., and a Todd-Bird Junior Research Fellowship from Oxford New College to O.M.C.

## **Author contributions**

B.S. and R.F.W conceived the study, K.M.S. and K.B. performed the hyperLOPIT experiments, K.B. A.G. and K.M.S. analyzed the hyperLOPIT experiments with support from O.M.C., A.G. performed the characterization experiments with contributions from K.M.S. to tagging and B.W. to U-ExM, L.B. performed electron microscopy, A.G. and B.S. wrote the manuscript with contributions from all authors.

## **Declaration of interests**

The authors declare no competing interests.

## **Main figure legends**

### **Figure 1: HyperLOPIT resolved *Cryptosporidium* sporozoite sub-compartments**

(A) Representation of the hyperLOPIT experiment design. Left, schematization of the sporozoite stage excysted from the oocyst with select compartments highlighted including the secretory organelles rhoptry, micronemes and dense granules. Right, representation of the gradient of cell lysate used to create 10 fractions labeled with TMT10plex tags and pooled for chromatography mass spectrometry. The bottom graphic shows the resulting protein abundance profiles with rhoptry and oocyst wall across the 10 fractions highlighted (from the second experiment). (B) Venn diagram showing the intersection of proteins quantified in the two hyperLOPIT experiments. (C) t-SNE projection of the concatenated data. Dots correspond to individual proteins and colors correspond to HDBSCAN-identified clusters. (D) Plotting of signal peptides



(green) onto the t-SNE map shows some clusters enriched in signal peptide containing proteins. (E) Plotting compartment marker proteins onto the t-SNE projection (see table S2, and panel f for colors). (F) Plotting the result of the Novelty-TAGM analysis. In addition to the 11 clusters used from the reference list, 11 more clusters, named cl.1 to 11, were identified. Dash circles correspond to the clusters enriched in signal peptide containing proteins highlighted in d, e and f and further investigated in this study.

### **Figure 2: Distinction between oocyst wall and crystalloid body proteins**

(A) Pie chart showing the fractions of previously known, predicted and newly assigned proteins for the oocyst wall and the crystalloid body. (B) Normalized transcript expression of asexual and female stages *in vitro* from Tandel *et al.*<sup>8</sup> for the oocyst wall, the crystalloid body and ribosome (40S+60S) clusters. (C-D) Immunofluorescence detection of two crystalloid body candidates in green, cgd8\_4310 (CB1) in (C) and cgd7\_1730 (cpCCP1) in (D) showing labeling basal to the parasite nucleus. Antibodies against the micronemal protein cp23 were used to label the parasite plasma membrane in red. Hoechst in blue, scale bar 1  $\mu$ m. (E) U-ExM on extracellular sporozoite showing the NHS-ester labeling pattern. g, granules; r, rhoptry; c, conoid. Scale bar 5  $\mu$ m. (F) U-ExM of extracellular tagged cgd8\_4310 (CB1) parasite (green). NHS-ester in red, Hoechst in blue, scale bar 5  $\mu$ m. (G) Widefield images of tagged candidates cgd7\_1730 (cpCCP1) and cgd8\_4310 (CB1) in green, 2 hours after invasion, showing all signals within the parasite. *Vicia Villosa* lectin (VVL) in red shows the parasitophorous vacuole as previously described<sup>66</sup>. Scale bar 1  $\mu$ m.

### **Figure 3: Repertoire of microneme and rhoptry proteins**

(A) Widefield image of a microneme candidate cgd7\_1960 in green showing a colocalization with the micronemal protein gp40/gp900. Gp40/gp900 in red, Hoechst in blue, scale bar 1  $\mu$ m. (B) Widefield image of a rhoptry candidate, cgd5\_20 in green, showing an apical staining pattern basal to the rhoptry neck marker CpPRP1. CpPRP1 in red, Hoechst in blue, scale bar 1  $\mu$ m. (C-D) U-ExM of (C) cgd7\_1960 (CpMIC2)- and (D) cgd5\_20 (ROP7)-stained cells showing respective protein location in numerous small compartments consistent with microneme organelles and the bulb of the single rhoptry. HA in green, NHS-ester in red, Hoechst in blue, scale bar 5  $\mu$ m. (E)

Graphic representing the normalized transcript expression levels at 2, 6 and 12 hours post infection from Mauzy *et al.*<sup>39</sup> of all the candidates identified in the microneme, rhoptry and ribosome (40+60S) clusters. The dash lines connect the medians. (F) Pie chart showing the fraction of previously confirmed, predicted and newly assigned proteins in microneme and rhoptry clusters. (G-H) Widefield images of tagged candidates cgd7\_1960 (CpMIC2) (G) and cgd5\_20 (ROP7) (H) in green in extracellular sporozoite or 2 hours after invasion, showing respectively protein locations at the parasite surface and the parasitophorous vacuole space. Cp23 or VVL in red, Hoechst in blue, scale bar 1  $\mu$ m.

#### **Figure 4: Identification of the dense granule proteins in *Cryptosporidium***

(A-B) Widefield images of dense granule candidates, cgd1\_590 (A) and cgd3\_1690 (B) in green, showing a protein location between the nucleus and the apical end. Cp23 in red, Hoechst in blue, scale bar 1  $\mu$ m. (C-D) U-ExM showing cgd1\_590 (C) and cgd3\_1690 (D) in green relative to the NHS-ester labeled granules (red). Hoechst blue, scale bar 5  $\mu$ m. (E) Zoomed images of the 4 tagged candidates within the NHS-labeled granules showing different patterns of co-location by U-ExM. Scale bar 1  $\mu$ m. (F) Immuno electron microscopy of DG2-tagged candidate (cgd5\_1440) in the dense granules of excysted sporozoites. m, micronemes; dg, dense granule. Scale bar, 100 nm.

#### **Figure 5: Dense granule exocytosis and localization during and post host invasion**

(A-B) Widefield images of cgd1\_590 (DG3) (A) and cgd3\_1690 (DG4) (B) secretion during the invasion process identifiable by the C-shaped parasite. Cp23 in red, HA in green, Hoechst in blue. Scale bar 1  $\mu$ m. (C-D) Widefield images 2 hours post invasion of DG candidates cgd1\_590 (DG3) (C) and cgd3\_1690 (DG4) (D) in green, showing an accumulation at the interface and in the parasitophorous vacuole respectively. VVL in red, Hoechst in blue, scale bar 1  $\mu$ m. (E) U-ExM of cgd5\_1440-HA (DG2) in green, showing the accumulation of the protein at the interface with the host cell also labeled by the NHS-ester. The image shows a parasite with 1 nucleus without granules in its cytoplasm. Scale bar, 5 $\mu$ m. (F) U-ExM image showing the accumulation of cgd5\_1440-HA (DG2) staining in nascent dense granules in addition to the interface of a 4-nucleus

parasite. Arrowhead, interface. Scale bar, 5µm. A 3D-reconstruction of E and F can be found in video S1. (G) ‘Swiss rolled’ intestinal image of a mouse infected with cgd1\_590 (DG3)-tagged strain confirming the accumulation of the protein at the interface *in vivo*. LDH in red, HA in green, Hoechst in blue. Scale bar, 10µm. (H) Zoom in on an image of an intestinal Swiss roll infected with cgd1\_590 (DG3) with the brush border highlighted by the host actin (white) confirming the DG3 location at the interface. LDH in red, HA in green, Phalloidin in white, Hoechst in blue. Scale bar, 1µm.

**Figure 6: Discovery of an additional secretory organelle in *Cryptosporidium*, the small granules**

(A-B) Widefield images of small granule candidates, cgd1\_3810 (SG1) (A) and cgd4\_3050 (SG2) (B) in green, showing staining proximal to but just apical of the nucleus. Cp23 in red, Hoechst in blue, scale bar 1 µm. (C-D) U-ExM of cgd1\_3810 (SG1) (C) and cgd4\_3050 (SG2) (D) showing overlap with a portion of the NHS-ester labeled granules. HA green, NHS-ester red, Hoechst blue, scale bar 5 µm (E-F) Widefield images 2 hours post invasion of SG candidates, cgd1\_3810 (SG1) (E) and cgd4\_3050 (SG2) (F) in green, showing an accumulation at the interface and in the parasitophorous vacuole respectively. Cp23 in red, Hoechst in blue, scale bar 1 µm. (G) ‘Swiss rolled’ intestinal image of a mouse infected with cgd1\_3810 (SG1)-tagged strain confirming the accumulation of the protein at the interface *in vivo*. LDH in red, HA in green, Hoechst in blue. Scale bar, 10µm. (H) Zoomed image of an intestinal staining of cgd1\_3810 (SG1) with the brush border highlighted by phalloidin labeling the cortical actin (white). LDH in red, HA in green, Hoechst in blue. Scale bar, 1µm. (I) Quantification of the percentage of parasites with small granules and dense granules present every two hours over the course of the first asexual replication cycle (12h) using SG1-HA and DG3-HA lines as shown in representative images in figure S5E and F.

**Figure 7: Comparative properties of secretory organelles in *Cryptosporidium***

(A) Representation of orthologues of *Plasmodium falciparum* and *Toxoplasma gondii* within each *Cryptosporidium* cluster. The five secretory compartments have more than 50% of their contents unique to *Cryptosporidium*. (B) Box-plots (first quartile and third quartile) with whiskers showing

527 the 10-90 percentile of the ratio of non-synonymous to synonymous single nucleotide  
528 polymorphism (SNP) for genes identified in each cluster using 118 genomes available on  
529 CryptoDB. The vertical line indicates the dataset mean at 0.58 and the clusters are ranked from  
530 higher (top) to lower (bottom) ratio. (C) Model showing the localization of secreted protein  
531 identified in this study and in our previous studies<sup>6,21</sup>. Three secretion destinations were  
532 observed, the parasitophorous vacuole, the host-parasite interface and the host cell cytoplasm.  
533 (D) Model of *Cryptosporidium* secretory organelles biogenesis and secretion across the 12-hours  
534 of asexual replication. The micronemes are secreted first when the parasite is extracellular while  
535 the rhoptry, dense and small granules content are secreted during invasion of the enterocyte.  
536 The small granules are started to be made between 2 to 4 hours while the dense granules aren't  
537 detected before 8-hours post invasion. The rhoptry and the micronemes are the last secretory  
538 organelles to be made at 10 to 12-hours post invasion. Parasites egressing from the host cell will  
539 then contain the 4 secretory organelles.

540 **STAR Methods**

541 **RESOURCE AVAILABILITY**

542 **Lead contact**

543 For access to reagents or parasite strains used in this study please contact the Lead Contact, Boris  
544 Striepen ([striepen@upenn.edu](mailto:striepen@upenn.edu)).

545

546 **Material availability**

547 All unique/stable reagents generated in this study are available from the Lead Contact with a  
548 completed Materials Transfer Agreement.

549

550 **Data and code availability**

551 The mass-spectrometry-based proteomics data have been deposited to the ProteomeXchange  
552 Consortium via PRIDE with the dataset identifier PXD015269. The data is also available through  
553 cryptoDB.

554 No original code was used in this study.

555 Any additional information required to reanalyze the data reported in this work paper is available  
556 from the Lead Contact upon request.

557

558 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

559 **Mouse models of infection**

560 *ifn $\gamma$ <sup>-/-</sup>* (stock no:002287) were purchased from Jackson Laboratory and used to build a breeder  
561 colony. For infection experiments both males and females (ages 6 to 8-weeks) were used in  
562 groups of four per cage (see Sateriale et al.<sup>67</sup> for additional detail). All protocols for animal  
563 maintenance, care and experimentation were approved by the Institutional Animal Care and Use  
564 Committee of the University of Pennsylvania (protocol #806292).

565

566 **Parasite strains**

567 *Cryptosporidium parvum* oocysts were purchased from Bunchgrass Farms. Transgenic strains  
568 were propagated within infected *ifn $\gamma$ <sup>-/-</sup>* mice (stock no:002287 bred in house). Oocysts were then

purified from fecal collections using sucrose flotation followed by a cesium chloride gradient and detailed in the Method Details.

## **Host cell culture**

HCT8 obtained from ATCC were cultured in DMEM supplemented with 10% FBS at 37c with 5% CO<sub>2</sub>.

## **METHOD DETAILS**

### **Generation of transgenic *Cryptosporidium***

Transgenic parasite lines were generated as previously described<sup>6</sup>. Briefly excysted oocysts were electroporated with a plasmid containing the cas9 and the guide with the DNA repair. The DNA repair is composed of a 30pb homology region around the guide induced double strand break, a triple hemagglutinin tag, a generic 3' UTR and the nanoluciferase-neomycin cassette under the enolase promoter (figure S1). The neomycin gene was used as a selection marker conferring resistance to paromomycin, while nanoluciferase was used to track the infection in mice by assaying enzyme activity in the feces. Mice were infected with transgenic parasites by gavage and paromomycin was supplied in drinking water. Parasites were purified from feces and used for downstream experiments.

DG1 Cgd3\_600-HA (Guide 025gdeF/R \_ Amplification 025repF/R \_ Integration 025F/R)

DG2 Cgd5\_1440-HA (Guide AG75/76 \_ Amplification AG77/78 \_ Integration AG107/108)

DG3 Cgd1\_590-HA (Guide 023gdeF/R \_ Amplification 023repF/R \_ Integration 023F/R)

DG4 Cgd3\_1690-HA (Guide AG407/408 \_ Amplification AG409/410 \_ Integration AG411/412)

ROP7 Cgd5\_20-HA (Guide AG118/119 \_ Amplification AG120/121 \_ Integration AG182/162)

CpMIC2 Cgd7\_1960-HA (Guide AG357/358 \_ Amplification Ag359/360 \_ Integration AG376/377)

SG1 Cgd1\_3810-HA (Guide AG463/464 \_ Amplification AG465/466 \_ Integration AG475/476)

SG2 Cgd4\_3050-HA (Guide AG467/468 \_ Amplification AG469/470 \_ Integration AG477/478)

cpCCP1 Cgd7\_1730-HA (Guide AG483/484 \_ Amplification AG485/486 \_ Integration AG491/492)

CB1 Cgd8\_4310-HA (Guide AG487/488 \_ Amplification AG489/490 \_ Integration AG493/494)

PCR mapping data for all modified loci can be found in figure S1, and the sequences for all primers used can be found in table S5.

### **Immunofluorescence microscopy**

Immunofluorescence assays were performed as previously described<sup>6</sup>. Briefly, oocysts were excysted for 1 hour and added to a coverslip with a HCT8 cell layer for 2 hours. Cells were fixed either with 4% PFA at room temperature for 30 minutes or 100% methanol at -20C for 7 minutes. Following permeabilization with 0.5% Triton X-100 for 15 minutes, the cells were blocked with 4% BSA for 1 hour. Primary antibodies were diluted in 1% BSA and cells were incubated for 1 hour, followed by 3 washes with PBS. Secondary antibodies were diluted in 1% BSA and cell were incubated for 1 hour, followed by 5 minutes incubation with Hoechst DNA dye diluted in PBS and 5 PBS washes. Coverslips were mounted using vectashield and imaged with either a widefield Leica DM6000B or a OMX SR DeltaVision microscope. Reference and dilution for all antibodies used are listed in the key resources table and table S6. All immunofluorescence experiments were performed at least twice independently, and more than 10 individual parasites were imaged.

### **Ultrastructure expansion microscopy**

Ultrastructure expansion microscopy was applied to *Cryptosporidium* sporozoites as described for *Toxoplasma*<sup>68</sup>. Briefly, excysted sporozoites were added to poly-D-lysine coated coverslips. Fixation with a combination of acrylamide and formaldehyde prevents crosslinking before embedding the sample in a water-based gel. Samples are denatured at 95 °C prior to expansion in water to around 4 times their original size. Gels are shrunk in PBS to reduce the quantity of antibodies and dye needed for labeling. Reference and dilution antibodies used are listed in the key resources table and table S6. The gels were expanded again in water and were imaged using the OMX SR DeltaVision microscope. A total of 12 experiments were performed on extracellular and 2 on intracellular parasites.

### **Immunoelectron microscopy**

Immunoelectron microscopy was conducted as previously described<sup>6</sup>. Briefly, excysted sporozoites were fixed in 4% paraformaldehyde for 30 minutes at room temperature, kept in fixative at 4°C. Samples were incubated in 0.1% glycine in phosphate buffer, pelleted, and embedded in 12% gelatine (porcine skin gelatine, Sigma). Gelatine blocks were infused for 24 hours in 2.3M sucrose at 4°C and frozen in liquid nitrogen. Sectioning was performed on a Leica UC7 cryo-ultramicrotome. 80 nm cryosections were picked-up in a 1:1 mixture of 2.3M sucrose and 2% methylcellulose in water, placed on nickel grids, and stored at 4°C. Grids were floated on PBS 2% gelatine for 30 minutes at 37°C to remove methylcellulose/sucrose mixture, then blocked with 1% skin-fish gelatine in PBS for 5 minutes. Successive incubation steps were performed on consecutive drops of rat monoclonal anti-HA (clone 3F10, Roche) in 1% BSA, rabbit polyclonal anti-rat IgG antibody (Sigma) in PBS 1% BSA, Protein A-gold (UMC Utrecht) in PBS 1% BSA. Four 2-minutes washes in PBS 0.1% BSA were performed between steps. After Protein A, grids were washed 4 times for 2 minutes with PBS, fixed for 5 minutes in 1% glutaraldehyde in water then washed 6 times for 2 minutes with distilled water. Grids were then incubated with 2% methylcellulose: 4% uranyl acetate 9:1 for 15 minutes on ice in the dark, picked-up on a wire loop and air-dried. Observation and image acquisition was performed on a Jeol 1200 EXII transmission electron microscope at the Electron Microscopy Platform of the University of Montpellier.

#### **Sample preparation for hyperLOPIT**

$8 \times 10^8$  oocysts for the first experiment and  $2.4 \times 10^9$  oocysts for the second experiment were excysted overnight using the previously described protocol<sup>6</sup> and detailed steps are summarized in table S1. The cells were centrifuged at 2000g for 20 minutes and resuspended in 5ml of the homogenization medium (HM: 0.24M sucrose, 10mM HEPES·KOH pH 7.4, 1mM EDTA) supplemented with proteinase inhibitors (Complete EDTA-free protease inhibitor cocktail, Roche). The cells were mechanically lysed by three consecutive rounds of nitrogen cavitation using the previously described method with minor modifications<sup>40</sup>. Briefly, a suspension of cells in HM at a number density of approximately  $1 \times 10^9$  ml<sup>-1</sup> was transferred into a pre-chilled cavitation vessel (model 4639, Parr Instruments), pressurized to 2800 PSI and incubated on wet ice for 15 min with periodic agitation. The content of the vessel was discharged at a flow rate of



1-2 droplets per second into a pre-chilled centrifugal tube of an appropriate volume and let settle for 1-2 min on ice. The collected homogenate was centrifuged for 15 min at  $2000 \times g_{\max}$  to sediment the unlysed material. The supernatant was collected and set aside, whereas the pellet was resuspended in HM to the original volume and subjected to another round of cavitation. The supernatants collected after three rounds of homogenization were pooled and fractionated further.

#### **HyperLOPIT subcellular fractionation**

The subcellular fractionation was performed as previously described<sup>22,40,69</sup>. All gradient media (GM) were prepared by mixing HM with the iodixanol working solution (IWS: 50% (w:v) iodixanol, 40 mM sucrose, 10 mM HEPES·KOH pH 7.4, 1 mM EDTA). To separate the subcellular membrane vesicles and particles from soluble proteins, the homogenate was underlaid with a discontinuous density gradient of 6% and 25% (w:v) of iodixanol and subjected to centrifugation for 1.5h at  $100,000 \times g_{\max}$ , 4°C (SW32Ti rotor, Optima L-80XP ultracentrifuge, Beckman) with the maximum acceleration and minimum deceleration. An aliquot of the supernatant enriched with soluble proteins was collected, mixed with six volumes of chilled acetone and removed to -20°C to induce protein precipitation. The opaque bands at the interfaces of the 6% and 25% iodixanol layers containing enriched subcellular membrane vesicles and particles were collected, diluted with HM to bring the iodixanol concentration below 6% (w:v), and centrifuged for 1h at  $200,000 \times g_{\max}$ , 4°C (SW55Ti rotor, Beckman) to further purify and concentrate the crude membranes. The membrane pellets were resuspended in GM containing 25% (w:v) iodixanol and carefully layered under a linear density gradient prepared in advance by diffusion of four layers of GM containing 8, 12, 16, and 20% (w:v) iodixanol. The density gradient loaded with the sample was centrifuged for 8h at  $100,000 \times g_{\max}$ , 4°C (VTi65.1 rotor, Beckman) with the maximum acceleration and minimum deceleration. The resolving gradient (total volume of 11.2 ml) was harvested by drop-wise dispensation from the bottom of the pierced centrifugation tube into 22-23 approximately equal-volume fractions. The density profile of the gradient was determined by measuring the refractive index of the fractions (Eclipse Handheld Refractometer 45-02, sugar 0-32%, Billingham and Stanley). The total protein content of the fractions was measured by the BCA protein assay

(Thermo Fischer Scientific) according to the manufacturer's instructions. The collected fractions were stored at  $-80^{\circ}\text{C}$  until used.

#### **Sample preparation for LC-MS/MS analysis**

Peptide samples for LC-MS/MS analysis were prepared as described by Barylyuk et al.<sup>40</sup> with minor modifications. Proteins were extracted from the gradient fractions by precipitation with 10% (w/v) trichloroacetic acid (TCA), as described elsewhere. The pellets were washed with 1 ml of 10% (w:v) aqueous TCA to remove residual iodixanol and polar components of GM, and 2×1ml of chilled acetone to extract residual TCA. To facilitate resolubilisation of proteins, the protein pellets were pulverized by sonication while suspended in chilled acetone in the final washing step (5 × 30s on, 30s off, Bioruptor Plus, Diagenode). The protein pellets, including the acetone-precipitated proteins from the soluble fraction (see above), were reconstituted in triethylammonium bicarbonate (TEAB) buffered solution (pH 8.3) containing 8M urea, 0.1% SDS assisted by sonication (5-10 cycles of 30s on, 30s off, Bioruptor Plus, Diagenode). Protein concentration was measured by the BCA assay. The gradient fractions were combined into nine pools to afford a total protein amount of 18-156  $\mu\text{g}$  per pool. These nine pools together with the soluble protein fraction constituted a 10plex.

The side chains of cysteine residues in proteins were reduced by incubation with 10 mM Tris(2-carboxyethyl)phosphine (TCEP; Sigma-Aldrich) for 1h at room temperature and alkylated by incubation with iodoacetamide (Sigma-Aldrich) at approximately 17 mM final concentration for 30 min at room temperature in the dark. The reduced and alkylated proteins were precipitated by the addition of six volumes of chilled acetone and incubation at  $-20^{\circ}\text{C}$  overnight. The protein pellets were reconstituted in 100 mM TEAB buffered solution (pH 8.3) with the assistance of sonication (5 × 30s on, 30s off, Bioruptor Plus, Diagenode) and digested with 2  $\mu\text{g}$  of sequencing-grade trypsin (Promega)  $37^{\circ}\text{C}$  overnight. The digests were centrifuged for 10 min at  $16,000 \times g_{\text{max}}$  at  $4^{\circ}\text{C}$ , and the supernatants were transferred to clean 1.5 ml Protein LoBind microcentrifuge tubes (Eppendorf) and labeled with TMT10plex isobaric tagging reagents (Thermo Fisher Scientific) according to the manufacturer's instructions. Briefly, 0.8 mg of TMT10plex reagents were brought to room temperature and dissolved in 41  $\mu\text{l}$  of LCMS-grade acetonitrile

immediately before use. The peptide solutions (approximately 100  $\mu$ l) were transferred to the TMT10plex reagent vials and the reaction mixtures were incubated at room temperature for 2h with constant agitation (800 RPM, PHMT thermomixer, Grant Bio Instruments). The reaction was stopped by adding 8  $\mu$ l of 5% (v/v) hydroxylamine solution and incubation for 15 min at room temperature with agitation. The TMT-labeled fractions were combined and reduced to dryness in a refrigerated (4°C) vacuum centrifuge (Labconco).

The peptides were fractionated by high-pH reverse-phase chromatography as per Barylyuk et al.<sup>40</sup> using a Waters Acquity UPLC system with a BEH C18 (1.7  $\mu$ m, 2.1 x 150 mm) column and BEH C18 (1.7  $\mu$ M, 2.1 x 5 mm) VanGuard pre-column. The eluents used were: 20 mM ammonium formate in HPLC-grade water, pH 10 (eluent A); 20 mM ammonium formate in LCMS-grade acetonitrile: HPLC-grade water 80:20 (v/v), pH 10 (eluent B). The flow rate was kept at 0.244 ml min<sup>-1</sup>. The peptides were reconstituted in 100  $\mu$ l of 95% eluent A: 5% eluent B blend and loaded onto the column equilibrated to the same solvent conditions. After the initial equilibration for 10 min. at 5% eluent B, a linear gradient of 5-75% eluent B was applied over 50 min followed by column washing with 100% eluent B for 5 min and equilibration into 5% eluent B for 10 min. The total time of the elution program was 77.5 min. Fifty fractions were collected every minute between minute 10 and 60 of the program, concatenated into 15 pools as per<sup>70</sup>, and reduced to dryness in a refrigerated vacuum centrifuge.

### **LC-MS/MS Analysis of peptides**

LC-MS/MS analyses were performed on an Orbitrap™ Fusion™ Lumos™ Tribrid™ mass spectrometer coupled to a Dionex Ultimate™ 3000 RSLCnano system (Thermo Fisher Scientific) as previously described<sup>40</sup>. The data were acquired in the positive-ion data-dependent acquisition mode with the SPS-MS<sup>3</sup> method with a total run time of 120 min. The peptide samples were resolubilized in 20  $\mu$ l of the sample loading solution (0.1% aqueous formic acid). Approximately 1  $\mu$ g of peptides was loaded onto a micro-precolumn (C18 PepMap 100, 300  $\mu$ m i.d. x 5 mm, 5  $\mu$ m particle size, 100 Å pore size, Thermo Fisher Scientific) with the sample loading solution for 3 min. Following the loading step, the valve was switched to the inject position, and the peptides were fractionated on an analytical Proxeon EASY-Spray column (PepMap, RSLC C18, 50 cm x 75

µm i.d., 2 µm particle size, 100 Å pore size, Thermo Fisher Scientific) using a linear gradient of 2-40 % (vol.) acetonitrile in aqueous 0.1% formic acid applied at a flow rate of 300 nl/min for 95 min, followed by a wash step (70% acetonitrile in 0.1% aqueous formic acid for 5 min) and a re-equilibration step. Peptide ions were analyzed in the Orbitrap at a resolution of 120,000 in an m/z range of 380-1,500 with a maximum ion injection time of 50 ms and an AGC target of 4E5 (MS<sup>1</sup> scan). Precursor ions with the charge states of 2-7 and the intensity above 5,000 were isolated in the quadrupole set to 0.7 m/z transmission window and fragmented in the linear ion trap via collision-induced dissociation (CID) at a 35% normalized collision energy, a maximum ion accumulation time of 50 ms and an AGC target of 1E4 (MS<sup>2</sup> scan). The selected and fragmented precursors were dynamically excluded for 70 s. Synchronous precursor selection (SPS) was applied to co-isolate ten MS<sup>2</sup>-fragments in the linear ion trap with an isolation window of 0.7 m/z in the range of m/z 400-1,200, excluding the parent ion and the TMT reporter ion series. The SPS precursors were activated at a normalized collision energy of 65% to induce fragmentation via high-energy collision-induced dissociation (HCD). The product ions were measured in the Orbitrap at a resolution of 50,000 in a detection range of m/z 100-500 with a maximum ion injection time of 86 ms and an AGC of 5E4 (MS<sup>3</sup> scan).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Raw LC-MS data processing and quantification

Raw LS-MS/MS data were processed with Proteome Discoverer v2.5 (Thermo Fisher Scientific). The spectra were searched against the annotated protein sequences of *Cryptosporidium parvum* isolate Iowa II (3,944 entries retrieved from CryptoDB.org release 52 on 27.05.2021)<sup>71</sup>. Common contaminant proteins – e.g., human keratins, bovine serum albumin, porcine trypsin – from the common Repository of Adventitious Proteins (cRAP, 115 entries, adapted from <https://www.thegpm.org/crap/>) were added to the database. The Sequest HT search engine was used. The precursor and fragment mass tolerances were set to 10 ppm and 0.6 Da, respectively. The enzyme was set to trypsin with up to two missed cleavages allowed. The static modifications were carbamidomethylation of cysteine and TMT6plex at the peptide N-terminus and lysine side chain. The dynamic modifications were TMT6plex at the side chains of serine, and threonine,

770 oxidation of methionine, and deamidation of asparagine and glutamine. In addition, the dynamic  
771 modifications of protein N-terminus considered were acetylation, the loss of methionine, and the  
772 combination of the two. The false discovery rate of peptide-to-spectrum matches (PSMs) was  
773 validated by Percolator v3.05.0 and only high-confidence peptides (FDR threshold 1%) of a  
774 minimum length of 6 amino acid residues were used for protein identification. The unmatched  
775 spectra were further searched against the same database with MS Amanda 2.0 using the same  
776 search parameters, except for not considering the protein N-terminal dynamic modifications. The  
777 PSM FDR was validated by Percolator as above. Strict parsimony was applied for protein  
778 grouping. The search results were merged globally by search engine, and the protein FDR was  
779 controlled at 5% cut-off.

780 TMT reporter ion abundances were obtained in Proteome Discoverer Reporter Ions Quantifier  
781 node using the most confident centroid method for peak integration with 20 p.p.m. tolerance  
782 window. The isotopic impurity correction as per the manufacturer's specification (TMT10plex Lot  
783 No. RJ239873) was applied. The results of data processing by Proteome Discoverer were exported  
784 as tab-delimited text files and processed further through an in-house data analysis pipeline in R  
785 version  $\geq 4.1$  (R Core Team), using Bioconductor packages MSnbase (v  $\geq 2.18.0$ ), pRoloc (v  
786  $\geq 1.32.0$ ), and tidyverse (v  $\geq 1.3.1$ ).

787 For protein quantification, the PSM-level output from Proteome Discoverer was filtered to retain  
788 only unique and unambiguous PSMs that were mapped to a single protein group not containing  
789 contaminant proteins, were matched to rank 1 peptide, and had no more than 50% precursor  
790 isolation interference, at least 50% of SPS precursor ions matched to sequence fragments, and  
791 an average reporter ion signal-to-noise ratio of at least 5. SMs containing missing reporter  
792 abundance values were removed if PSMs with complete quantitation data were matched to the  
793 same protein group, or if more than 4 values in the middle 8 TMT channels (TMT127N to  
794 TMT130C) were missing. The missing values in the TMT126 and TMT131 channels were  
795 considered missing not at random and imputed using a left-censored method 'MinDet' available  
796 with the function 'MSnBase::impute'. The missing values in the remaining channels were  
797 considered missing at random and were imputed with a hot-deck method 'knn' in the function  
798 'MSnBase::impute'. Prior to kNN-imputation of missing values, the individual PSM abundance

values were normalized to their sum across the TMT channels to eliminate the absolute intensity variance. The row-sum-normalized abundances of all PSMs attributed to a given protein group were summarized to the protein-level abundances using the median values across the TMT channels. Finally, proteins with complete abundance data in both 10plex datasets were concatenated to produce a 20plex dataset on 1742 proteins for further analysis.

### **Dimensionality reduction and unsupervised clustering**

The normalized 20plex dataset was visualized by principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) as described earlier by Barylyuk et al.<sup>25,40</sup>. Briefly, the data were PCA-transformed, reduced to top 14 principal components explaining approximately 99% of the variance and whitened, and embedded into two dimensions by running t-SNE for 5000 iterations with the parameters 'theta' and 'perplexity' set to 0 and 42, respectively.

Unsupervised clustering of proteins was performed on the normalized abundance data using the HDBSCAN\* algorithm<sup>72</sup> available through the Python library 'hdbscan' (<https://hdbscan.readthedocs.io/en/latest/index.html>). We used the distance to the 7<sup>th</sup> nearest neighbor to define the core distance for each protein ('min\_samples = 7') and allowed a minimum of 10 proteins per cluster ('min\_cluster\_size = 10'). Flat clusters from the cluster tree hierarchy were selected using 'cluster\_selection\_method = 'leaf''. All other parameters were kept at default values. The result was verified to include no clusters with zero stability ( $\lambda > 0$ ).

### **Bayesian novelty-detection analysis of the spatial proteomic data by Novelty-TAGM**

Semi-supervised prediction of protein subcellular localization was performed using a Novelty-TAGM model as described by Crook et al.<sup>27</sup>. Novelty-TAGM provides a Bayesian framework for the detection of protein clusters in the spatial proteomics data alongside with the ones specified by the markers, and uses Markov-chain Monte-Carlo (MCMC) methods to determine the posterior allocation of proteins to the marker-defined or discovered classes. The collapsed Gibbs sampler was run in parallel for 8 MCMC chains for 20,000 iterations per chain. The first 10,000 iterations were discarded ('burnin = 10000'), and the remaining 10,000 iterations were sampled

to retain every 10<sup>th</sup> state ('thin = 10'). The number of marker-defined components was 11, and the maximum number of novel components in the model was set to 15. Convergence was assessed by monitoring the number of proteins allocated to the outlier component of the model at each iteration of parallel MCMC chains and computing the Gelman-Rubin's diagnostic between the chains. Based on a value of the statistic below 1.2 indicating convergence, two MCMC chains were retained.

To analyze the uncertainty patterns in the number of components and protein allocation in the Novelty TAGM model, we computed a posterior similarity matrix ( $M$ ) by calculating the probability of any two proteins to be allocated to the same component. We then clustered proteins by applying the partitioning around medoids method to the dissimilarity matrix ( $1 - M$ ) with 26 centres (11 marker-defined components and up to 15 novel components). For the proteins attributed to the same clusters as the marker proteins, the class labels of markers were used. All the additional protein clusters were ordered by the number of proteins assigned, from the largest to the lowest, and labeled as 'PAM cluster  $n$ ', where  $n$  is the cluster rank. For the proteins assigned to the marker-defined classes, we defined the localisation probability ( $P_{loc.}$ ) as a product of the allocation probability ( $P_{allocation}$ ) and the complement of the outlier probability ( $P_{outlier}$ ):  $P_{loc.} = P_{allocation} \cdot (1 - P_{outlier})$ . For the proteins attributed to the discovered clusters, we defined the localisation probability ( $P_{loc.}$ ) as a product of the discovery ( $P_{discovery}$ ) probability and the complement of the outlier probability ( $P_{outlier}$ ):  $P_{loc.} = P_{discovery} \cdot (1 - P_{outlier})$ . We then applied a threshold of 99% to the localisation probability to retrieve the most confidently assigned proteins and labeled all other proteins as 'unassigned'. Only marker-defined classes and the additional clusters with at least 7 proteins assigned (PAM clusters 1-11) were considered further. GO terms result using curated and computed biological process with a p-values cut-off of 0.05 was generated using CryptoDB tool. Table S4 show the result of the top 10-fold enriched GO Terms per Novelty-TAGM cluster.

#### **Prediction of signal peptides and TM spans**

Protein sequence feature prediction was performed as described earlier<sup>40</sup>. SignalP 5.0<sup>26</sup> was used to predict cleaved signal peptides in *C. parvum* IOWA II annotated protein sequences with the

organism group set to Eukarya. Transmembrane <sup>TM</sup> span prediction was performed using TMHMM 2.0<sup>73</sup>. It is known that signal peptides can create TM false positives, so the results of TMHMM and SignalP analyses were compared and whenever a protein was predicted to have an N-terminal TM span and simultaneously a signal peptide the first TM span predicted by TMHMM was removed.

### Gene orthology analysis and ratio of non-synonymous to synonymous mutations

Inference of orthologous groups between *Cryptosporidium parvum* Iowa II, *Plasmodium falciparum* 3D7, and *Toxoplasma gondii* ME49 was performed with OrthoFinder version 2.5.2<sup>74</sup> using the default parameters. The annotated protein sequences were retrieved from CryptoDB.org, PlasmoDB.org, and ToxoDB.org, respectively. OrthoFinder assigned 15817 genes (73.3% of total) to 4757 orthogroups. Fifty percent of all genes were in orthogroups with 3 or more genes (G50 was 3) and were contained in the largest 2352 orthogroups (O50 was 2352). There were 1714 orthogroups with all species present and 1357 of these consisted entirely of single-copy genes. The patterns of conservation of the subcellular proteomes as defined by the Novelty-TAGM classification result were summarized by computing the number of proteins exclusive to *C. parvum* or conserved across Apicomplexa (homologues identified in *P. falciparum* and/or *T. gondii*). The ratio of non-synonymous to synonymous mutations (dN/dS) was obtained from CryptoDB using 118 genomes available and display per clusters as Box-Plots.

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-HA	Cell Signaling	3724S; RRID:AB_1549585
Anti-HA (for IEM)	Roche	Clone 3F10; RRID:AB_390918
Anti-cp23	LS Bio	LS-C137378-100; RRID:AB_10947007
Anti-gp40/gp900	Honorine Ward laboratory	Cevallos et al. <sup>36</sup>
Anti-Rabbit polyclonal	Fabio Tosini laboratory	Trasarti et al. <sup>75</sup>
Anti-CpPRP1	Furio Spano laboratory	Valentini et al. <sup>37</sup>
Anti-LDH	Guan Zhu laboratory	Zhang et al. <sup>44</sup>



Chemicals, Peptides, and Recombinant proteins		
TMT10plex™ Isobaric Label Reagent Set, 3x0.8mg	Thermo Fisher	90111
OptiPrep™ Density Gradient Medium	Sigma Aldrich	D1556
Sequencing Grade Modified Trypsin	Promega	V5111
Benzonase Nuclease HC	Merck Millipore	71205-3
cOmplete™ EDTA-free protease inhibitor cocktail	Sigma Aldrich	11873580001
Bacterial and virus strains		
<i>Escherichia coli</i> GC5	Genesee Scientific	42-653
Critical commercial assays		
Quick-DNA Fecal/Soil Microbe Kit	Zymo Research	D6010
Pierce BCA Protein Quantitation Assay	Thermo Fisher	23227
Deposited Data		
Raw LC-SPS-MS <sup>3</sup> data, peptide and protein proteomic identification and quantification results	This paper and PRIDE Archive	PRIDE ID: PXD038366
hyperLOPIT spatial proteome map of <i>C. parvum</i> extracellular sporozoite	This paper and CryptoDB	<a href="https://cryptodb.org">https://cryptodb.org</a>
Experimental models: Cell lines		
HCT8	ATCC	CCL-244
Experimental models: Organisms/strains		
<i>Cryptosporidium parvum</i>	Bunchgrass Farms, ID	IOWA strain
<i>ifn<math>\gamma</math></i> <sup>-/-</sup> mice	Jackson Laboratory	Strain 002287
Oligonucleotides		
Primers for cloning see table 5	This paper	N/A
Software and algorithms		
PRISM	GraphPad	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
ImageJ	ImageJ	<a href="https://imagej.net/Welcome">https://imagej.net/Welcome</a>
R	R Core Team, 2020	<a href="https://www.R-project.org/">https://www.R-project.org/</a>
Rstudio	Rstudio Team, 2020	<a href="http://www.rstudio.com/">http://www.rstudio.com/</a>
TAGM-MAP and TAGM-MCMC	Crook et al. <sup>27</sup>	<a href="https://doi.org/10.18129/B9.bioc.pRoloc">https://doi.org/10.18129/B9.bioc.pRoloc</a>
t-SNE	Van der Maaten and Hinton <sup>25</sup>	<a href="https://cran.r-project.org/package=Rtsne">https://cran.r-project.org/package=Rtsne</a>
HDBSCAN	Campello et al. <sup>72</sup>	<a href="https://hdbscan.readthedocs.io/en/latest/index.html">https://hdbscan.readthedocs.io/en/latest/index.html</a>
Orthofinder	Emms and Kelly <sup>74</sup>	<a href="https://github.com/davidemms/OrthoFinder">https://github.com/davidemms/OrthoFinder</a>
SignalP	Almagro Armenteros et al. <sup>26</sup>	<a href="https://services.healthtech.dtu.dk/service.php?SignalP-5.0">https://services.healthtech.dtu.dk/service.php?SignalP-5.0</a>
TMHMM	Krogh et al. <sup>73</sup>	<a href="https://services.healthtech.dtu.dk/service.php?TMHMM-2.0">https://services.healthtech.dtu.dk/service.php?TMHMM-2.0</a>

Proteome Discoverer	Thermo Fisher	<a href="https://www.thermofisher.com/ch/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html">https://www.thermofisher.com/ch/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html</a>
Other		
VVL-FITC	Vector	FL1231
Phalloidin-647	Thermo Fisher	A22287
Fluorescein-Phalloidin	Thermo Fisher	F432
AlexaFluo-568-NHS-Ester	Thermo Fisher	A20003
Hoechst	Thermo Fisher	H3570

879

880

## References

1. 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., et al. (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. 10.1016/S0140-6736(13)60844-2.
2. 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., et al. (2015). A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *Lancet Infect Dis* 15, 85-94. 10.1016/S1473-3099(14)70772-8.
3. Choy, R.K.M., and Huston, C.D. (2020). Cryptosporidiosis should be designated as a tropical disease by the US Food and Drug Administration. *PloS neglected tropical diseases in press*.
4. Striepen, B. (2013). Parasitic infections: Time to tackle cryptosporidiosis. *Nature* 503, 189-191.
5. Guérin, A., and Striepen, B. (2020). The Biology of the Intestinal Intracellular Parasite *Cryptosporidium*. *Cell Host Microbe* 28, 509-515. 10.1016/j.chom.2020.09.007.
6. Guérin, A., Roy, N.H., Kugler, E.M., Berry, L., Burkhardt, J.K., Shin, J.B., and Striepen, B. (2021). *Cryptosporidium* roptry effector protein ROP1 injected during invasion targets the host cytoskeletal modulator LMO7. *Cell Host Microbe* 29, 1407-1420 e1405. 10.1016/j.chom.2021.07.002.
7. English, E., Guérin, A., Tandel, J., and B., S. (2021). Live imaging of the *Cryptosporidium parvum* lifecycle reveals direct development of male and female gametes from type one meronts. *bioRxiv* 2021.12.07.471643.
8. Tandel, J., English, E.D., Sateriale, A., Gullicksrud, J.A., Beiting, D.P., Sullivan, M.C., Pinkston, B., and Striepen, B. (2019). Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium parvum*. *Nat Microbiol.* 10.1038/s41564-019-0539-x.
9. Elliott, D.A., and Clark, D.P. (2000). *Cryptosporidium parvum* induces host cell actin accumulation at the host-parasite interface. *Infect Immun* 68, 2315-2322.
10. Ostrovska, K., and Paperna, I. (1990). *Cryptosporidium* sp. of the starred lizard *Agama stellio*: ultrastructure and life cycle. *Parasitol Res* 76, 712-720.
11. Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., et al. (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441-445.
12. Zhu, G. (2007). Biochemistry. In *Cryptosporidium and Cryptosporidiosis*, R. Fayer, and L. Xiao, eds. (CRC Press), pp. 57-77.
13. Pawlowic, M.C., Somepalli, M., Sateriale, A., Herbert, G.T., Gibson, A.R., Cuny, G.D., Hedstrom, L., and Striepen, B. (2019). Genetic ablation of purine salvage in *Cryptosporidium parvum* reveals nucleotide uptake from the host cell. *Proc Natl Acad Sci U S A* 116, 21160-21165. 10.1073/pnas.1908239116.

14. Ben Chaabene, R., Lentini, G., and Soldati-Favre, D. (2021). Biogenesis and discharge of the rhoptries: Key organelles for entry and hijack of host cells by the Apicomplexa. *Mol Microbiol* *115*, 453-465. 10.1111/mmi.14674.
15. Frenal, K., Dubremetz, J.F., Lebrun, M., and Soldati-Favre, D. (2017). Gliding motility powers invasion and egress in Apicomplexa. *Nat Rev Microbiol* *15*, 645-660. 10.1038/nrmicro.2017.86.
16. Frickel, E.M., and Hunter, C.A. (2021). Lessons from Toxoplasma: Host responses that mediate parasite control and the microbial effectors that subvert them. *J Exp Med* *218*. 10.1084/jem.20201314.
17. Hakimi, M.A., Olias, P., and Sibley, L.D. (2017). Toxoplasma Effectors Targeting Host Signaling and Transcription. *Clinical Microbiology Reviews* *30*, 615-645. 10.1128/Cmr.00005-17.
18. Dubois, D.J., and Soldati-Favre, D. (2019). Biogenesis and secretion of micronemes in Toxoplasma gondii. *Cell Microbiol* *21*, e13018. 10.1111/cmi.13018.
19. Griffith, M.B., Pearce, C.S., and Heaslip, A.T. (2022). Dense granule biogenesis, secretion, and function in Toxoplasma gondii. *J Eukaryot Microbiol*, e12904. 10.1111/jeu.12904.
20. Mercier, C., and Cesbron-Delauw, M.F. (2015). Toxoplasma secretory granules: one population or more? *Trends Parasitol* *31*, 60-71. 10.1016/j.pt.2014.12.002.
21. Dumaine, J.E., Sateriale, A., Gibson, A.R., Reddy, A.G., Gullicksrud, J.A., Hunter, E.N., Clark, J.T., and Striepen, B. (2021). The enteric pathogen Cryptosporidium parvum exports proteins into the cytosol of the infected host cell. *Elife* *10*. 10.7554/eLife.70451.
22. Christoforou, A., Mulvey, C.M., Breckels, L.M., Geladaki, A., Hurrell, T., Hayward, P.C., Naake, T., Gatto, L., Viner, R., Martinez Arias, A., and Lilley, K.S. (2016). A draft map of the mouse pluripotent stem cell spatial proteome. *Nat Commun* *7*, 8992. 10.1038/ncomms9992.
23. Vinayak, S., Pawlowic, M.C., Sateriale, A., Brooks, C.F., Studstill, C.J., Bar-Peled, Y., Cipriano, M.J., and Striepen, B. (2015). Genetic modification of the diarrhoeal pathogen Cryptosporidium parvum. *Nature* *523*, 477-480. 10.1038/nature14651.
24. Trotter, M.W., Sadowski, P.G., Dunkley, T.P., Groen, A.J., and Lilley, K.S. (2010). Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions. *Proteomics* *10*, 4213-4219. 10.1002/pmic.201000359.
25. Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research* *9*.
26. Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* *37*, 420-423. 10.1038/s41587-019-0036-z.
27. Crook, O.M., Geladaki, A., Nightingale, D.J.H., Vennard, O.L., Lilley, K.S., Gatto, L., and Kirk, P.D.W. (2020). A semi-supervised Bayesian approach for simultaneous protein sub-cellular localisation assignment and novelty detection. *PLoS Comput Biol* *16*, e1008288. 10.1371/journal.pcbi.1008288.
28. Spano, F., Puri, C., Ranucci, L., Putignani, L., and 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.

966 29. Templeton, T.J., Lancto, C.A., Vigdorovich, V., Liu, C., London, N.R., Hadsall, K.Z., and  
967 Abrahamsen, M.S. (2004). The *Cryptosporidium* oocyst wall protein is a member of a  
968 multigene family and has a homolog in *Toxoplasma*. *Infection and Immunity* 72, 980-987.  
969 10.1128/iai.72.2.980-987.2004.

970 30. Samuelson, J., Bushkin, G.G., Chatterjee, A., and Robbins, P.W. (2013). Strategies To  
971 Discover the Structural Components of Cyst and Oocyst Walls. *Eukaryotic Cell* 12, 1578-  
972 1587. 10.1128/Ec.00213-13.

973 31. Jenwithisuk, R., Kangwanransan, N., Tachibana, M., Thongkukiatkul, A., Otsuki, H.,  
974 Sattabongkot, J., Tsuboi, T., Torii, M., and Ishino, T. (2018). Identification of a PH domain-  
975 containing protein which is localized to crystalloid bodies of *Plasmodium* ookinetes. *Malar*  
976 *J* 17, 466. 10.1186/s12936-018-2617-6.

977 32. Lemgruber, L., and Lupetti, P. (2012). Crystalloid body, refractile body and virus-like  
978 particles in Apicomplexa: what is in there? *Parasitology* 139, 285-293.  
979 10.1017/S0031182011002034.

980 33. Tetley, L., Brown, S.M., McDonald, V., and Coombs, G.H. (1998). Ultrastructural analysis  
981 of the sporozoite of *Cryptosporidium parvum*. *Microbiology* 144 ( Pt 12), 3249-3255.

982 34. Tremp, A.Z., Saeed, S., Sharma, V., Lasonder, E., and Dessens, J.T. (2020). *Plasmodium*  
983 *berghei* LAPs form an extended protein complex that facilitates crystalloid targeting and  
984 biogenesis. *J Proteomics* 227, 103925. 10.1016/j.jprot.2020.103925.

985 35. Tosini, F., Agnoli, A., Mele, R., Gomez Morales, M.A., and Pozio, E. (2004). A new modular  
986 protein of *Cryptosporidium parvum*, with ricin B and LCCL domains, expressed in the  
987 sporozoite invasive stage. *Mol Biochem Parasitol* 134, 137-147.  
988 10.1016/j.molbiopara.2003.11.014.

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

993 37. Valentini, E., Cherchi, S., Possenti, A., Dubremetz, J.F., Pozio, E., and Spano, F. (2012).  
994 Molecular characterisation of a *Cryptosporidium parvum* rhoptry protein candidate  
995 related to the rhoptry neck proteins TgRON1 of *Toxoplasma gondii* and PfASP of  
996 *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 183, 94-99.  
997 10.1016/j.molbiopara.2012.02.004.

998 38. Putignani, L., Possenti, A., Cherchi, S., Pozio, E., Crisanti, A., and Spano, F. (2008). The  
999 thrombospondin-related protein CpMIC1 (CpTSP8) belongs to the repertoire of  
1000 micronemal proteins of *Cryptosporidium parvum*. *Mol Biochem Parasitol* 157, 98-101.  
1001 10.1016/j.molbiopara.2007.09.004.

1002 39. Mauzy, M.J., Enomoto, S., Lancto, C.A., Abrahamsen, M.S., and Rutherford, M.S. (2012).  
1003 The *Cryptosporidium parvum* transcriptome during in vitro development. *PLoS One* 7,  
1004 e31715. 10.1371/journal.pone.0031715.

1005 40. Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O.M., Lassadi, I., Gupta, V., Tromer,  
1006 E., Mourier, T., Stevens, T.J., et al. (2020). A Comprehensive Subcellular Atlas of the  
1007 *Toxoplasma* Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell*  
1008 *Host Microbe* 28, 752-766 e759. 10.1016/j.chom.2020.09.011.

- 1009 41. Bonnin, A., Gut, J., Dubremetz, J.F., Nelson, R.G., and Camerlynck, P. (1995). Monoclonal  
1010 antibodies identify a subset of dense granules in *Cryptosporidium parvum* zoites and  
1011 gamonts. *J Eukaryot Microbiol* 42, 395-401. 10.1111/j.1550-7408.1995.tb01601.x.
- 1012 42. de Koning-Ward, T.F., Gilson, P.R., Boddey, J.A., Rug, M., Smith, B.J., Papenfuss, A.T.,  
1013 Sanders, P.R., Lundie, R.J., Maier, A.G., Cowman, A.F., and Crabb, B.S. (2009). A newly  
1014 discovered protein export machine in malaria parasites. *Nature* 459, 945-949.  
1015 10.1038/nature08104.
- 1016 43. El Bakkouri, M., Pow, A., Mulichak, A., Cheung, K.L., Artz, J.D., Amani, M., Fell, S., de  
1017 Koning-Ward, T.F., Goodman, C.D., McFadden, G.I., et al. (2010). The Clp chaperones and  
1018 proteases of the human malaria parasite *Plasmodium falciparum*. *J Mol Biol* 404, 456-477.  
1019 10.1016/j.jmb.2010.09.051.
- 1020 44. Zhang, H., Guo, F., and Zhu, G. (2015). *Cryptosporidium* Lactate Dehydrogenase Is  
1021 Associated with the Parasitophorous Vacuole Membrane and Is a Potential Target for  
1022 Developing Therapeutics. *PLoS Pathog* 11, e1005250. 10.1371/journal.ppat.1005250.
- 1023 45. Janouskovec, J., Paskerova, G.G., Mioliubova, T.S., Mikhailov, K.V., Birley, T., Aleoshin,  
1024 V.V., and Simdyanov, T.G. (2019). Apicomplexan-like parasites are polyphyletic and widely  
1025 but selectively dependent on cryptic plastid organelles. *Elife* 8. 10.7554/eLife.49662.
- 1026 46. Mathur, V., Kolisko, M., Hehenberger, E., Irwin, N.A.T., Leander, B.S., Kristmundsson, A.,  
1027 Freeman, M.A., and Keeling, P.J. (2019). Multiple Independent Origins of Apicomplexan-  
1028 Like Parasites. *Curr Biol* 29, 2936-2941 e2935. 10.1016/j.cub.2019.07.019.
- 1029 47. Boisard, J., Duvernois-Berthet, E., Duval, L., Schrevel, J., Guillou, L., Labat, A., Le Panse, S.,  
1030 Prensier, G., Ponger, L., and Florent, I. (2022). Marine gregarine genomes reveal the  
1031 breadth of apicomplexan diversity with a partially conserved glideosome machinery. *BMC*  
1032 *Genomics* 23, 485. 10.1186/s12864-022-08700-8.
- 1033 48. Mathur, V., Wakeman, K.C., and Keeling, P.J. (2021). Parallel functional reduction in the  
1034 mitochondria of apicomplexan parasites. *Curr Biol* 31, 2920-2928 e2924.  
1035 10.1016/j.cub.2021.04.028.
- 1036 49. Gubbels, M.J., and Duraisingh, M.T. (2012). Evolution of apicomplexan secretory  
1037 organelles. *Int J Parasitol* 42, 1071-1081. 10.1016/j.ijpara.2012.09.009.
- 1038 50. Valigurova, A., Hofmannova, L., Koudela, B., and Vavra, J. (2007). An ultrastructural  
1039 comparison of the attachment sites between *Gregarina steini* and *Cryptosporidium muris*.  
1040 *J Eukaryot Microbiol* 54, 495-510. 10.1111/j.1550-7408.2007.00291.x.
- 1041 51. Ebert, D., and Fields, P.D. (2020). Host-parasite co-evolution and its genomic signature.  
1042 *Nat Rev Genet* 21, 754-768. 10.1038/s41576-020-0269-1.
- 1043 52. Cova, M.M., Lamarque, M.H., and Lebrun, M. (2022). How Apicomplexa Parasites Secrete  
1044 and Build Their Invasion Machinery. *Annu Rev Microbiol* 76, 619-640. 10.1146/annurev-  
1045 micro-041320-021425.
- 1046 53. Forney, J.R., DeWald, D.B., Yang, S., Speer, C.A., and Healey, M.C. (1999). A role for host  
1047 phosphoinositide 3-kinase and cytoskeletal remodeling during *Cryptosporidium parvum*  
1048 infection. *Infect Immun* 67, 844-852.
- 1049 54. Wetzel, D.M., Schmidt, J., Kuhlenschmidt, M.S., Dubey, J.P., and Sibley, L.D. (2005). Gliding  
1050 motility leads to active cellular invasion by *Cryptosporidium parvum* sporozoites.  
1051 *Infection and Immunity* 73, 5379-5387. 10.1128/iai.73.9.5379-5387.2005.

55. Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S., Thiru, P., Saeij, J.P.J., Carruthers, V.B., Niles, J.C., and Lourido, S. (2016). A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes. *Cell* 166, 1423-1435 e1412. 10.1016/j.cell.2016.08.019.
56. Gao, X., Yin, J., Wang, D., Li, X., Zhang, Y., Wang, C., Zhang, Y., and Zhu, G. (2021). Discovery of New Microneme Proteins in *Cryptosporidium parvum* and Implication of the Roles of a Rhomboid Membrane Protein (CpROM1) in Host-Parasite Interaction. *Front Vet Sci* 8, 778560. 10.3389/fvets.2021.778560.
57. Deng, M., Templeton, T.J., London, N.R., Bauer, C., Schroeder, A.A., and Abrahamsen, M.S. (2002). *Cryptosporidium parvum* genes containing thrombospondin type 1 domains. *Infect Immun* 70, 6987-6995. 10.1128/IAI.70.12.6987-6995.2002.
58. Kremer, K., Kamin, D., Rittweger, E., Wilkes, J., Flammer, H., Mahler, S., Heng, J., Tonkin, C.J., Langsley, G., Hell, S.W., et al. (2013). An overexpression screen of *Toxoplasma gondii* Rab-GTPases reveals distinct transport routes to the micronemes. *PLoS Pathog* 9, e1003213. 10.1371/journal.ppat.1003213.
59. Morlon-Guyot, J., El Hajj, H., Martin, K., Fois, A., Carrillo, A., Berry, L., Burchmore, R., Meissner, M., Lebrun, M., and Daher, W. (2018). A proteomic analysis unravels novel CORVET and HOPS proteins involved in *Toxoplasma gondii* secretory organelles biogenesis. *Cell Microbiol* 20, e12870. 10.1111/cmi.12870.
60. Cui, H., Xu, R., Li, Y., Guo, Y., Zhang, Z., Xiao, L., Feng, Y., and Li, N. (2022). Characterization of Dense Granule Metalloproteinase INS-16 in *Cryptosporidium parvum*. *Int J Mol Sci* 23. 10.3390/ijms23147617.
61. Beck, J.R., Muralidharan, V., Oksman, A., and Goldberg, D.E. (2014). PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. *Nature* 511, 592-595. 10.1038/nature13574.
62. Bougdour, A., Durandau, E., Brenier-Pinchart, M.P., Ortet, P., Barakat, M., Kieffer, S., Curt-Varesano, A., Curt-Bertini, R.L., Bastien, O., Coute, Y., et al. (2013). Host Cell Subversion by *Toxoplasma* GRA16, an Exported Dense Granule Protein that Targets the Host Cell Nucleus and Alters Gene Expression. *Cell Host & Microbe* 13, 489-500. 10.1016/j.chom.2013.03.002.
63. Mair, G.R., Lasonder, E., Garver, L.S., Franke-Fayard, B.M.D., Carret, C.K., Wiegant, J.C.A.G., Dirks, R.W., Dimopoulos, G., Janse, C.J., and Waters, A.P. (2010). Universal Features of Post-Transcriptional Gene Regulation Are Critical for *Plasmodium* Zygote Development. *Plos Pathogens* 6. ARTN e1000767 10.1371/journal.ppat.1000767.
64. Wu, X., Yan, R., Cao, P., Qian, H., and Yan, N. (2022). Structural advances in sterol-sensing domain-containing proteins. *Trends Biochem Sci* 47, 289-300. 10.1016/j.tibs.2021.12.005.
65. Tandel, J., Walzer, K.A., Byerly, J.H., Pinckston, B., Beiting, D.P., and Striepen, B. (2023). Genetic ablation of a female specific Apetala 2 transcription factor blocks oocyst shedding in *Cryptosporidium parvum*. *mBio*, 2022.2011.2023.517783. 10.1128/mbio.03261-22.
66. Gut, J., and Nelson, R.G. (1999). *Cryptosporidium parvum*: synchronized excystation in vitro and evaluation of sporozoite infectivity with a new lectin-based assay. *J Eukaryot Microbiol* 46, 56S-57S.

67. Sateriale, A., Pawlowic, M., Vinayak, S., Brooks, C., and Striepen, B. (2020). Genetic Manipulation of *Cryptosporidium parvum* with CRISPR/Cas9. *Methods Mol Biol* 2052, 219-228. 10.1007/978-1-4939-9748-0\_13.
68. Dos Santos Pacheco, N., and Soldati-Favre, D. (2021). Coupling Auxin-Inducible Degron System with Ultrastructure Expansion Microscopy to Accelerate the Discovery of Gene Function in *Toxoplasma gondii*. *Methods Mol Biol* 2369, 121-137. 10.1007/978-1-0716-1681-9\_8.
69. Mulvey, C.M., Breckels, L.M., Geladaki, A., Britovsek, N.K., Nightingale, D.J.H., Christoforou, A., Elzek, M., Deery, M.J., Gatto, L., and Lilley, K.S. (2017). Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nature Protocols* 12, 1110-1135. 10.1038/nprot.2017.026.
70. Yang, F., Shen, Y., Camp, D.G., 2nd, and Smith, R.D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev Proteomics* 9, 129-134. 10.1586/epr.12.15.
71. Amos, B., Aurecochea, C., Barba, M., Barreto, A., Basenko, Evelina Y., Bazant, W., Belnap, R., Blevins, A.S., Böhme, U., Brestelli, J., et al. (2021). VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research* 50, D898-D911. 10.1093/nar/gkab929.
72. Campello, R.J.G.B., Moulavi, D., and Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. held in Berlin, Heidelberg, (Springer Berlin Heidelberg), pp. 160-172.
73. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305, 567-580. 10.1006/jmbi.2000.4315.
74. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, 238. 10.1186/s13059-019-1832-y.
75. Trasarti, E., Pizzi, E., Pozio, E., and Tosini, F. (2007). The immunological selection of recombinant peptides from *Cryptosporidium parvum* reveals 14 proteins expressed at the sporozoite stage, 7 of which are conserved in other apicomplexa. *Mol Biochem Parasitol* 152, 159-169. 10.1016/j.molbiopara.2006.12.010.



















