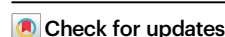


SKSR1 identified as key virulence factor in *Cryptosporidium* by genetic crossing

Received: 28 February 2024

Accepted: 14 May 2025

Published online: 20 May 2025



Wei He^{1,2,4}, Lianbei Sun^{1,4}, Tianyi Hou^{1,4}, Zuwei Yang¹, Fuxian Yang¹, Shengchen Zhang¹, Tianpeng Wang^{1,2}, Xinran Wang¹, Na Li¹, Yaqiong Guo¹, L. David Sibley³✉, Yaoyu Feng¹✉ & Lihua Xiao¹✉

Cryptosporidium is a major cause of severe diarrhea. Although *Cryptosporidium* isolates exhibit significant differences in infectivity and virulence, the genetic determinants for these traits are not clear. In this study, we use classical genetics to cross two *Cryptosporidium parvum* isolates of different virulence and use bulk segregant analysis of whole-genome sequences from the progeny to identify quantitative trait loci (QTL) associated with *Cryptosporidium* infectivity and virulence. Of the 23 genes in three QTL, two have loss-of-function mutations in the low-virulence isolates, including the *SKSR1* gene encoding a variant secretory protein. Deletion of the *SKSR1* gene or expression of the frame-shifted sequence reduces the pathogenicity of the virulent isolate. *SKSR1* is expressed in small granules and secreted into the parasite-host interface during invasion. These results demonstrate that *SKSR1* is an important virulence factor in *Cryptosporidium*, and suggest that the extended *SKSR* protein family, encoded by clusters of subtelomeric genes, may contribute to pathogenesis.

Cryptosporidiosis is a leading cause of severe diarrhea, leading to death and malnutrition in many children in low- and middle-income countries¹. It is also a major cause of food- and waterborne disease in high-income countries². The disease is particularly severe in young children and immunocompromised individuals, leading to malnutrition and significant mortality³. The pathogenesis of cryptosporidiosis is poorly understood.

Cryptosporidium hominis and *Cryptosporidium parvum* are the dominant species responsible for more than 90% of human cases of cryptosporidiosis⁴. The two differ significantly in their host range, with the former being predominantly an anthroponotic species in low- and middle-income countries and the latter being a zoonotic species and the dominant species in both humans and calves in high-income countries⁴. The two species share ~96% sequence identity in whole genome sequences⁵. Due to the lack of convenient culture and animal

models for *C. hominis*, most biological studies of *Cryptosporidium* spp. have been conducted with *C. parvum*⁶.

C. parvum isolates vary widely in infectivity and virulence. In experimental infections of healthy adults, three *C. parvum* isolates differed in half infectious doses (ID₅₀), infection intensity, attack rates, and duration of diarrhea⁷. Differences in virulence among *C. parvum* isolates have also been observed in experimental infections of mice and young livestock^{8–10}. For example, at the subtype level, IIaA15G2R1 has become the dominant *C. parvum* in both humans and animals in most high-income countries^{4,11}. However, the genetic determinants of *Cryptosporidium* virulence and infectivity are not clear, although the results from comparative genomic analysis have provided some clues¹⁰.

In *Plasmodium* spp. and *Toxoplasma gondii*, linkage mapping of genetic crosses of isolates has been effective in studies of virulence

¹State Key Laboratory for Animal Disease Control and Prevention, Center for Emerging and Zoonotic Diseases, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China. ²School of Biology and Agriculture, Shaoguan University, Shaoguan, China. ³Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA. ⁴These authors contributed equally: Wei He, Lianbei Sun, Tianyi Hou.

✉ e-mail: sibley@wustl.edu; yyfeng@scau.edu.cn; lxiao1961@gmail.com

factors^{12,13}. It has led to the identification of several rhoptry proteins as major determinants of virulence in *T. gondii*¹⁴ and point mutations in the erythrocyte binding like protein that impact red blood cell invasion and virulence in *Plasmodium yoelii*¹⁵. However, traditional linkage analysis requires the comparative analysis of numerous recombinant progeny. In recent years, a more efficient linkage mapping technique, bulk segregant analysis (BSA), has been used to characterize genetic crosses of *Plasmodium falciparum* isolates^{16,17}. It measures changes in allele frequencies in cross progeny under different selection pressures, allowing the identification of single nucleotide polymorphisms (SNPs) associated with specific phenotypic traits¹⁸.

In this report, we describe the identification of genes underlying the virulence differences using BSA of progeny from genetic crosses of two *C. parvum* isolates. The role of candidate genes in infection and virulence is validated using gene depletion and replacement techniques. The results show that the small granule (SG) protein SKSR1 is a key virulence factor in *C. parvum*.

Results

Coinfection of *C. parvum* isolates of different virulence produces recombinant progeny

We used two *C. parvum* isolates of different subtypes of the 60 kDa glycoprotein (GP60) gene in our cross studies, including IIdA20G1-HLJ (HLJ) and IIdA19G1-GD (GD)¹⁰. They differ in infectivity ($ID_{50} = 0.6$ and 5.1 oocysts, respectively) and infection pattern (HLJ produces over 10^7 oocysts per gram of feces, OPG, for more than 8 weeks compared to a peak OPG of 10^6 for less than 2 weeks for GD) in interferon- γ knockout (GKO) mice. HLJ is highly virulent in this animal model, causing arched backs, lethargy, weight loss, and significant mortality. In contrast, GD is avirulent in GKO mice¹⁰.

To create a genetic cross between the virulent HLJ and the avirulent GD, we endogenously tagged the two isolates with different fluorescent proteins using CRISPR/Cas9 as previously described¹⁹. In the HLJ isolate, we replaced the uracil phosphoribosyltransferase (*UPRT*) gene on chromosome 1 with a sequence encoding the tdTomato protein. Similarly, in the GD isolate, we replaced the thymidine kinase (*TK*) gene on chromosome 5 with a sequence encoding the mNeonGreen protein (Supplementary Fig. 1a). Fluorescence microscopy of purified oocysts showed that all (100/100) oocysts from HLJ and GD contained mNeonGreen (green) and tdTomato (red) fluorescent proteins, respectively (Supplementary Fig. 1b). PCR analysis showed correct integration of the replacement cassette (Supplementary Fig. 1c).

We examined the difference in pathogenicity between fluorescently tagged GD (GD-mNeonGreen) and HLJ (HLJ-tdTomato) lines. When treated with paromomycin (PRM) after infection with transgenic lines carrying the neomycin resistance (neo) gene, the HLJ-tdTomato line showed higher oocyst shedding (Supplementary Fig. 1d), caused significant clinical signs of infection (Supplementary Fig. 1e), and resulted in 100% mortality of the GKO mice (Supplementary Fig. 1f). In contrast, GD-mNeonGreen is avirulent in GKO mice (Supplementary Fig. 1e, f).

We performed four genetic crosses of GD-mNeonGreen and HLJ-tdTomato in GKO mice (Fig. 1a). Prior to infection, GKO mice were pretreated with antibiotics to increase the infection rate and the intensity of oocyst shedding of the parasite²⁰. Co-infection of the two isolates resulted in the generation of recombinant oocysts expressing both mNeonGreen and tdTomato fluorescent proteins. We purified the progeny oocysts from the intestinal contents of infected mice on day 4 post infection (DPI) (Fig. 1b). In the first cross, we found 19% colorless oocysts in the cross progeny (Supplementary Fig. 2a), possibly due to the presence of some colorless sporozoites in newly tagged parental oocysts. Subsequent crosses of parental oocysts from subsequent passages showed that approximately 32.6% of the purified oocysts showed both tdTomato and mNeonGreen fluorescence (yellow in

merged images), while the remaining oocysts showed predominantly tdTomato fluorescence, with a small number of oocysts showing mNeonGreen fluorescence or being colorless (Fig. 1c). We enriched and collected oocysts expressing two fluorescent proteins by flow cytometric sorting (Supplementary Fig. 2b). Microscopic examination of 100 sorted oocysts and sporozoites showed that all oocysts and sporozoites had two fluorescent proteins (Supplementary Fig. 2c), some of which may be residues of the parental protein²¹.

Prior to analysis, we generated a chromosome-level HLJ genome using PacBio in combination with Illumina sequencing. The whole genome assembly together with full annotation has been deposited in DDBJ/ENA/GenBank databases under accession number JBJGDY000000000. The new HLJ genome is comparable in completeness to several other *C. parvum* genomes sequenced using third-generation sequencing technology²², with 13 telomeric ends. Three missing telomeric regions are located at the 5' end of chromosome 1 and the 3' ends of chromosomes 7 and 8 (Supplementary Fig. 2d). There are 1065 SNP differences distributed between the HLJ and GD genomes (Supplementary Fig. 2e). Based on PCR analysis of polymorphic loci, we confirmed the presence of mixed sequence types in oocysts (DPI 4 progeny pool) expressing two fluorescent proteins (Supplementary Fig. 2f), indicating the presence of genomic sequences from both HLJ and GD. Whole genome sequencing (WGS) was performed on 12 individual oocysts, yielding good sequence data from 11 of them. Allele frequencies and mapping depth of *TK* and *UPRT* locus analysis showed that the single oocyst genomes were all heterozygous (Supplementary Fig. 3a and Supplementary Fig. 4), confirming the recombinant nature of the progeny as in previous studies¹⁹. Mosaic sequences at polymorphic loci were found in comparative genomic analyses of sequencing data from two single oocysts (Fig. 1d), suggesting that the yellow oocysts contained pools of recombinant progeny of HLJ and GD. Frequent crossovers of the HLJ and GD sequences were observed along the eight chromosomes of the 11 single-oocyst genomes (Supplementary Fig. 3a). However, due to the heterozygosity of the single-oocyst genomes, we used allele frequency differences >0.8 between neighbouring SNP loci to identify sequence crossovers. The lack of a consistent pattern of genetic recombination across the eight chromosomes suggests that the current threshold setting may underestimate the number of crossover events, resulting in some recombination not being detected (Supplementary Fig. 3b).

Progeny BSA identifies quantitative trait loci associated with parasite growth

To identify genes associated with parasite growth, we performed BSA of the progeny from genetic crosses of the HLJ and GD isolates. Fecal samples were collected from GKO mice every 6 days after infection with the progeny pool for oocyst purification and WGS analysis (Fig. 2a). Microscopic examination of oocysts collected at different time points revealed that although only yellow oocysts were used to inoculate mice and PRM was used to maintain fluorescence tagged parasites, there was a rapid decrease in the proportion of yellow oocysts during the early course of infection. This change was accompanied by the appearance of tdTomato, mNeonGreen fluorescence, and colorless oocysts (Fig. 2b, c). G-statistics analysis of the WGS data was used to identify SNPs enriched in recombinant *C. parvum* progeny over the course of infection. Calculation of G' values for different batches and replicates resulted in a threshold value of 91.4 for the identification of quantitative trait loci (QTL). Gradual increases in G' values were observed in the subtelomeric regions of chromosomes 1, 7, and 8 as infection progressed, suggesting an increased frequency of some alleles in these regions (Fig. 2d and Supplementary Fig. 5a).

In addition, the delta SNP index in these regions was close to -1 , suggesting an enrichment of HLJ alleles (Fig. 2e and Supplementary Tables 1–4). The SNP index is another algorithm used to localize QTL in BSA²³. It creates a segregating population from the progeny, selects

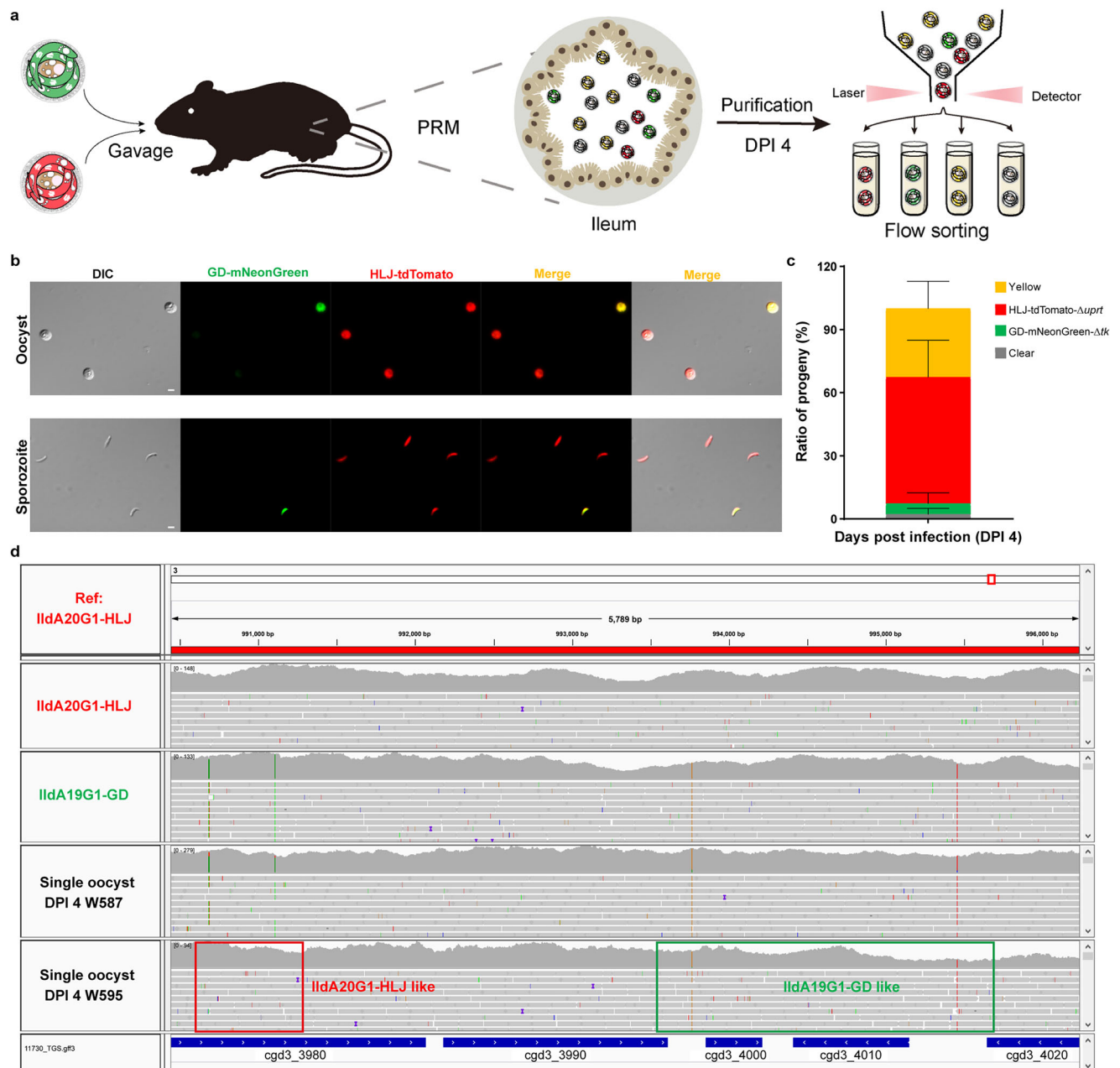


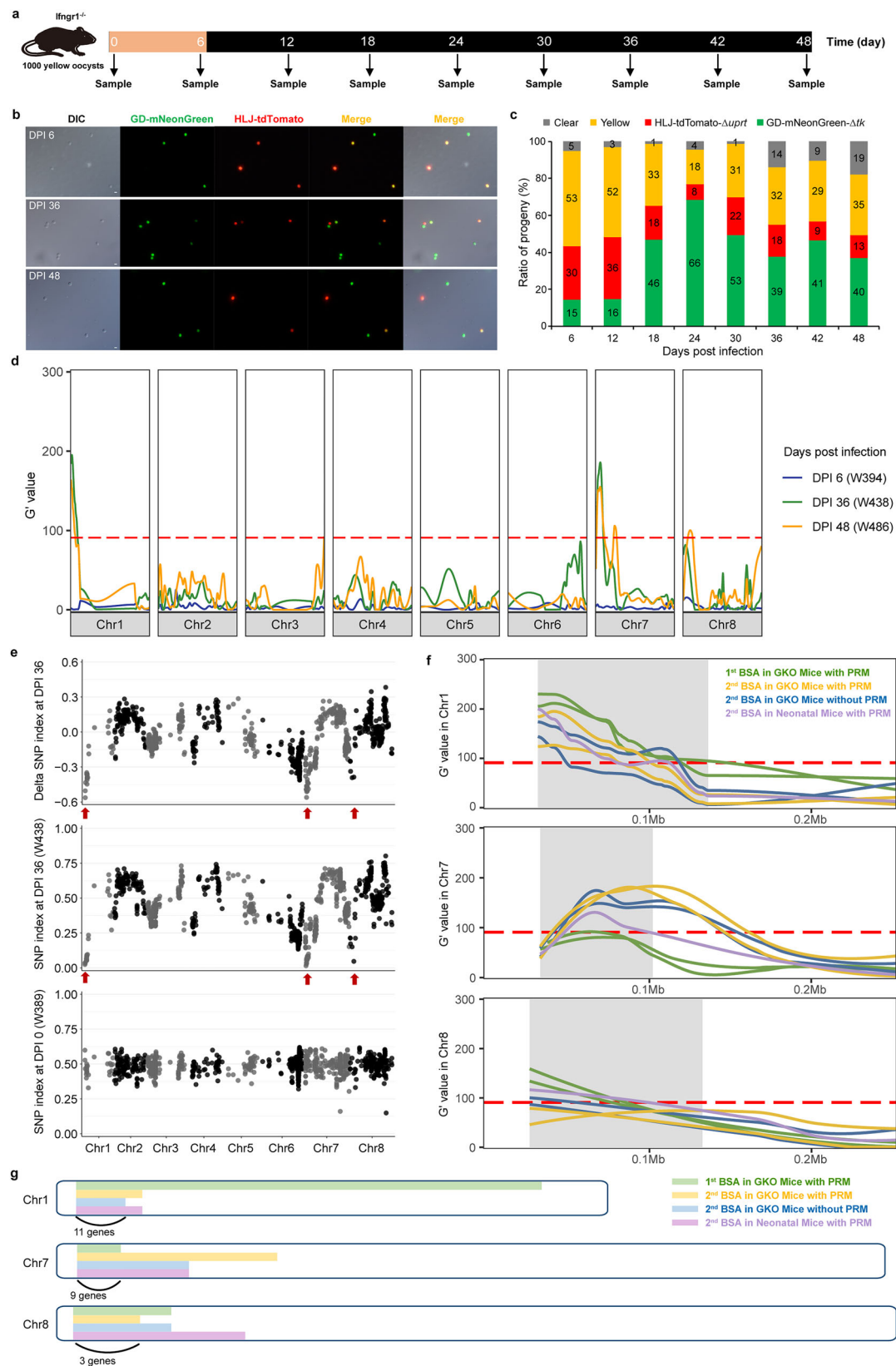
Fig. 1 | Genetic crosses between virulent HLJ and avirulent GD strains of *Cryptosporidium parvum* in vivo. **a** Diagram of genetic crosses of HLJ and GD in mice and flow cytometric sorting of the progeny. **b** Image of oocysts harvested from the intestinal mucosa 4 days after coinfection of oocysts tagged with different fluorescent proteins and the sporozoites excysted from them. Scale bars = 2 μ m. **c** Ratio of oocysts of each color in intestinal content of mice 4 days after coinfection, as determined by microscopy of samples (mean \pm SD) from three independent

experiments (the 2nd, 3rd, and 4th genetic crosses). **d** Confirmation of genetic recombination in single oocysts of progeny by read mapping of whole genome sequences. W587 and W595 represent the IDs of the sequenced single oocyst genomes. Two polymorphic sites relative to the parental sequences on the genomes in oocyst W595 are marked, indicating crossover of sequence types in this region. Source data are provided as a Source Data file.

individuals with extreme phenotypes to form mixed pools, and then genotypes these pools for SNPs. The allele frequencies within each mixed pool are analyzed and compared to one of the parental genotypes. The SNP index at a given locus is determined by the proportion of genotypes that differ from the parent. A larger delta SNP index (the difference in SNP index between two mixed pools) indicates a stronger association of the SNP with the trait of interest.

In total, we performed four infection studies with two genetic crosses in GKO and neonatal mice with and without the use of PRM (Supplementary Figs. 6–8). BSA of WGS data collected at different time

points in the infection courses identified similar enrichment of HLJ alleles in three regions on chromosomes 1, 7, and 8 (Fig. 2f). The GKO mouse is an appropriate model for *Cryptosporidium* infection, as our isolates show differences in virulence in this model¹⁰. However, it uses immunodeficient mice, which may support *Cryptosporidium* growth differently than immunocompetent mice. Therefore, we used neonatal C57 BL/6 mice to validate the BSA results obtained from GKO mice. Interestingly, the same three regions were identified as QTL associated with differences in virulence between the two *C. parvum* isolates. However, the use of no PRM selection during progeny pool infection of



GKO mice identified an additional enrichment of HLJ alleles on chromosome 5 (Supplementary Fig. 7), which was not present in the other three BSA experiments. In total, three QTL were identified in all four BSA studies, encompassing and contained 23 genes on chromosomes 1, 7, and 8. They were associated with the difference in parasite growth between HLJ and GD (Fig. 2g and Supplementary Fig. 5b).

Major candidate QTL genes have deleterious nucleotide insertion or substitution in the coding region

The 23 genes in the three QTL regions were closely examined for sequence differences between HLJ and GD and for the level of effect of these sequence variations. Most of these genes had only 1–2 SNPs in the coding regions, which were judged by SnpEff to have a low to

Fig. 2 | Screening of genes potentially associated with *Cryptosporidium* growth using bulk segregant analysis (BSA). **a** Diagram showing the design of the BSA study in GKO mice and the time of sample collection for WGS analysis. **b** Image of purified oocysts at different time points of the second BSA infection with PRM. Images of purified oocysts at DPI 6, DPI 36, and DPI 48 are shown. Scale bars = 5 μ m. **c** Ratio of oocysts of different colors at different time points of infection as determined by microscopy. **d** Distribution of G' values of *Cryptosporidium* genomes at DPI 6, DPI 36, and DPI 48 of the BSA study. **e** Distribution of the SNP index of *Cryptosporidium* genomes collected at DPI 0 (progeny pool) and DPI 36. The delta SNP indices in the subtelomeric regions of chromosomes 1, 7, and 8 are close to -1, indicating the enrichment of HLJ alleles ($n = 3$ mice; data from one representative replicate at DPI 36 are shown). The first row is the delta SNP index of the DPI 36 (W438) and DPI 0 (W389) genomes, which is the result after removing the

background noise, i.e., the pre-screening interference (the SNP index of DPI 0). The second and third rows are the SNP indices of the DPI 36 and DPI 0 genomes, respectively. W389 and W438 are the IDs of the genomes. Data in (b–e) are from the second BSA infection study with the progeny pool from the second genetic cross. Altogether, four BSA studies were performed using progeny from two genetic crosses. **f** Identification of three regions on chromosomes 1, 7, and 8 as the QTLs underlying the growth differences between HLJ and GD, based on a 95% confidence interval of data from four BSA studies. The shaded areas are overlapping QTL regions obtained from different BSA studies after Venn analysis of the data. **g** Identification of 23 candidate genes associated with growth using physical and Venn plots of data from four BSA studies. Source data are provided as a Source Data file.

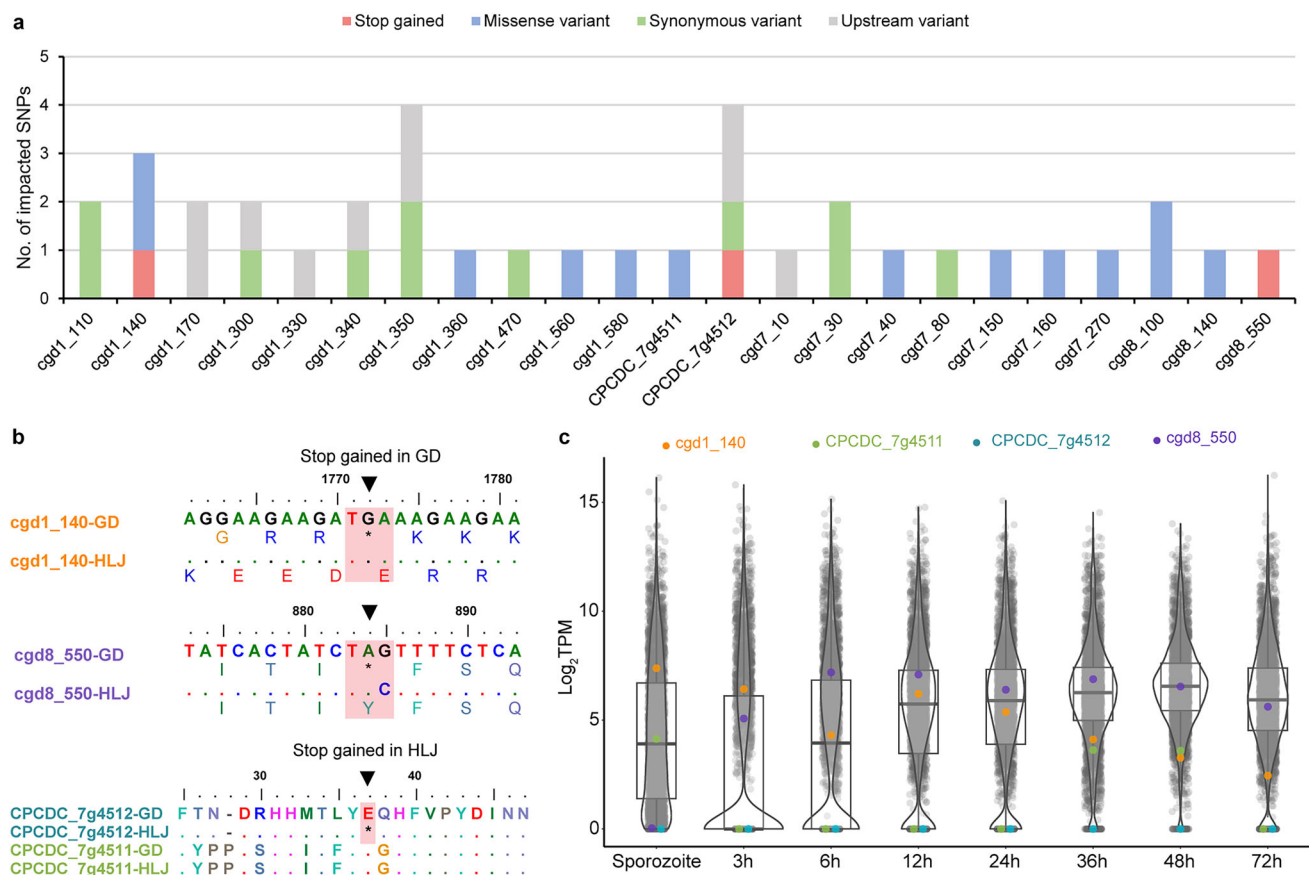


Fig. 3 | Sequence and expression characteristics of candidate *Cryptosporidium* growth-associated genes. **a** Distribution of different types of SNPs in 23 genes. **b** Alignment of the partial nucleotide and amino acid sequences of 3 candidate virulence genes. Insertion of base A in *cgd1_140* results in premature termination of *SKSRI*-GD transcription (the arrowhead). The other two arrowheads indicate the position of the base mutation causing a termination codon in *cgd8_550*-GD and *CPCDC_7g4512*-HLJ. **c** Box and violin plots of the expression of potential virulence

genes. The *CPCDC_7g4512* gene is not expressed in all life cycle stages examined. Each dot represents one gene ($n = 4$ for the RNA-seq analysis of each culture point). The bounds and horizontal bar of the box in each plot represent quartile and median expression level, while the density curves in the violin plot shows the distribution of gene expression levels. Source data are provided as a Source Data file, and the RNA-seq data are available at NCBI under BioProject No. PRJNA1011005.

moderate effect on the function of the proteins they encode (Supplementary Table 5). However, three genes had nucleotide insertions or substitutions that resulted in the formation of stop codons that are likely to significantly affect the function of the proteins they encode (Fig. 3a). These included *cgd1_140* (encoding the protein *SKSRI*), a newly predicted paralog (gene ID: *CPCDC_7g4512*) of the *cgd5_4510* gene (encoding a hypothetical protein; gene ID: *CPCDC_7g4511*) on chromosome 7 of the *Ila-IOWA 43IA8* genome²², and *cgd8_550* (encoding a hypothetical protein). Allele frequency analyses of these genes showed a rapid enrichment of the HLJ sequence during the BSA infection (Supplementary Fig. 9).

Among these candidate QTL genes, *SKSRI*-GD had an A-base insertion after nucleotide 1753 compared to *SKSRI*-HLJ, resulting in premature termination of *SKSRI*-GD translation (Fig. 3b). In contrast, the *CPCDC_7g4512* gene in virulent HLJ had a G to T substitution, resulting in the formation of a stop codon 51 bp after the start codon. The *CPCDC_7g4512* gene had high sequence identity to *CPCDC_7g4511*, which differed by one nucleotide between HLJ and GD and had no stop codon-generating nucleotide substitution (Supplementary Table 5 and Fig. 3b). Similarly, there was only one nucleotide difference between HLJ and GD in the *cgd8_550* gene resulting in the formation of a termination codon in *cgd8_550*-GD (Fig. 3b).

Since CPCDC_7g4512 is a recently predicted gene, we analyzed its expression level. RNA-seq analysis of the transcriptome across all developmental stages of HLJ in HCT-8 cultures and sporozoites showed no evidence of CPCDC_7g4512 expression. In contrast, its paralog CPCDC_7g4511 showed modest expression in sporozoites and at 36 and 48 h in culture (Fig. 3c). Although the *cgd8_550* gene was not expressed in sporozoites, its expression gradually increased in HCT-8 culture and remained at high levels during 6–48 h. In contrast, the expression of the *SKSR1* gene was high in sporozoites and at 3, 12, and 24 h in culture (Fig. 3c). The appearance of the stop codon in the virulent HLJ isolate and the absence of transcription suggested that CPCDC_7g4512 was unlikely to be involved in the virulence differences between HLJ and GD. Therefore, our subsequent validation studies had focused on the *SKSR1* and *cgd8_550* genes.

Protein encoded by *cgd8_550* is not a virulence factor

We investigated the expression and function of the hypothetical protein encoded by the *cgd8_550* gene. We generated *cgd8_550*-tagged (*cgd8_550*-3HA), C-to-G mutant (*cgd8_550*^{m(C-G)}-3HA), and gene deletion (Δ *cgd8_550*) lines of the virulent HLJ strain using CRISPR/Cas9 (Supplementary Figs. 10a, c and 11a). PCR and fluorescence analysis confirmed the successful deletion of the *cgd8_550* gene in the Δ *cgd8_550* line, with oocysts from this line being green due to the replacement of the gene with the mNeonGreen sequence (Supplementary Fig. 10e). Immunofluorescence analysis (IFA) showed no *cgd8_550* expression in sporozoites and high expression near the nuclei in meronts (Supplementary Fig. 12a). We investigated the role of *cgd8_550* in *Cryptosporidium* growth and host pathogenicity using HCT-8 culture and GKO mouse models. Deletion of the *cgd8_550* gene did not significantly affect *C. parvum* growth in vitro and in vivo (Supplementary Fig. 12b, c), and GKO mice infected with the *cgd8_550*-3HA and Δ *cgd8_550* lines had similar weight gain and survival (Supplementary Fig. 12d, e). Because the *cgd8_550* gene was not associated with the high virulence of the HLJ strain, we did not subsequently validate the effect of the C-to-G substitution (*cgd8_550*^{m(C-G)}-3HA) on virulence.

SKSR1 is a small granule protein encoded by a variant subtelomeric gene

The *cgd1_140* gene is a subtelomeric gene on chromosome 1, encoding the *Cryptosporidium*-specific secretory protein SKSR1. Comparison of *cgd1_140* sequences from *C. parvum* showed that the gene is polymorphic, with 21–23 SNPs between IIa and IIc subtypes in the 2892 bp coding sequence. Within the IIc subtypes, the A insert between NT1743 and NT1744 was seen in all IIcA19G1 isolates, but was absent from the sequences of other IIc isolates. This insert was also not detected in the IIa and IIc isolates analyzed. The sequence polymorphism resulted in the formation of two subclades within the IIc cluster and three subclades within the IIa cluster (Supplementary Fig. 13a).

We generated SKSR1-tagged (SKSR1-3HA), A-base inserted mutant (SKSR1^{m(+A)}-3HA), and SKSR1 deletion (Δ *sksr1*) lines of the HLJ isolate using CRISPR/Cas9 (Supplementary Figs. 10b, d and 11b). Fecal luciferase monitoring of mice infected with these lines and PCR analysis of purified oocysts confirmed successful integration of the replacement templates (Supplementary Figs. 10d and 13b). SKSR1 expression was observed near the nucleus of sporozoites and merozoites in a pattern distinct from canonical dense granules (Fig. 4b and Supplementary Fig. 13c). In the SKSR1^{m(+A)}-3HA line, due to the insertion of an A in the *SKSR1* gene, there was no detection of SKSR1 in IFA analysis of all stages examined (Supplementary Fig. 13c). This was confirmed by Western blot analysis, showing the expression of SKSR1-3HA but not SKSR1^{m(+A)}-3HA (Supplementary Fig. 13d). Fluorescence microscopy further confirmed the successful deletion of the *SKSR1* gene in the Δ *sksr1* line, as oocysts of the mutant line all contained the replacement tdTomato protein (Supplementary Fig. 10f). This finding was supported by read

mapping of the WGS data from the Δ *sksr1* line (Supplementary Fig. 13e).

We performed immunoelectron microscopy (IEM) of the ileal tissue from mice infected with the SKSR1-3HA line and SG1-3HA lines, and the results showed the accumulation of gold particles for SKSR1 in small vesicles near the nucleus that matched the location, shape, and size of SG1-3HA-positive vesicles (mean \pm SD = 113.9 \pm 17.3 and 118.3 \pm 28.2 nm, respectively; Supplementary Fig. 14a, b) instead of the dense granules (mean = 202 \pm 22.7 nm)²⁴. In addition, ultrastructure expansion microscopy (U-ExM) confirmed SKSR1 expression in SG around the nucleus, away from dense granules (Fig. 4b). Consistent with the gene expression pattern of SG proteins²⁴, *SKSR1* expression was detected in sporozoites as well as at most time points of the asexual development of *C. parvum* (Fig. 3d).

SKSR1 translocates to the parasite-host interface during invasion

We used immunofluorescence to follow SKSR1 secretion during sporozoite invasion. The results showed that SKSR1 was not secreted during sporozoite gliding. SKSR1 remained in the SG during host cell invasion until the parasite was completely engulfed by the host cell membrane (Supplementary Fig. 15a). Once a cup-like structure was formed, SKSR1 was detected at the apical side of the parasite (Fig. 4c). The secretion of SKSR1 at the apical region of sporozoites was confirmed by generating 3D rendered Z-stacks (Fig. 4d). In addition, SKSR1 was detected at low levels at the host-parasite interface by IEM (Fig. 4e). In the intracellular stages, confocal microscopy revealed the aggregation of SKSR1 on the parasite surface (Supplementary Fig. 14c). SKSR1 was further localized to the parasite-host interface by U-ExM and IEM (Supplementary Fig. 14d, e). Taken together, these results suggest that SKSR1 is transported from the SG to the host-parasite interface during the late phase of the invasion process (Fig. 4f). This pattern is similar to the secretion of SG1 during host cell invasion²⁴.

We used an attachment-invasion assay to determine whether SKSR1 is involved in host cell invasion by *C. parvum* sporozoites²⁵. HCT-8 cells were infected with sporozoites from the SKSR1-HA, SKSR1^{m(+A)}-3HA and Δ *sksr1* lines for 2.5 h. After the removal of unattached sporozoites by washing, samples were fixed and probed with rabbit anti-*C. parvum* antigens (anti-Cp) followed by goat anti-rabbit IgG-Alexa Fluor 488. The specimens were then permeabilized and reprobed with rabbit anti-Cp and goat anti-rabbit IgG-Alexa Fluor 594. Attached but non-invasive parasites were stained yellow, reflecting both detection steps, while intracellular parasites were stained red only. The number of intracellular and extracellular parasites was quantified by image analysis, and data from SKSR1^{m(+A)}-3HA, and Δ *sksr1*-infected cultures were normalized to the SKSR1-HA control. The data obtained showed that host cell invasion in cultures infected with SKSR1^{m(+A)}-3HA and Δ *sksr1* parasites was comparable to that in cultures infected with SKSR1-HA parasites (Supplementary Fig. 15b), indicating that SKSR1 is not involved in parasite invasion.

SKSR1 is a virulence factor in *C. parvum*

We investigated the role of SKSR1 in *C. parvum* growth and virulence using in vitro and in vivo infection models. Compared to the SKSR1-3HA line, deletion of the *SKSR1* gene (Δ *sksr1*) significantly reduced *C. parvum* growth in vitro, and the A base insertion in the gene (SKSR1^{m(+A)}-3HA) also had a similar effect (Fig. 5a). We examined the differences in growth and pathogenicity between the Δ *sksr1*, SKSR1-3HA, and SKSR1^{m(+A)}-3HA lines in three independent experiments using the GKO mouse model, with five mice per group in each experiment. There was no significant difference in oocyst shedding among the three groups during the early stage of infection (DPI 6). However, the oocyst shedding intensity of the Δ *sksr1* and SKSR1^{m(+A)}-3HA groups was significantly lower than that of the SKSR1-3HA group from DPI 8 to DPI

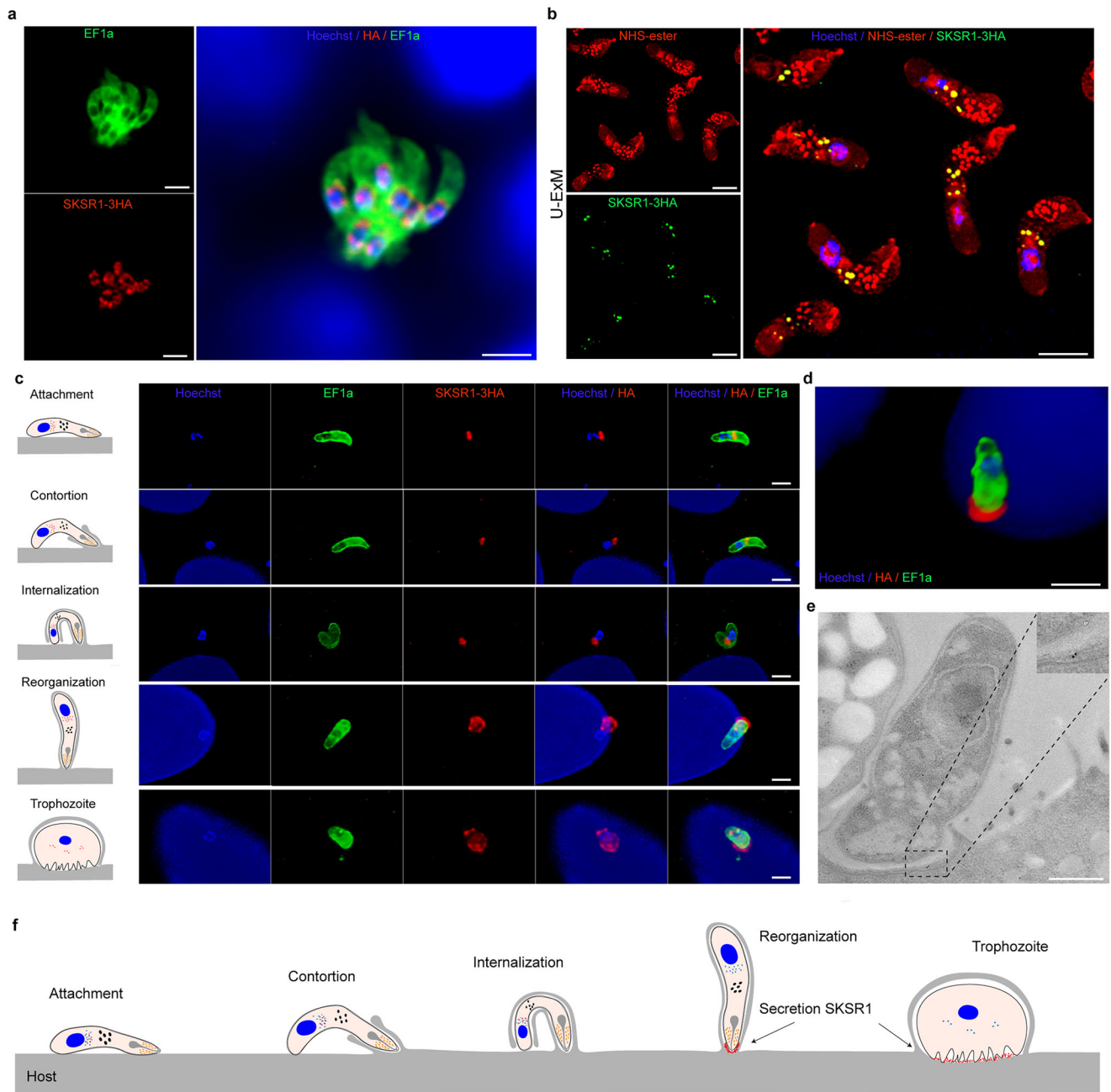


Fig. 4 | Localization of SKSR1 expression and translocation of the protein during invasion. **a** Identification of SKSR1-3HA expression in 8 egressing merozoites. SKSR1 (red) is localized near the nucleus; EF1a in green; nuclei (stained with Hoechst) in blue; scale bar = 5 μ m. **b** Ultrastructure expansion microscopy (U-ExM) showing SKSR1 in green relative to NHS ester-stained organelles (red); nuclei in blue; scale bar = 5 μ m. **c** Dynamics of SKSR1 expression during invasion. SKSR1 is translocated from small granules to the apical region of *C. parvum* sporozoites during the late phase of the host cell invasion. The cartoon images to the left of IFA

panels illustrate the process of sporozoite invasion. Scale bar = 5 μ m.

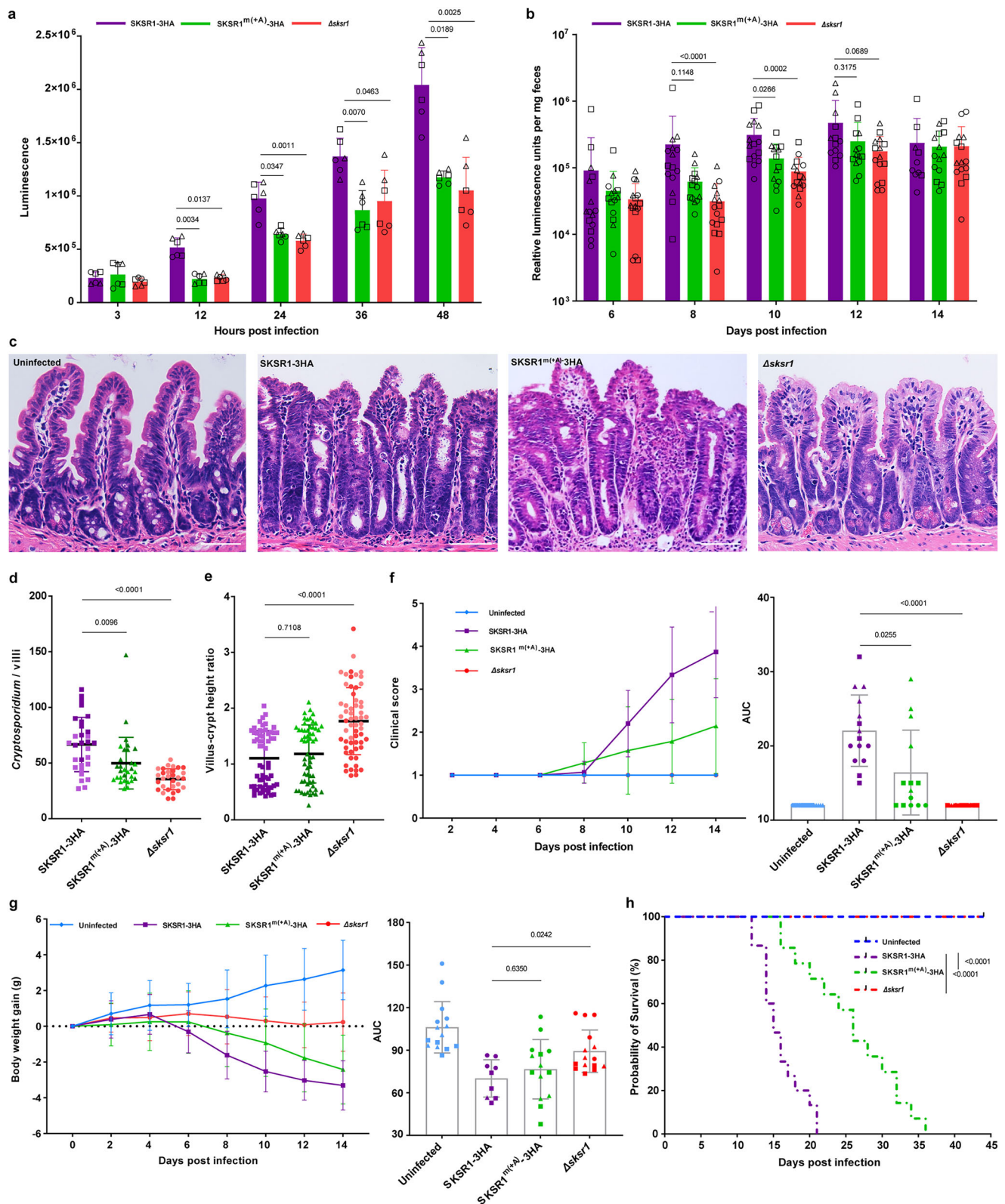
d Immunofluorescence localization of SKSR1 at the stage when sporozoites form a cup-like structure (side view) with the host cell membrane. Scale bar = 5 μ m.

e Immunoelectron microscopic localization of SKSR1. SKSR1 is localized at the parasite-host interface. Scale bars = 500 nm. **a–e** Each experiment was performed at least twice with similar results. **f** A model of SKSR1 secretion during *Cryptosporidium* invasion. The SKSR1 protein is shown in red. Source data are provided as a Source Data file.

12. After the peak of oocyst shedding at DPI 12, there was no significant difference in oocyst shedding intensity among the three infection groups (Fig. 5b). When the intestinal tissues of infected mice were analyzed in two infection studies, the parasite load on the villus surface was significantly different between the SKSR1-3HA and Δ sksr1 groups, with the SKSR1^{mt(A)}-3HA group having a parasite load intermediate between the two groups (Fig. 5c, d). SKSR1-3HA and SKSR1^{mt(A)}-3HA caused severe intestinal damage, whereas Δ sksr1 caused only mild villous atrophy (Fig. 5c). Mice infected with the Δ sksr1 line had a higher

ratio of villus height to crypt depth ratio than those infected with the SKSR1-3HA and SKSR1^{mt(A)}-3HA lines (Fig. 5e).

We used a scoring system to rank the severity of clinical signs of infected mice in each group, with higher scores representing worse conditions (Fig. 5f). Mice in the SKSR1-3HA group showed significant clinical signs and weight loss at DPI 8, with all animals moribund by DPI 21, whereas mice in the SKSR1^{mt(A)}-3HA group had a significantly delayed onset of clinical signs and weight loss (starting at DPI 12) and showed less weight loss than those in the SKSR1-3HA group (Fig. 5f–h).



In contrast, Δ sksr1 mice show no clinical signs and weight loss by DPI 14, and all mice in these groups survived to the end of three infection studies at DPI 45 (Fig. 5f–h).

To validate the involvement of SKSR1 in parasite competition between SKSR1-3HA and SKSR1^{m(+A)}-3HA in vivo. Equal numbers of SKSR1-3HA and SKSR1^{m(+A)}-3HA oocysts were used to infect three GKO mice. Analysis of the trace files from Sanger sequencing of the PCR products

from the fecal samples showed a clear double peak in the A-base insertion region, but due to the presence of competing nucleotides in this region, we were unable to determine the temporal changes of the two lines during the course of infection (Supplementary Fig. 16a). Two mice became moribund at DPI 11 and the remaining mouse succumbed to infection at DPI 15 (Supplementary Fig. 16b). The survival curves of the co-infections were similar to those of the SKSR1-3HA infections, suggesting that the SKSR1-3HA parasites may outcompeted the

Fig. 5 | Involvement of SKSR1 in the virulence of *Cryptosporidium parvum*.

a Growth pattern of SKSR1-3HA, SKSR1^{mi(A)}-3HA, and Δ sksr1 lines in HCT-8 cultures. The Δ sksr1 and SKSR1^{mi(A)}-3HA lines had significantly slower growth at 24 h and 48 h post infection (dots represent data from a total of six replicates in three independent experiments, and each bar represents the mean \pm SD). **b** Oocyst shedding pattern of Δ sksr1, SKSR1-3HA, and SKSR1^{mi(A)}-3HA lines in GKO mice. Dots represent data from three independent experiments with a total of 15 mice per group, and each bar represents the mean \pm SD. **c** Hematoxylin and eosin (H&E) microscopic images. Scale bar = 100 μ m. **d** Parasite load per villus on the ileal surface under H&E microscopy. **e** Villus length and crypt depth ratio of the ileum of uninfected and infected mice. Data in (**d** and **e**) are the mean \pm SD from two independent experiments (n = 15 villi per mouse), and different color shades represent two independent experiments. **f** Differences in clinical score (left) and the area under the curve (AUC) (right) (mean \pm SD) between groups. At 14 days post

infection, in three independent experiments, six mice in the SKSR1-3HA-infected group were moribund, which we defined as animals with the worst clinical signs. **g** Body weight gain and AUC (mean \pm SD) during the course of infection with different lines. Dead mice were not included in the AUC statistics for body weight gain. Statistical analysis in (**a**–**g**) was performed using the Kruskal–Wallis test. **h** Survival curves of infected mice. The time of death of mice in the SKSR1-3HA group was significantly different from that in the Δ sksr1 and SKSR1^{mi(A)}-3HA groups by the Gehan–Breslow–Wilcoxon test (two-tailed). Data in (**b** and **f**–**h**) are from three independent infection studies with five mice per group in each experiment, the different dots represent data from three independent experiments. One mouse in the SKSR1^{mi(A)}-3HA group died of a non-infection-related cause during the first infection study and its data were not included in subsequent analyses. Source data are provided as a Source Data file.

SKSR1^{mi(A)}-3HA parasites in vivo in the co-infection, leading to the lack of delay in death of infected mice.

Discussion

Virulence factors play a key role in the pathogenesis of microbial pathogens²⁶. They are often involved in the initial pathogen–host interactions, induce cellular damage, or alter host cellular responses. Different pathogens use different strategies to mediate virulence. In apicomplexan parasites, *T. gondii* uses a family of rhoptry-secreted polymorphic kinases as key virulence factors¹³. They are injected into host cells and mediate immune evasion and other changes in host cell signaling^{26,27}. In contrast, *P. falciparum* uses a family of variant PfEMP1 surface antigens encoded by the subtelomeric *var* genes. In addition to inducing immune evasion, these proteins mediate the adhesion of infected erythrocytes to endothelial cells in various organs, leading to different types of severe malaria²⁸. Linkage mapping of genetic crosses of isolates of different virulence has played a major role in the identification of virulence factors^{12,13}.

In the present study, we have successfully used linkage mapping of genetic crosses to identify one virulence factor in *C. parvum*. We have generated genetic crosses of fluorescently tagged *C. parvum* isolates HLJ and GD, which differ significantly in infectivity, infection pattern, and virulence in GKO mice¹⁰. Linkage mapping using BSA of genomes collected during infection with progeny pools has identified three QTL containing 23 genes that potentially control the growth difference between the two isolates. Reverse genetic studies of the two best candidates have confirmed the involvement of SKSR1, a secretory SG protein encoded by a polymorphic subtelomeric gene, in *C. parvum* virulence.

The generation of genetic crosses of *C. parvum* is greatly facilitated by the endogenous fluorescent tagging of isolates. Previous attempts at genetic crosses of *Cryptosporidium* isolates have been hampered by the lack of genetic tagging tools, making it difficult to isolate recombinant progeny difficult without fluorescent tagging and drug selection^{29,30}. Recent advances in genetic manipulation of *Cryptosporidium* spp. allow for endogenous tagging of isolates with fluorescent proteins and incorporation of selectable markers and luciferase sequences for easy detection and enrichment of recombinant progeny⁹. This approach has been successfully used to generate genetic crosses of a *C. parvum* isolate tagged with two different fluorescent proteins^{21,31,32}. Research has also shown that genetic crosses between species are possible. Crosses between *C. parvum* and *C. tyzzeri* have resulted in progeny with a recombinant genome derived from both species that continue to reproduce vigorously sexually³³. Despite the fact that these two species hybridize, large fragments or even entire chromosomes have been involved in genetic recombination, limiting the ability to identify genes related to host infectivity. In the present study, we have tagged two *C. parvum* isolates of different virulence and generated genetic crosses in mice, and purified the recombinant progeny using flow cytometric sorting. WGS analysis of

individual oocysts has confirmed the presence of frequent crossovers in the genomic sequences of the progeny, thus facilitating genetic mapping and gene identification.

The identification of virulence determinants in *Cryptosporidium* spp. in the present study is achieved by a new approach to perform linkage mapping of recombinant progeny using BSA of WGS data. The technique examines genomic changes during the course of infection using pools of progeny from genetic crosses. Alleles enriched in late infection are identified by read mapping of WGS data¹⁸. This eliminates the need to clone the progeny of genetic crosses, which is difficult to perform for *Cryptosporidium* spp. due to the lack of effective culture systems and the high ID₅₀ of isolates in laboratory animal models. This strategy has been used in combination with crosses to map genes associated with drug resistance, virulence, and other phenotypic traits in *P. falciparum*^{16,18}. In the present study, by BSA of 68 sets of WGS data collected from GKO and neonatal mice infected with progeny pools of two genetic crosses of *C. parvum* isolates of low and high virulence, we have identified the evolution toward the more virulent parent as the infection persisted. This is consistent with the observation in a previous study of genetic crosses of two *C. parvum* isolates with different host preferences³⁰. Using this approach, we have identified three QTL containing 23 genes on chromosomes 1, 7, and 8 that are potentially involved in the growth and virulence difference between the two *C. parvum* isolates studied.

While the strategy used in our BSA studies is designed to enrich for genes involved in rapid parasite growth, there is an established relationship between the severity of cryptosporidiosis and the intensity of infection in humans and animals. For example, in human volunteer studies of *C. parvum* infection, individuals with diarrhea had significantly higher oocyst shedding levels than those without diarrhea³⁴, and one high virulence isolate induced a higher intensity of oocyst shedding than two low virulence isolates⁷. Similarly, immunodeficient mice infected with virulent strains of *C. parvum* shed significantly more oocysts than those infected with avirulent strains, even when a lower number of oocysts from the virulent strain was used to establish infection^{9,10}. In calves experimentally infected with *C. parvum*, there is a high correlation between the severity of diarrhea and the intensity of oocyst shedding^{35,36}. This is supported by a recent treatment study in calves experimentally infected with *C. parvum* in which elimination of oocyst shedding with a *Cryptosporidium* PI(4)K inhibitor resulted in rapid resolution of diarrhea³⁷.

Most of the 23 candidate genes identified by BSA have only a small number of SNPs that are less likely to significantly affect the functions of the proteins they encode, making it difficult to intuitively determine their effect on the difference in virulence between the two *C. parvum* isolates studied. However, three of these genes, including cgd1_140 (SKSR1), CPCDC_7g4512, and cgd8_550, have premature stop codons due to single base insertion and loss-of-function SNPs. Among them, CPCDC_7g4512 has the deleterious nucleotide substitution in the highly virulent HLJ. Since this newly predicted gene is not expressed at

different developmental stages, CPCDC_7g4512 is unlikely to be involved in virulence. In contrast, cgd1_140 (*SKSR1*) and cgd8_550 have premature stop codons in the low virulence isolate GD, making them candidate virulence determinants.

Results from gene tagging and deletion studies suggest that the cgd8_550 gene is unlikely to be involved in virulence. The sequence characteristics of the gene are not related to those of other QTL identified in the study. In addition, the protein it encodes is a cytosolic protein that is predominantly expressed in the asexual stages rather than in the invasive sporozoites, suggesting that it is unlikely to interact with host cells as expected for known virulence factors in apicomplexans. Indeed, depletion of the gene in the virulent HLJ has minimal effect on the growth and pathogenicity of the isolate.

Studies using genetically modified lines of HLJ have confirmed the involvement of *SKSR1* in *C. parvum* virulence. Deletion of the *SKSR1* gene significantly reduced *C. parvum* growth in vitro, and attenuated parasite virulence in vivo. In particular, mice infected with the *Δsksr1* line had significantly reduced infection intensity and no weight loss and mortality. More importantly, replacement of the *SKSR1* gene in HLJ with GD also reduced the infection intensity and pathogenicity of the wild-type isolate, resulting in reduced weight loss and improved survival of infected mice. *SKSR1* shares some characteristics of virulence factors in other apicomplexans, including being a secretory protein encoded by a polymorphic subtelomeric gene, and being secreted into the parasite-host interface during invasion and early development^{38,39}. The *SKSR1* family is one of a small number of multigene families in *Cryptosporidium* spp. They are secreted proteins encoded by 8–11 subtelomeric genes. These genes are polymorphic within and between *Cryptosporidium* species, and are mainly found in species closely related to the human-pathogenic *C. parvum* and *C. hominis*⁴⁰. In *C. parvum*, gain and loss of *SKSR* genes is associated with the host range of isolates⁴¹. However, very few studies have investigated the biological function of *SKSR* proteins. One study genetically tagged *SKSR7* and *SKSR8*, but their locations and functions remain unknown⁴².

The preliminary characterizations performed in this study may shed light on the biological function of *SKSR1*. First, we have identified *SKSR1* as a member of the secretory proteins in the newly identified SG. Recently, two SG proteins were identified in *C. parvum* using spatial proteomics and gene tagging. Although the functions of these proteins remain unclear, SG1 has been shown to be secreted into the parasite-host interface soon after invasion and thus may be involved in host-pathogen interactions²⁴. Second, *SKSR1* is secreted into the parasite-host interface during the late phase of the invasion process. In *T. gondii*, several dense granule proteins are secreted into the PVM to modulate host cellular responses as part of parasite immune evasion⁴³. Third, *SKSR1* is one of several variant secretory *SKSR* proteins encoded by two gene clusters, suggesting that it may function in concert with other putative *C. parvum* virulence factors⁴⁴. Whether *SKSR1* acts as a virulence factor by regulating host cellular responses will require future studies.

Methods

Parasite isolates and mice

The GD and HLJ isolates of *C. parvum* used in this study were originally obtained from dairy calves in Guangdong (GD) and Heilongjiang (HLJ) provinces, China, respectively¹⁰ and propagated in interferon- γ knockout (GKO) C57BL/6j mice. *C. parvum* oocysts were purified from the feces of infected mice by sucrose flotation and cesium chloride gradient centrifugation as described¹⁰.

GKO mice were purchased from The Jackson Laboratory (Bar Harbor, USA). Mice were housed in clean filter-top cages with a 12:12 light-dark cycle, 50–60% humidity, and room temperature (22 °C) according to standard protocols under the regulations of the Laboratory Animal Center of the South China Agricultural University. This

study was approved by the Institutional Animal Care and Use Committee of South China Agricultural University (No. 2021C076). Animals used in the study were housed and handled in accordance with established ethical principles. Female and male mice (4–6 weeks old) were housed in individually ventilated cages (one mouse per cage) and used in a 1:1 ratio to generate and propagate stable transgenic parasites.

Construction of CRISPR/Cas9 plasmids

Homologous repair templates and CRISPR/Cas9 plasmids were generated as previously described³². The CRISPR/Cas9 plasmids were constructed by inserting an sgRNA targeting the gene of interest into the *C. parvum* Cas9/U6 linear plasmid amplified from pACT1:Cas9-GFP, U6:sgTK using Gibson assembly cloning. To generate the GD-*Δtk-mNeonGreen* and HLJ-*Δuprt*-tdTomato lines, we used mNeonGreen and tdTomato tags to replace GFP and mCh in TK-GFP-Nluc-P2A-neo-TK and UPRT-mCh-Nluc-P2A-neo-UPRT plasmids, respectively. To generate gene knockout lines (*Δsksr1* or *Δcgd8_550*), homology repair fragments were generated by PCR amplification of TK-mNeonGreen-Nluc-P2A-neo-TK and UPRT-tdTomato-Nluc-P2A-neo-UPRT. These fragments were assembled with corresponding regions of the gene of interest, including the 3' and 5' UTR sequences. To tag genes with the 3 × HA epitope, we modified the pINS1-3HA-Nluc-P2A-neo⁴⁵ by replacing its *INS1* C-terminal and 3' UTR sequences with regions of the gene of interest using a four-fragment Gibson assembly. In addition, to prevent erroneous cleavage of the repair plasmid by the CRISPR/Cas9 plasmid, we modified the gRNA or protospacer adjacent motif sequence at the 5' homology arm of the repair plasmid using a two-fragment Gibson assembly. In addition, we constructed repair templates and CRISPR/Cas9 plasmids for SG1 as previously described²⁴, which were subsequently used to generate an SG1-3HA line. The primers were synthesized by Sangon Biotech (Shanghai, China) and are listed in Supplementary Table 6.

Generation of transgenic parasites

Transgenic parasites were generated using CRISPR/Cas9 as previously described^{19,46}. Briefly, oocysts were treated with 25% Clorox (1.3% sodium hypochlorite) on ice for 10 min and excysted by treatment with sodium taurodeoxycholate at 37 °C for 60 min. The excysted sporozoites were transfected with an appropriate CRISPR/Cas9 plasmid containing a parasite-specific guide RNA (gRNA) and a targeting plasmid containing the desired genetic manipulation flanked by DNA homologous to the regions surrounding the gRNA cut site. The transfected sporozoites were administered by oral gavage to GKO mice pretreated with oral sodium bicarbonate to neutralize gastric acid. Transgenic parasites in infected mice were selected by continuous administration of 16 g/L PRM in the drinking water at 18 h post infection (HPI 18).

Measurement of parasite burden by luciferase assay

To quantify transgenic parasites expressing luciferase, fecal pellets from infected mice were weighed, placed in microfuge tubes containing 1 mL luciferase lysis buffer and glass beads, and vortexed for 3 min. The tubes were briefly centrifuged at 10,000 × g to pellet debris, and 100 μ L of the supernatant was transferred to a 96-well round-bottom plate (Costar, USA). Next, 100 μ L of a 1:50 mixture of nanoluciferase substrate:buffer (Promega, USA) was added to each sample, and luminescence was measured using a BioTek Synergy H1 Hybrid plate reader (BioTek, USA).

PCR analysis of transgenic parasites to confirm gene integration

DNA was extracted from oocysts of transgenic parasites using the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Germany). PCR primers were designed to anneal outside the 5' and 3' homology arms directing homologous recombination, the luciferase reporter gene

and the neomycin selection marker. In addition, primers were designed for the coding and mutant regions of the gene of interest to verify successful knockout and mutation of genes.

Genetic crossing of isolates with different virulence

Two GKO mice were treated orally with a suspension containing 10,000 each of HLJ-*Δuprt*-tdTomato oocysts and GD-*Δtk*-mNeonGreen oocysts. Infected mice were treated continuously with 16 g/L PRM via drinking water at HPI 18. They were monitored for fecal luciferase at DPI 4. If the mice were positive, they were euthanized and the small intestine of the mice was collected. The small intestine was rinsed several times with chilled PBS, and the suspension was filtered through 70 μm and 40 μm filters (Falcon, Cat#352350 and Cat#352350). The filtrate was centrifuged at 10,000 × g for 10 min, and the pellet was resuspended in 500 μL PBS for fluorescence microscopy on a BX53 microscope (Olympus, Japan). The oocyst suspension was also analyzed by flow cytometry on a FACSaria III (BD Biosciences, USA), with recombinant oocysts of dual color harvested by flow sorting. Two genetic crosses of the two isolates were performed during this study.

Isolation of single oocysts for whole genome sequencing

The progeny pool was diluted to a final concentration of 1 oocyst/μL, and 1 μL of the oocyst suspension was placed on a low adsorption slide for light microscopy. After confirming the presence of a single oocyst, the droplet was transferred to a PCR tube with 2.5 μL PBS from the droplet wash. The harvested single oocyst was repeatedly freeze-thawed in liquid nitrogen and a 55 °C water bath to lyse the oocysts. The released genomic DNA was amplified using the REPLI-g Single Cell Kit (QIAGEN) and analyzed by PCR targeting three genes that were polymorphic between the GD and HLJ isolates. Positive DNA from individual oocysts was sequenced for whole genomes as described below. The PCR sequences used are listed in Supplementary Table 6.

Infection of mice with progeny of genetic crosses

In the first infection study, GKO mice ($n = 3$) were orally gavaged with 1000 yellow oocysts from the progeny pool of the genetic cross and treated with PRM as described above. Fecal samples were collected from the infected mice every 6 days for 36 days starting at DPI 6. In the second infection study, two infection groups were established with and without PRM treatment, and mice ($n = 3$) in each group were infected with progeny pool from the second cross. A total of 49 samples were collected at multiple time points of infection as described in the first infection experiment. In the third infection study, one liter of 5-day-old C57BL/6j mice were infected as in the second infection study and 4 samples were collected at DPI 12, DPI 18, DPI 24, and DPI 30. Oocysts in fecal samples collected during the infection studies were purified by cesium chloride and sucrose gradient centrifugation for WGS.

Library preparation and sequencing

Genomic DNA was extracted from the purified oocysts using the QIAGEN DNA Mini Kit and amplified using the REPLI-g Midi Kit (QIAGEN). Whole-genome amplification products were purified using the EasyPure PCR Purification Kit (TransGen Biotech) and randomly sheared into fragments of ~350 bp. Libraries were constructed using the Rapid Plus DNA Lib Prep Kit for Illumina (RK20208) and sequenced using Illumina 150 bp paired-end technology on a Nova-seq S4 or HiSeq X sequencer to achieve >100-fold coverage of the *Cryptosporidium* genome. In addition to sequencing 12 single oocysts from the progeny pool, 68 oocyst pools from the progeny pool infection studies were sequenced during the study. To generate a chromosome-level reference genome, the parental HLJ strain was also sequenced using standard Pacific Biosciences (PacBio) ccs procedures after whole-genome amplification of the extracted DNA as described above.

Assembly and annotation of the reference HLJ genome

Sequence reads from PacBio were filtered for chimeras using 3rd-ChimeraMiner⁴⁷. Hybrid de novo assembly was performed on the PacBio and Illumina reads of HLJ using wengan v0.2 with M model. The primary assembly underwent two rounds of polishing with PacBio reads and four rounds of polishing with Illumina sequencing data using NextPolish⁴⁸. The sequencing data were manually checked against the assembled genome to obtain the final genome assembly. AUGUSTUS and GeneMark-ES v4 were used for de novo gene prediction, and homology-based prediction GeMoMa v1.9 and the rapid annotation transfer tool Miniprot v 0.13 were used to transfer gene annotation information from the *C. parvum* IOWA-43IA8 reference genome to the HLJ genome. Orthofinder v 2.5.5 was used to analyze the annotated amino acid sequences of HLJ with the amino acid sequences of *C. parvum* IOWA-43IA8²² and IOWA-ATCC⁴⁹ to identify orthologous genes and gene families to obtain the final genome annotation file. To identify the structural differences among the *C. parvum* strains, we aligned the fully assembled genomes of IOWA-43IA8, IOWA-ATCC, IOWA-BGF, Ila-KWI52, and HLJ using Mauve (<https://darlinglab.org/mauve/mauve.html>).

Mapping, genotyping and BSA analysis

Fastq files from WGS were checked for quality using FastQC-v0.11.05, and sequence reads were trimmed for poor quality and adapters using Trimmomatic v0.39⁵⁰ (MINLEN:60, SLIDINGWINDOW:4:15, ILLUMINA-LIP:Truseq3-PE.fa:2:30:10). The cleaned reads were aligned to the *C. parvum* HLJ genome using BWA-MEM2 v2.2⁵¹. The resulting alignments were sorted and duplicate reads were removed using SAMtools v1.7 and Sambamba v0.8.2^{52,53}. Read mapping results were viewed using the Integrative Genomics View⁵⁴. We identified sequence cross-overs from single oocysts in the progeny pool using variant allele frequency (VAF) = AD/DP (AD stands for allele depth, DP stands for total depth); the difference of VAF > 0.8 between adjacent SNPs in single-oocyst genomes indicates that a sequence crossover has occurred in the region.

For BSA, the previously described approach was used¹⁶. Briefly, variants were called and filtered using BCFtools v1.12⁵² and annotated using SnpEff v4⁵⁵, which provides a simple assessment of the putative impact of the variant (e.g., HIGH, MODERATE or LOW impact). The BCFtools filter was used to remove variants with Phred scores <30 and average read depth ≤25. The modules *runQTLseqAnalysis* and *runG-primeAnalysis* in the R package QTLseqr were used to calculate the delta SNP index and G' values⁵⁶. We calculated the G-statistic for each locus of the genome using data from each BSA experiment and used the mean plus three standard deviations of the G' values as the cutoff for each locus. Because the average G-statistic value for different BSA studies was used to identify QTL, some regions with marginal G-statistic values in individual BSA studies may have been missed. Regions with G' > the threshold were identified as extreme QTL. Once a QTL was detected, genes within the QTL regions were extracted based on the intersection of the QTL regions obtained from the BSA replicate experiments using jvenn (<https://jvenn.toulouse.inra.fr/app/example.html>). The expression patterns of the candidate genes were determined based the transcriptomic data we previously described⁵⁷.

Assessment of transgenic parasite development in vitro and IFA

Human ileocecal adenocarcinoma cells (HCT-8; ATCC CCL-244) were seeded onto 48-well plates and grown to 80% confluence. Bleach-treated transgenic parasite oocysts (10,000 oocysts per well) with identical luminescence levels were then used to infect the HCT-8 monolayer for various time periods: 3, 12, 24, 36, and 48 h. Cultures infected with transgenic oocysts were then washed twice with PBS at HPI 3 and replenished with fresh medium containing 2% fetal bovine serum. At multiple time points, the culture medium was removed from the wells, 100 μL of lysis buffer was added to the wells, and the plate

was incubated at 37 °C for 10 min. After incubation at 37 °C for 10 min, the lysates were collected by centrifugation at $15,000 \times g$ for 3 min and subjected to luciferase assay as described previously.

For IFA, 12-mm diameter glass coverslips (Thermo Fisher Scientific) were placed in 24-well plates and seeded with HCT-8 cells. The cells were washed with PBS, fixed by incubation with 4% paraformaldehyde (Thermo Fisher Scientific), and permeabilized by treatment with 0.5% Triton X-100 (Bio-rad) for 15 min. Coverslips were then blocked with 1% bovine serum albumin (BSA; Sigma-Aldrich) and incubated with antibodies diluted in the blocking solution, including a rabbit monoclonal antibody against HA (1:800; Cell Signaling Technology) as the primary antibody and a goat anti-rabbit polyclonal Alexa Fluor 594 (1:400; Invitrogen) as secondary antibody, along with direct staining of parasite stages using Vicia Villosa Lectin (1:1000; VVL, Vector). Host cell and parasite nuclei were probed with Hoechst (5 µg/mL, Thermo Fisher Scientific). For sporozoite IFA, sporozoites were resuspended in 10 µL PBS and spread on sterile coverslips pretreated with poly-L-lysine (Sigma-Aldrich). After fixation and permeabilization, they were treated with mouse anti-Cp23 (1:200; Laboratory Preparation) or anti-EF1a (1:200; Laboratory Preparation) and rabbit monoclonal anti-HA (1:800; Cell Signaling Technology) as primary antibodies and goat anti-rabbit polyclonal Alexa Fluor 594 (1:400; Invitrogen) and goat anti-mouse polyclonal Alexa Fluor 488 (1:400; Invitrogen) as secondary antibodies. Each analysis was performed in duplicate for at least two biological replicates. Slides were examined using either a Leica STELLARIS 5 or an Olympus BX53 microscope.

Gliding and invasion assays

For the gliding assay⁵⁸, 100 µL of poly-L-lysine was added to 12 mm diameter glass coverslips and incubated for 30 min at room temperature. After washing with PBS, 50 µL of 1×10^6 SKSR1-HA sporozoite suspension was added to coverslips and incubated at 37 °C for 30 min, and the liquid on the coverslips was washed with PBS. After fixation and permeabilization, SKSR1 protein was identified using rabbit monoclonal anti-HA (1:800) as the primary antibody and goat anti-rabbit polyclonal Alexa Fluor 594 (1:400) as the secondary antibody; mouse anti-Cp23 (1:200) was used as the primary antibody (1:200) and goat anti-mouse polyclonal Alexa Fluor 488 (1:400) as the secondary antibody to detect the parasite and glide track.

An attachment and invasion assay²⁵ was used to evaluate the effect of SKSR deletion on *C. parvum* invasion of host cells. Briefly, a suspension of 2×10^5 sporozoites was added to HCT-8 cells in 24-well plates. After 2.5 h of infection, all wells were washed three times with PBS and fixed. Extracellular parasites were labeled with rabbit anti-Cp and goat anti-rabbit IgG Alexa Fluor 488. The wells were then permeabilized with 0.05% Triton X-100. After permeabilization, all parasites were labeled with rabbit anti-Cp and goat anti-rabbit IgG Alexa Fluor 594 and nuclei were stained with Hoechst 33342 nuclear stain. Slides were examined under an Olympus BX53 microscope to count the number of green fluorescent parasites (extracellular parasites) and two-color fluorescent parasites (intracellular parasites). The invasion efficiency of the parasites was calculated using previously described methods.

Ultrastructure expansion microscopy

Ultrastructure-expansion microscopy (U-ExM) was performed on *Cryptosporidium* sporozoites and meronts as described for *T. gondii*⁵⁹. Briefly, excysted sporozoites were added to poly-D-lysine-coated coverslips and fixed with 1.4% formaldehyde and 2% acrylamide in PBS. The samples were embedded in a water-based gel and denatured at 95 °C. The gels were then probed with rat monoclonal antibody (3F10) against HA (1:200; Roche) as the primary antibody and rabbit anti-rat IgG AF488 (1:200; Invitrogen) as the secondary antibody, along with direct staining of parasite organelles using NHS ester (10 µg/mL; Sigma-Aldrich). Hoechst was used to stain host cell

and parasite nuclei. The gels were imaged using STELLARIS 5 (Leica Microsystems).

Western blot analysis of transgenic parasite oocysts

Purified transgenic oocysts (5×10^6) were treated with bleach, washed with cold PBS, and resuspended in lysis buffer (ThermoFisher Scientific) containing protease inhibitors (Sigma-Aldrich). The mixture was incubated overnight at 4 °C, combined with protein loading buffer, and boiled for 10 min. Proteins in the lysate were then fractionated by SDS-PAGE and transferred to a PVDF membrane (Merck Millipore). After blocking with 1% nonfat milk overnight at 4 °C, the membrane was incubated with a rabbit monoclonal antibody against HA (1:2000; Cell Signaling Technology) for 1 h, and washed three times with PBS containing 0.1% Tween 20 (PBST). After incubation with HRP-conjugated anti-rabbit IgG (H + L) (1:10,000; Cell Signaling Technology), washed three times with PBST, and treated with High-sig ECL Western Blotting Substrate (Tanon), the membrane was analyzed using a Tanon 5200. The membrane was further probed with a mouse polyclonal antibody (1:2000) against CP23 of *C. parvum* and HRP-conjugated goat anti-mouse IgG (H + L) (1:10,000; Cell Signaling Technology).

Histologic and immunoelectron microscopic analysis of infected tissue

For studies of the biological significance of virulence-related genes, one mouse from each experimental group was selected and euthanized on DPI 14 at the peak of oocyst shedding. The small intestine was dissected and washed with cold PBS. The ileum was collected for conventional hematoxylin-eosin (H&E) and IEM. For H&E microscopy, 30 villi were randomly selected for measurement of villus length and crypt height and calculation of villus length/crypt height ratio. Parasite burden was also determined in 15 intestinal villi. For IEM, rabbit anti-HA (1:20; Cell Signaling Technology) and goat anti-rabbit IgG conjugated with 10 nm colloidal gold (1:20; Sigma-Aldrich) were used as primary and secondary antibodies, respectively, as described⁴⁵. Processed sections were examined on a Talos L120C (ThermoFisher Scientific) by a technician blinded to ileal tissue group assignment.

Assessment of virulence of transgenic *C. parvum* in mice

To assess the biological significance of each candidate gene, GKO mice in each infection group were orally gavaged with 10,000 oocysts of transgenic *C. parvum*, and received PRM via drinking water as described above. To assess the relevance of SKSR1 for virulence, we performed three independent experiments with six mice per group in each experiment. Intestinal tissues from one mouse per experimental group in each experiment were used for histological and immunoelectron microscopic analyses as described above. After infection, fecal luciferase activity and body weight of each mouse were determined every other day, and mortality and other clinical signs were recorded daily. A scoring system was used to evaluate the severity of the clinical signs in infected mice, and this scoring was blinded: 1: mice were in good health and physically active; 2: mice appeared lethargic and moved less frequently; 3: mice were obviously lethargic, had arched backs, and moved only when touched by hand; 4: mice were hunched, had rough hair, and were immobile even when touched; 5: mice are in extremely poor mental condition. Survival curves were plotted at the end of the infection study.

Statistical analysis

GraphPad Prism (<https://www.graphpad.com/>) was used for all statistical analyses. Group means were obtained from at least two independent experiments or three biological replicates. Unless otherwise noted, two-tailed Mann-Whitney test was used to assess differences between two groups, and Kruskal–Wallis test was used to assess differences between three or more groups. The Gehan-Breslow-Wilcoxon

test was used to compare the survival curves. P values < 0.05 were considered statistically significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All whole-genome sequence data generated in this study have been deposited in the NCBI Short Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) under the BioProject accession numbers PRJNA1069297. The fully assembled and annotated genome of the IIdA20G1-HLJ strain is deposited in DDBJ/ENA/GenBank databases under the accession number JBJGDY000000000. All data necessary to evaluate the conclusions of the paper are included present in the paper and/or in the Supplementary Materials. Source data are provided with this paper.

Code availability

The code used in analysis and data analyzed are available at GitHub and Code Ocean through the following links: <https://github.com/tyhou/BSA.Cpar.IId.GDxHLJ>; <https://codeocean.com/capsule/0441727/tree/v1>.

References

- Kotloff, K. L. et al. 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 (2013).
- Gharpure, R. et al. Cryptosporidiosis outbreaks - United States, 2009–2017. *MMWR Morb. Mortal. Wkly. Rep.* **68**, 568–572 (2019).
- Checkley, W. et al. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. *Lancet Infect. Dis.* **15**, 85–94 (2015).
- Feng, Y., Ryan, U. M. & Xiao, L. Genetic diversity and population structure of *Cryptosporidium*. *Trends Parasitol.* **34**, 997–1011 (2018).
- Guo, Y. et al. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *BMC Genomics* **16**, 320 (2015).
- Guerin, A. & Stripen, B. The biology of the intestinal intracellular parasite *Cryptosporidium*. *Cell Host Microbe* **28**, 509–515 (2020).
- Okhuysen, P. C., Chappell, C. L., Crabb, J. H., Sterling, C. R. & DuPont, H. L. Virulence of three distinct *Cryptosporidium parvum* isolates for healthy adults. *J. Infect. Dis.* **180**, 1275–1281 (1999).
- Bartley, P. M. et al. Differences in virulence and oocyst shedding profiles in lambs experimentally infected with different isolates of *Cryptosporidium parvum*. *Curr. Res. Parasitol. Vector Borne Dis.* **4**, 100127 (2023).
- Sayed, F. G., Hamza, A. I., Gala, L. A., Sayed, D. M. & Gaber, M. Virulence of geographically different *Cryptosporidium parvum* isolates in experimental animal model. *Ann. Parasitol.* **62**, 221–232 (2016).
- Jia, R. et al. High infectivity and unique genomic sequence characteristics of *Cryptosporidium parvum* in China. *PLoS Negl. Trop. Dis.* **16**, e0010714 (2022).
- Chalmers, R. M., Robinson, G., Elwin, K. & Elson, R. Analysis of the *Cryptosporidium* spp. and gp60 subtypes linked to human outbreaks of cryptosporidiosis in England and Wales, 2009 to 2017. *Parasit. Vectors* **12**, 95 (2019).
- Su, X. Z., Lane, K. D., Xia, L., Sa, J. M. & Wellems, T. E. *Plasmodium* genomics and genetics: new insights into malaria pathogenesis, drug resistance, epidemiology, and evolution. *Clin. Microbiol. Rev.* **32**, e00019–e00019 (2019).
- Behnke, M. S., Dubey, J. P. & Sibley, L. D. Genetic mapping of pathogenesis determinants in *Toxoplasma gondii*. *Annu. Rev. Microbiol.* **70**, 63–81 (2016).
- Taylor, S. et al. A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science* **314**, 1776–1780 (2006).
- Abkhallo, H. M. et al. Rapid identification of genes controlling virulence and immunity in malaria parasites. *PLoS Pathog.* **13**, e1006447 (2017).
- Amambua-Ngwa, A. et al. Chloroquine resistance evolution in *Plasmodium falciparum* is mediated by the putative amino acid transporter AAT1. *Nat. Microbiol.* **8**, 1213–1226 (2023).
- Mok, S. et al. Mapping the genomic landscape of multidrug resistance in *Plasmodium falciparum* and its impact on parasite fitness. *Sci. Adv.* **9**, eadi2364 (2023).
- Vendrey, K. M., Kumar, S., Li, X. & Vaughan, A. M. Humanized mice and the rebirth of malaria genetic crosses. *Trends Parasitol.* **36**, 850–863 (2020).
- Vinayak, S. et al. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* **523**, 477–480 (2015).
- Charania, R., Wade, B. E., McNair, N. N. & Mead, J. R. Changes in the microbiome of *Cryptosporidium*-infected mice correlate to differences in susceptibility and infection levels. *Microorganisms* **8**, 879 (2020).
- Kimball, A. et al. Mendelian segregation and high recombination rates facilitate genetic analyses in *Cryptosporidium parvum*. *PLoS Genet.* **20**, e1011162 (2024).
- Huang, W. et al. Sequence introgression from exogenous lineages underlies genomic and biological differences among *Cryptosporidium parvum* IOWA lines. *Water Res.* **254**, 121333 (2024).
- Takagi, H. et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**, 174–183 (2013).
- Guerin, A. et al. *Cryptosporidium* uses multiple distinct secretory organelles to interact with and modify its host cell. *Cell Host Microbe* **31**, 650–664 (2023).
- Akey, M. E., Xu, R., Ravindran, S., Funkhouser-Jones, L. & Sibley, L. D. Apical secretory glycoprotein complex contributes to cell attachment and entry by *Cryptosporidium parvum*. *mBio* **14**, e0306422 (2023).
- Johnson, R., Mylona, E. & Frankel, G. *Typhoidal Salmonella*: distinctive virulence factors and pathogenesis. *Cell Microbiol.* **20**, e12939 (2018).
- Rastogi, S., Cygan, A. M. & Boothroyd, J. C. Translocation of effector proteins into host cells by *Toxoplasma gondii*. *Curr. Opin. Microbiol.* **52**, 130–138 (2019).
- Wahlgren, M., Goel, S. & Akhouri, R. R. Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat. Rev. Microbiol.* **15**, 479–491 (2017).
- Feng, X., Rich, S. M., Tzipori, S. & Widmer, G. Experimental evidence for genetic recombination in the opportunistic pathogen *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* **119**, 55–62 (2002).
- Tanriverdi, S., Blain, J. C., Deng, B., Ferdig, M. T. & Widmer, G. Genetic crosses in the apicomplexan parasite *Cryptosporidium parvum* define recombination parameters. *Mol. Microbiol.* **63**, 1432–1439 (2007).
- Hanna, J. C. et al. Mode of action studies confirm on-target engagement of lysyl-tRNA synthetase inhibitor and lead to new selection marker for *Cryptosporidium*. *Front. Cell Infect. Microbiol.* **13**, 1236814 (2023).
- Wilke, G. et al. A Stem-cell-derived platform enables complete *Cryptosporidium* development in vitro and genetic tractability. *Cell Host Microbe* **26**, 123–134 (2019).

33. Shaw, S. et al. Genetic crosses within and between species of *Cryptosporidium*. *Proc. Natl. Acad. Sci. USA* **121**, e2313210120 (2024).
34. Chappell, C. L., Okhuysen, P. C., Sterling, C. R. & DuPont, H. L. *Cryptosporidium parvum*: intensity of infection and oocyst excretion patterns in healthy volunteers. *J. Infect. Dis.* **173**, 232–236 (1996).
35. Operario, D. J., Bristol, L. S., Liotta, J., Nydam, D. V. & Houpt, E. R. Correlation between diarrhea severity and oocyst count via quantitative PCR or fluorescence microscopy in experimental cryptosporidiosis in calves. *Am. J. Trop. Med. Hyg.* **92**, 45–49 (2015).
36. Belloso, M. L. et al. A comparison of fecal percent dry matter and number of *Cryptosporidium parvum* oocysts shed to observational fecal consistency scoring in dairy calves. *J. Parasitol.* **97**, 349–351 (2011).
37. Manjunatha, U. H. et al. *Cryptosporidium* PI(4)K inhibitor EDIO48 is a gut-restricted parasitocidal agent to treat paediatric enteric cryptosporidiosis. *Nat. Microbiol.* **9**, 2817–2835 (2024).
38. Sanchez, S. G. & Besteiro, S. The pathogenicity and virulence of *Toxoplasma gondii*. *Virulence* **12**, 3095–3114 (2021).
39. Walker, I. S. & Rogerson, S. J. Pathogenicity and virulence of malaria: sticky problems and tricky solutions. *Virulence* **14**, 2150456 (2023).
40. Xu, Z., Guo, Y., Roellig, D. M., Feng, Y. & Xiao, L. Comparative analysis reveals conservation in genome organization among intestinal *Cryptosporidium* species and sequence divergence in potential secreted pathogenesis determinants among major human-infecting species. *BMC Genomics* **20**, 406 (2019).
41. Wang, T. et al. Sympatric recombination in zoonotic *Cryptosporidium* leads to emergence of populations with modified host preference. *Mol. Biol. Evol.* **39**, msac150 (2022).
42. Dumaine, J. E. et al. The enteric pathogen *Cryptosporidium parvum* exports proteins into the cytosol of the infected host cell. *Elife* **10**, e70451 (2021).
43. Panas, M. W. & Boothroyd, J. C. Seizing control: how dense granule effector proteins enable *Toxoplasma* to take charge. *Mol. Microbiol.* **115**, 466–477 (2021).
44. Tichkule, S. et al. Global population genomics of two subspecies of *Cryptosporidium hominis* during 500 years of evolution. *Mol. Biol. Evol.* **39**, msac056 (2022).
45. Xu, R., Feng, Y., Xiao, L. & Sibley, L. D. Insulinase-like protease 1 contributes to macrogamont formation in *Cryptosporidium parvum*. *mBio* **12**, e03405–e03420 (2021).
46. Sateriale, A., Pawlowic, M., Vinayak, S., Brooks, C. & Striepen, B. Genetic manipulation of *Cryptosporidium parvum* with CRISPR/Cas9. *Methods Mol. Biol.* **2052**, 219–228 (2020).
47. Lu, N. et al. Exploration of whole genome amplification generated chimeric sequences in long-read sequencing data. *Brief Bioinform.* **24**, bbad275 (2023).
48. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
49. Baptista, R. P. et al. Long-read assembly and comparative evidence-based reanalysis of *Cryptosporidium* genome sequences reveal expanded transporter repertoire and duplication of entire chromosome ends including subtelomeric regions. *Genome Res.* **32**, 203–213 (2022).
50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
52. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Giga-Science* **10**, giab008 (2021).
53. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
54. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
55. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly. (Austin)* **6**, 80–92 (2012).
56. Mansfeld, B. N. & Grumet, R. QTLseql: an R package for bulk segregant analysis with next-generation sequencing. *Plant Genome* **11**, 180006 (2018).
57. Li, M. et al. Variant surface protein GP60 contributes to host infectivity of *Cryptosporidium parvum*. *Commun. Biol.* **7**, 1175 (2024).
58. Wetzel, D. M., Schmidt, J., Kuhlenschmidt, M. S., Dubey, J. P. & Sibley, L. D. Gliding motility leads to active cellular invasion by *Cryptosporidium parvum* sporozoites. *Infect. Immun.* **73**, 5379–5387 (2005).
59. Dos Santos Pacheco, N. & Soldati-Favre, D. 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 (2021).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (32030109 to L.X., 31972697 to Y.G., 31820103014 to L.X. and 32150710530 to L.X.), Guangdong Major Project of Basic and Applied Basic Research (2020B0301030007 to L.X.), 111 Project (D20008 to L.X.) and Double First-class Discipline Promotion Project (2023B10564003 to L.X.). We thank Jilei Huang, Chuanhe Liu, and Xiaoxian Wu of the Instrumental Analysis & Research Center, South China Agricultural University for assistance with electron microscopy. The authors also thank the Institute of Hematology, Jinan University, for technical assistance in flow cytometric sorting.

Author contributions

Conceptualization: L.X., Y.F. and L.D.S.; Methodology and investigation: W.H., L.S., T.H., Z.Y., F.Y., S.Z., T.W., X.W., N.L. and Y.G.; Formal Analysis: W.H., L.S. and T.H.; Supervision: L.X., Y.F. and L.D.S.; Writing—original draft: W.H., L.X. and Y.F.; Writing—review and editing: All authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60088-7>.

Correspondence and requests for materials should be addressed to L. David Sibley, Yaoyu Feng or Lihua Xiao.

Peer review information *Nature Communications* thanks Jessica Kissinger, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025