Specific pathway abundances in the neonatal calf faecal microbiome are associated with susceptibility to Cryptosporidium parvum infection: A metagenomic analysis

Miriam F Hares (mhares@liverpool.ac.uk)
University of Liverpool

Bethany E Griffiths
University of Liverpool

Flora Johnson
University of Liverpool

Charlotte Nelson
University of Liverpool

Samuel Haldenby
University of Liverpool

Christopher J Stewart
Newcastle University

Jennifer S Duncan
University of Liverpool

Georgios Oikonomou
University of Liverpool

Janine Coombes
Robert Gordon University

Research Article

Keywords: Cryptosporidium parvum, bovine, cryptosporidiosis, faecal microbiome, metagenome, pathway abundances, functional profiling

Posted Date: April 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2844476/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Cryptosporidium parvum is the main cause of calf scour globally. With limited therapeutic options and research compared to other Apicomplexa, it is important to understand the parasites’ biology and interactions with the host and microbiome in order to develop novel strategies against this infection. The age-dependent nature of symptomatic cryptosporidiosis suggests a link to the undeveloped immune response, the immature intestinal epithelium, and its associated microbiota. This led us to hypothesise that specific features of the early life microbiome could predict calves’ susceptibility to C. parvum infection.

Results

In this study, faecal samples were collected from ≤ 1-week-old calves (n = 346). A retrospective case-control approach was taken whereby a metagenomic analysis was conducted on healthy calves (Control group; n = 30) and calves that went on to develop diarrhoea and test positive for C. parvum infection (Cryptosporidium-positive group; n = 30). Taxonomic analysis showed no significant differences in alpha diversity, beta diversity, and taxa relative abundance between controls and Cryptosporidium-positive groups. Analysis of functional potential showed pathways related to isoprenoid precursor, haem and purine biosynthesis were significantly higher in abundance in calves that later tested positive for C. parvum (q ≤ 0.25). These pathways are uniquely lacking in the C. parvum parasites, unlike the other Apicomplexa. Though the de novo production of isoprenoid precursors, haem and purines are absent, C. parvum has been shown to encode enzymes that catalyse the downstream reactions of these pathway metabolites, indicating that C. parvum may scavenge those products from an external source.

Conclusions

The host has previously been put forward as the source of essential metabolites, but our study suggests that C. parvum may also be able to harness specific metabolic pathways of the microbiota in order to survive and replicate. This finding is important as components of these microbial pathways could be exploited as potential drug targets for the prevention or mitigation of cryptosporidiosis in bovine neonates.

Background

Cryptosporidium parvum is an apicomplexan, protozoan parasite that invades the small intestinal epithelium of neonatal calves. It causes an acute diarrheal disease known as cryptosporidiosis, which is characterised by watery diarrhoea, dehydration, weight loss and even death in severe cases. Cryptosporidiosis leads to approximately 37% of all diarrhoea events and 20% of co-infections in calves in the UK, culminating in production losses of approximately £130 per calf affected and poorer overall animal welfare [1, 2]. Consequently, it is a serious veterinary issue which requires effective therapies to combat infection. With no vaccine currently available against bovine cryptosporidiosis, the current therapeutic options in cattle are limited to the antibiotic, paromomycin, and the anti-cryptosporidial, FDA-approved drug, halofuginone, which is believed to target the merozoite and sporozoite stages [3]. Unfortunately, both halofuginone and paromomycin have been found to have variable efficacy against cryptosporidiosis in calves. While the cryptosporidiostatic effect can both reduce oocyst shedding and severity of diarrhoea, these treatments are not lethal to Cryptosporidium and oocyst shedding and diarrhoea will often commence on drug withdrawal [4–10]. In addition, halofuginone has high toxicity at twice the recommended dose, leading to adverse side effects, therefore calves must be weighed in order to administer an effective, non-lethal dose [3, 11–13]. In light of this, the development of new effective therapies against cryptosporidiosis in calves is crucial, not only from an animal welfare point of view but also from an economic perspective.

Notably, age is a major risk factor that determines symptomatic C. parvum infection in cattle. Neonatal calves exhibit acute diarrheal disease and shed high numbers of oocysts, whilst infected adults remain asymptomatic though are still a source of oocyst environmental contamination [14]. With age comes naturally reoccurring challenge from the parasite in the environment, leading to the development of immunological memory and long-term resistance. However, the intestinal mucosa displays features independent of the immune system that may determine calf predisposition to clinical infection. According to one study,
when inoculated with the scrapings from adult cow and rat intestinal mucosa, neonatal rats were protected from \textit{C. parvum} infection, but interestingly, scrapings from naïve calves and calves with a prior infection had no such effect \cite{15}. A follow-up study identified the active component of the adult cow and rat intestinal scrapings to be leucine aminopeptidase \cite{16}. This work implies that local features of the adult intestinal mucosa confer protection against \textit{C. parvum} infection, irrespective of prior challenge. Though if an immune response was generated in calves, it appears to be ineffective.

A preliminary intestinal microbial colony is formed within hours of birth from maternally and environmental sources, which then goes on to rapidly change and diversify throughout the first year of life, ultimately becoming the stable population of microbes found in the gut of adult cattle \cite{17}. The mature adult microbiome provides protection against infection through interactions with the intestinal epithelial cells that fortify the mucosal barrier \cite{18}. Unfortunately, there is limited data on the maturation of the calf microbiota with existing studies only containing small sample sizes and therefore we do not yet fully understand how the composition of the intestinal flora impacts susceptibility to infectious diarrhoea \cite{19–21}.

We hypothesised that specific features of the intestinal microbiota of calves prior to infection could predict calf susceptibility to cryptosporidiosis. In a retrospective case-control study, we conducted a metagenomic analysis of faecal samples collected from calves during the first week of life, that aimed to determine any pre-disposing taxonomic and functional characteristics of the microbiome that are associated with susceptibility to cryptosporidiosis in neonatal calves.

**Methods**

The study was conducted following ethical approval by the University of Liverpool Research Ethics Committee (VREC927) and procedures regulated by the Animals (Scientific Procedures) Act were conducted under a UK Home Office License (P191F589B).

**Animals**

65 female Holstein dairy calves were used in this study from 3 farms (Farm 1, 2, and 3) based in North Wales and Cheshire, as part of a prospective cohort study that enrolled 346 calves. Calves that had received routine antibiotic and/or anti-cryptosporidial prophylactic treatment were included as this is common practice on UK farms. All calves received a similar dietary regime of cow colostrum in the first 24 hours of life, followed by milk replacer and were then weaned onto a standard cereal and hay-based diet. The breed and farm management of the sample population of calves on all farms was considered by the veterinary team as representative of the UK dairy calf population. The calves were monitored throughout the study by body condition score (BCS), and the Wisconsin Scoring System as well as a scoring system developed by the sample collector to determine the health status of the calves \cite{22, 23}. In addition, blood serum total protein was measured within 7 days of birth and thoracic ultrasonography was used to identify respiratory disease post-weaning.

All calves included in the study displayed no clinical signs of cryptosporidiosis in the first week of life. The study design is presented in Fig. 1.

**Sample Collection**

Faecal samples were collected from 346 calves less than one week old by rectal swab (Sterilin Regular Nylon Flocked Swabs 552C, Scientific Laboratory Supplies), prior to the development of any clinical signs of cryptosporidiosis, and stored on dry ice immediately after collection. Samples were transferred to -80°C within a few hours from collection and stored until DNA extraction. The health monitoring conducted over the course of the study included a faecal score which was used to determine if a diarrhoea event had occurred \cite{23}. A diarrhoea event was defined as any faecal score of 2 or more which is described as watery stool that sifts through bedding. Calves that exhibited a diarrhoea event were tested for infectious agents using a lateral flow test (Farmacy Rainbow Calf Scour Diagnostic Test) designed to detect \textit{Rotavirus}, \textit{Coronavirus}, \textit{E. coli} F5 (K99) and \textit{Cryptosporidium parvum}. Once an appropriate number of the calves developed \textit{C. parvum} infection after week 1 sampling (n=32), healthy control calves were selected from the remaining sampled cohort (n=33). \textit{Cryptosporidium}-positive calves were selected on the basis that they showed clinical signs of diarrhoea and received a positive lateral flow test for \textit{Cryptosporidium parvum} after week 1 sampling. Healthy calves were selected on the basis that they showed no clinical signs of diarrhoeal
disease during the sampling period, though calves with mild respiratory disease signs or who had received routine prophylaxis (Diatrim, Synulox and Halocur) were allowed to be included in the study. The control group was matched to the Cryptosporidium-positive group by age, sex, farm, and breed and as closely matched for date and type of prophylactic treatment as possible. From here onwards, selected calves that did not experience a diarrhoea event will be referred to as the control group (n=33) and calves that experienced a diarrhoea event and received a positive test result for Cryptosporidium after week 1 sampling will be referred to as the Cryptosporidium-positive group (n=32).

**Sample Preparation**

Faecal swab samples were placed directly into bead beating tubes provided in the DNA extraction kit (QIAGEN DNeasy PowerLyzer PowerSoil Kit). Excess plastic applicator was cut using scissors, sterilised with 100% ethanol and a Bunsen burner between samples in order for swabs to fit into the tubes. DNA extraction was performed on all samples following the manufacturers protocol with the following adjustments; 500µL of Powerbead solution was added to each tube along with 60µL of solution C1. Swabs were bead beaten for 15 minutes in a tube adaptor on the Vortex Genie 2 at 7.5 speed. C2 and C3 were mixed 1:1 and 300µL of this solution was added to the sample supernatant and placed at 4°C for 5 minutes. 50µL of C6 Elution Buffer was added to the spin column membrane and incubated at room temperature for 5 minutes to elute the gDNA. Negative extraction controls were provided in the form of empty bead beating tubes and were processed alongside the samples.

DNA was quantified using the Nanodrop and Qubit 3.0 to determine DNA concentration. DNA quality was also determined using the Nanodrop and gel electrophoresis using a 1Kb ladder. Samples with gDNA quality and quantity that did not meet the Centre for Genomic Research (CGR, University of Liverpool) QC requirements were excluded from the study (n=5), resulting in a total of 60 control (n=30) and Cryptosporidium-positive (n=30) samples as well as 3 negative extraction controls.

**Shotgun Metagenomic Sequencing**

60 gDNA samples and 3 negative extraction control samples underwent shotgun metagenomic sequencing and analysis at the Centre of Genomic Research (CGR), University of Liverpool. The Illumina fragment library was prepared from the gDNA samples using the Illumina NEBNext Ultra II FS kit on the Mosquito platform using the 1/10 reduced volume protocol. 50ng of DNA was used as input material where available, followed by size selection of Adaptor-ligated DNA. Following 8 cycles of amplification, the libraries were purified using Ampure XP beads. These final libraries were pooled and the quantity and quality of the pool was assessed by Qubit and the Bioanalyzer and later by qPCR using the Illumina Library Quantification Kit from Kapa on a Roche Light Cycler LC480II according to manufacturer's instructions. After calculation of the molarity using qPCR data, template DNA was diluted to 300pM and denatured for 8 minutes at room temperature using freshly diluted 0.2N sodium hydroxide (NaOH) and the reaction was subsequently terminated by the addition of 400mM TrisCl pH=8. To improve sequencing quality control, 1% PhiX was spiked-in. The libraries were sequenced on the Illumina® NovaSeq 6000 platform (Illumina®, San Diego, USA) following the XP workflow on 2 lanes of an S4 flow cell, generating 2 x 150 bp paired-end reads. See BioProject: PRJNA935534 to access raw sequence data.

**Data Processing**

Initial processing and quality assessment of the sequence data was performed. Briefly, base calling and de-multiplexing of indexed reads was performed by CASAVA version 1.8.2 (Illumina). The resulting raw fastq files were trimmed to remove Illumina adapter sequences using Cutadapt version 1.2.1 [24]. The reads were further trimmed to remove low quality bases, using Sickle version 1.2 with a minimum window quality score of 20 [25]. After trimming, reads shorter than 20 bp were removed. Statistics for the total number of reads obtained for each sample and the distribution of trimmed read lengths for the forward (R1), reverse (R2) and singlet (R0) reads were generated using fastq-stats from EAUtils (Supplementary Fig. 1 and 2) [26].

Prior to analysis, host reads were removed from all samples by aligning reads to the Bos taurus and Homo sapiens combined reference genomes, using the short-read alignment tool, Bowtie2 [27]. The resulting alignment file was processed to extract and retain read pairs where neither read aligned to the host genome, using custom scripts (Supplementary File 1). The retained reads for each sample are shown in Table 1.
Samples underwent taxonomic profiling whereby Kraken2 was used to assign a taxonomic ID to each sequence read and Bracken was used to convert the raw counts into predicted relative abundances for each taxon [28, 29]. Bracken relative abundance tables were parsed by taxonomic rank from species up to phylum level using a custom script (Supplementary File 2). The species relative abundance tables were filtered to 0.1% abundance in at least 1 sample to remove low abundance species (Supplementary File 3).

Prior to functional profiling, read pairs from each sample were analysed to detect overlaps and merged accordingly using PEAR [30]. The samples underwent functional profiling using a MetaPhlAn2 generated relative abundance table (Supplementary File 4) and HUMAnN3 to produce gene family and MetaCyc pathway relative abundances. The gene family relative abundances were converted to GO then GO-Slim term abundances (biological processes, molecular functions, and cellular components) [31-33]. Following processing with HUMAnN3, pathway abundances and GO-Slim terms were renormalised as counts per million reads (CPM).

**Statistical Analysis**

Diversity was measured and plotted using R version 4.2.2 and R packages: tidyverse 1.3.2, vegan 2.6.2, ape 5.6.2 and ggpubr 0.4.0, ggssignif 0.6.4, ggtext 0.1.2, glue 1.6.2, and scales 1.2.1 [34-42]. Comparisons were made between the diversity of the control and Cryptosporidium-positive groups as well as between the swab collection days by grouping the samples into the first half of the week (Day 1-3) and the latter half of the week (Day 4-7). The alpha diversity of samples was measured using species richness and the Shannon index. Normality tests showed that richness data was normal and Shannon diversity data was not normally distributed. The unpaired T-test was used to determine significant differences in species richness between groups. The unpaired Wilcoxon test was applied to determine significant differences in Shannon diversity between groups. Beta diversity of samples was measured using Bray-Curtis PCoA ordination. A PERMANOVA was used to ascertain whether there was a significant distance between centroids.

Taxa relative abundance stacked bar charts were plotted in R version 4.2.2 and R packages: tidyverse 1.3.2, vegan 2.6.2, RColorBrewer 1.1-3, egg 0.4.5, ggtext 0.1.2, and markdown 1.4 [37, 38, 40, 42-45]. The Multivariate Association with Linear models 2 (MaAsLin2) package version 1.8.0 was used to conduct statistical analysis of HUMAnN3, Bracken and MetaPhlAn2 relative abundance outputs. Comparisons between the control and Cryptosporidium-positive groups were performed to reveal any significant taxa or functional data, whilst correcting for confounding variables [46]. Samples were also grouped by covariates including farm (1, 2, and 3), antibiotic/anti-cryptosporidial treatment (Diatrim, Synulox, and Halocur) and swab day within the first week of life (Day 1-7). These covariates were all included as fixed effects. Parameters were kept the same for HUMAnN3, Bracken and MetaPhlAn2 data. The minimum abundance was set to 0.0001 and minimum prevalence was set to 0.1. P-values were adjusted by MaAsLin2 for multiple comparisons using Benjamini-Hochberg procedure (False Discovery Rate). The Q-value cut-off was kept at the default value of 0.25 for taxonomic and functional profiling. Datasets were normalised by Total Sum Scaling (TSS) and the transformation parameter was set to none. Significant results ($q \leq 0.25$) from the MaAsLin2 analysis were visualised using GraphPad Prism 9.3.1 [47].

**Results**

**Microbial diversity does not predict susceptibility to *C. parvum* infection**

DNA extracted from faecal samples collected from 60 ≤1-week-old calves prior to onset of infection underwent shotgun metagenomic sequencing, processing, and taxonomic and functional profiling. Samples were grouped by calves that remained healthy for the duration of the study (Control group; n=30) and calves that displayed clinical signs and tested positive for Cryptosporidium infection following sampling (Cryptosporidium-positive group; n=30). A metagenomic analysis was performed to compare various aspects of the microbiomes of control and Cryptosporidium-positive groups, to determine features associated with susceptibility to infection. Taxonomic profiling down to species level provided species relative abundance tables that were used to measure alpha and beta diversity of the calf faecal microbiome to determine their impact, if any, on susceptibility to bovine cryptosporidiosis. The early microbial diversity between calves was extremely varied. Control and Cryptosporidium-positive groups showed no significant differences in species richness (T-test, p=0.55), Shannon diversity
(Wilcoxon, p=0.96) or beta diversity (PERMANOVA, p=0.5; Fig. 2A-C). However, calves sampled on Day 1-3 versus Day 4-7 had a significant difference in species richness (T-test, p=0.0047), Shannon diversity (Wilcoxon, p=0.0007) and beta diversity (PERMANOVA, p=0.001; Fig. 2D-F). Calves sampled on Day 1-3 exhibited significantly lower alpha diversity compared to calves sampled on Day 4-7. Calves sampled on Day 1-3 and Day 4-7 showed significant dissimilarity in the Bray-Curtis PCoA. This could be attributed to the rapid diversification of the microbiome in the first week of life observed in the existing literature [48, 49].

**Taxa abundance does not predict susceptibility to *C. parvum* infection**

Species relative abundance tables generated by Kraken2/Bracken were parsed by taxonomic rank from species up to phylum level using a custom script and used to compare the relative abundance of different taxa at different taxonomic levels between the control and *Cryptosporidium*-positive groups. The predominant phyla (≥1% relative abundance) present in all samples were *Firmicutes* (30.8%), *Bacteroidetes* (27.7%), *Proteobacteria* (23.4%), *Actinobacteria* (16.1%), and *Fusobacteria* (1.96%). The control group had higher relative abundances of *Bacteroidetes* (31.8% vs 23.7%) and *Actinobacteria* (18.2% vs 14.1%), and lower relative abundances of *Firmicutes* (27.4% vs 34.1%) and *Proteobacteria* (20.2% vs 26.6%) compared to the *Cryptosporidium*-positive group (Fig. 3A). At the genus level, the control group had higher relative abundances of *Bacteroides* (31.6% vs 23.5%) and *Bifidobacterium* (12.5% vs 9.88%), and lower relative abundances of *Escherichia* (16.7% vs 22.9%) and *Faecalibacterium* (9.49% vs 10.8%) compared to the *Cryptosporidium*-positive group (Fig. 3B). However, these differences did not reach statistical significance in the MaAsLin2 analysis (Supplementary Files 5 and 6).

Two taxonomic profiling tools were used to determine the composition of the calf microbiome; Bracken and MetaPhlAn2. When the Bracken species relative abundance data was analysed using MaAsLin2, *Veillonella rodentium* was found to be significantly less abundant in the control group compared to the *Cryptosporidium*-positive group in the Bracken data (q=0.13; Supplementary Fig. 5A; Supplementary File 7). However, the only non-zero relative abundance samples for this dataset were one in the control group versus eight in the *Cryptosporidium*-positive group. MetaPhlAn2 was run as part of the functional profiling pipeline and the relative abundance data was also run through MaAsLin2. *Veillonella sp. CAG 933* relative abundance was found to be significantly lower in the control group compared to the *Cryptosporidium*-positive group in the MetaPhlAn2 data (q=0.19; Supplementary Fig. 5B; Supplementary File 8). The non-zero samples for this species were 13 in the control group and 10 in the *Cryptosporidium*-positive group. Both Bracken and MetaPhlAn2 data have comparable results as the same genus was found to be significant for both datasets. Otherwise, no significant differences in taxa were found in Bracken or MetaPhlAn2 relative abundance data between control and *Cryptosporidium*-positive groups, when adjusting for potential confounding variables using MaAsLin2 (Supplementary Files 7 and 8). This implies that no single taxon was associated with susceptibility to *C. parvum* infection.

**Specific pathway abundances may predict susceptibility to *C. parvum* infection**

Though there were no significant differences in diversity or taxa prior to onset of infection, metagenomic analysis also allows for determination of functional potential of specific taxa. Functional profiling was performed using HUMAnN3 with MetaPhlAn2 and significant differences between the control and *Cryptosporidium*-positive groups in the resulting destratified functional relative abundance tables were determined (Supplementary File 9). The multivariate analysis showed that 12 MetaCyc pathway relative abundances between the control and *Cryptosporidium*-positive groups were significantly different (q≤0.25; Fig. 4, Table 1; Supplementary Fig. 6; Supplementary File 10). The majority of these pathways were related to the methylerythritol phosphate (MEP) pathway for the biosynthesis of isoprenoid precursors. Others were related to purine salvage and degradation, and haem biosynthesis.

**Table 1. Significant pathway relative abundances in the control group (n=30) versus the *Cryptosporidium*-positive group (n=30).**
<table>
<thead>
<tr>
<th>MetaCyc Pathway</th>
<th>Biological Function</th>
<th>Coefficient Value</th>
<th>Associated with susceptibility to <em>C. parvum</em></th>
<th>Taxa contributing to pathway</th>
<th>P-Value</th>
<th>Q-Value (Adjusted P-Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEICHOICACID-PWY: teichoic acid (polyglycerol) biosynthesis</td>
<td>Gram-positive bacterial cell wall biosynthesis</td>
<td>1.81E-04</td>
<td>Yes</td>
<td><em>Escherichia coli, Staphylococcus condimenti, unclassified.</em></td>
<td>0.002293</td>
<td>0.020765</td>
</tr>
<tr>
<td>PWY-5920: superpathway of haem biosynthesis from glycine</td>
<td>Haem is a cofactor for cytochromes</td>
<td>3.20E-04</td>
<td>Yes</td>
<td><em>Escherichia coli, unclassified.</em></td>
<td>0.007606</td>
<td>0.057446</td>
</tr>
<tr>
<td>PWY-7392: taxadiene biosynthesis (engineered)</td>
<td>Terpene produced from MEP pathway metabolites</td>
<td>1.59E-04</td>
<td>Yes</td>
<td><em>Unclassified.</em></td>
<td>0.009125</td>
<td>0.066171</td>
</tr>
<tr>
<td>PWY-7560: methylerythritol phosphate pathway II</td>
<td>MEP Isoprenoid precursor biosynthesis</td>
<td>4.15E-04</td>
<td>Yes</td>
<td><em>Escherichia coli, Klebsiella pneumonia, Proteus mirabilis, unclassified.</em></td>
<td>0.014846</td>
<td>0.102318</td>
</tr>
<tr>
<td>PWY-5695: urate biosynthesis/inosine 5-phosphate degradation</td>
<td>Purine nucleotide degradation.</td>
<td>-6.14E-04</td>
<td>No</td>
<td><em>Alistipes, Allisonella, Anaerostipes, Anaerotignum, Bacteroides, Bibersteinia, Bilophila, Ruminococcus, Catenibacterium, Citrobacter, Clostridioiodes, Clostridium, Desulfovibrio, Enterobacter, Enterococcus, Clostridium, Escherichia, Flavonifractor, Fusicatenibacter, Fusobacterium, Gallibacterium, Hafnia, Intestinibacter, Klebsiella, Kluyvera, Kocuria, Lactococcus, Mannheimia, Megamonas, Megasphaera, Morganella, Parabacteroides, Pasteurella, Prevotella, Proteus, Providencia, Pseudoflavonifractor, Raoultella, Staphylococcus, Streptococcus, Terrisporobacter, Vagococcus, Veillonella, and unclassified spp.</em> (Species are grouped by genera here for brevity).</td>
<td>0.016182</td>
<td>0.110565</td>
</tr>
<tr>
<td>Pathway ID</td>
<td>Description</td>
<td>KEGG Category</td>
<td>KEGG Pathway ID</td>
<td>Isoprenoid Precursor Biosynthesis</td>
<td>Status</td>
<td>Percent Change</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td>--------</td>
<td>---------------</td>
</tr>
<tr>
<td>PWY-5121: superpathway of geranylgeranyl diphosphate biosynthesis II (via MEP)</td>
<td>MEP Isoprenoid precursor biosynthesis</td>
<td>2.39E-04</td>
<td>Yes</td>
<td>Escherichia coli, Klebsiella pneumonia, Proteus mirabilis, unclassified.</td>
<td>0.017217</td>
<td>0.115303</td>
</tr>
<tr>
<td>PWY-6859: all-trans-farnesol biosynthesis</td>
<td>Isoprenoid precursor biosynthesis</td>
<td>1.75E-04</td>
<td>Yes</td>
<td>Citrobacter portucalensis, Citrobacter youngae, Enterobacter cloacae complex, Escherichia coli, Escherichia fergusonii, Klebsiella oxytoca, Klebsiella pneumonia, Morganella morganii, Proteus mirabilis, Proteus vulgaris, Providencia stuartii, unclassified.</td>
<td>0.01747</td>
<td>0.116663</td>
</tr>
<tr>
<td>NONMEVIPP-PWY: methylerythritol phosphate pathway</td>
<td>MEP Isoprenoid precursor biosynthesis</td>
<td>4.72E-04</td>
<td>Yes</td>
<td>Escherichia coli, Klebsiella pneumonia, Proteus mirabilis, unclassified.</td>
<td>0.019814</td>
<td>0.128331</td>
</tr>
<tr>
<td>PWY66-409: superpathway of purine nucleotide salvage</td>
<td>Purine nucleotide salvage</td>
<td>4.86E-04</td>
<td>Yes</td>
<td>Clostridium butyricum, Clostridium perfringens, Clostridium sp 7 2 43FAA, Enterobacter cloacae complex, Enterococcus faecalis, Enterococcus faecium, Escherichia coli, Klebsiella oxytoca, Klebsiella pneumonia, Proteus mirabilis, Providencia stuartii, Streptococcus equinus, Streptococcus galolyticus, Streptococcus pasteurianus, unclassified.</td>
<td>0.023986</td>
<td>0.15325</td>
</tr>
<tr>
<td>PWY-6270: isoprene biosynthesis I</td>
<td>Isoprenoid precursor biosynthesis</td>
<td>2.73E-04</td>
<td>Yes</td>
<td>Escherichia coli, Klebsiella pneumonia, Proteus mirabilis, unclassified.</td>
<td>0.034367</td>
<td>0.206205</td>
</tr>
<tr>
<td>PWY-6383: mono-trans, poly-cis decaprenyl phosphate biosynthesis</td>
<td>Isoprenoid precursor biosynthesis</td>
<td>1.35E-04</td>
<td>Yes</td>
<td>Unclassified</td>
<td>0.035707</td>
<td>0.212223</td>
</tr>
<tr>
<td>PWY-5910: superpathway of geranylgeranyl diphosphate biosynthesis I (via mevalonate)</td>
<td>MVA Isoprenoid precursor biosynthesis</td>
<td>3.49E-05</td>
<td>Yes</td>
<td>Unclassified</td>
<td>0.039785</td>
<td>0.229955</td>
</tr>
</tbody>
</table>

Isoprenoid precursor related pathways are associated with susceptibility to *C. parvum* infection.
Several isoprenoid precursor-associated MetaCyc pathway relative abundances were significantly lower in the control group compared to the Cryptosporidium-positive group (q≤0.25; Fig.4; Table 1). The HUMAnN3 stratified data further shows that *Escherichia coli*, *Klebsiella pneumoniae* and *Proteus mirabilis* species were responsible for the lower abundance of MEP-related pathways in the control group compared to the Cryptosporidium-positive group, though the remainder of the species involved in these pathways were unclassified (Table 1; Supplementary File 11). TEichoicacid-PWY, PWY-7392, PWY-7560, PWY-5121, PWY-6859, NONmevipp-PWY, PWY-6270, and PWY-6383 were the significant pathways related to the MEP pathway (Supplementary Fig. 7) and PWY-5910 was linked to the mevalonate (MVA) pathway.

In addition, the data shows that *Escherichia coli*, *Staphylococcus condimenti* and unclassified species were responsible for the significantly lower abundance of the teichoic acid (poly-glycerol) biosynthesis pathway in the control group compared to the Cryptosporidium-positive group (q=0.02; Table 1; Supplementary File 11).

**Haem biosynthesis pathway is associated with susceptibility to *C. parvum* infection**

The MetaCyc pathway relative abundance, PWY-5920: superpathway of haem biosynthesis from glycine, was significantly lower in the control group compared to the Cryptosporidium-positive group (q=0.06; Fig. 4; Supplementary Fig. 8). This pathway was attributed to *Escherichia coli* (Table 1; Supplementary File 11). Otherwise, the rest of the abundances were unclassified in the stratified data.

**Purine nucleotide salvage is associated with susceptibility to *C. parvum* infection and inosine 5-phosphate degradation is associated with health**

The MetaCyc pathway relative abundance of PWY-5695: urate biosynthesis/inosine 5-phosphate (IMP) degradation was significantly higher in abundance in the control group compared to the Cryptosporidium-positive group (q=0.11; Fig. 4; Supplementary Fig. 9). The PWY66-409: superpathway of purine nucleotide salvage pathway abundance was significantly lower in the control group versus the Cryptosporidium-positive group (q=0.15; Fig. 4; Supplementary Fig. 9). Multiple species, including *Clostridium perfringens*, *Enterobacter cloacae complex*, *Enterococcus faecalis*, *Escherichia coli*, *Klebsiella pneumonia* and *Proteus mirabilis* were responsible for the significantly lower relative abundance of the purine nucleotide salvage pathway in the control group versus the Cryptosporidium-positive group (Table 1; Supplementary File 11). This was also the case for the IMP degradation pathway abundances in the control group, with numerous species of the microbiome contributing to this pathway (Table 1; Supplementary File 11).

**Other variables impact the taxonomic and functional composition of the microbiome**

The factor that had the most profound effect on microbiome species composition was day of sampling. Calves were sampled in the first week of life, however, the day of swabbing within this time frame varied between 1-7 days for the sampled population. Calves that had swabs taken closer to their day of birth (Day 1-3) had significantly different gut microbiomes compared to calves that had swabs taken in the latter part of the week (Day 4-7; Fig. 2D-F, Supplementary Fig. 3).

Though the impact of routine antibiotic/anti-cryptosporidial use on the calf microbiome was not the focus of this study, the inclusion of calves that had been treated prior to sampling was unavoidable. This was taken into consideration during selection of controls which were matched as closely as possible to Cryptosporidium-positive calves by prophylactic treatment which included Diatrim, Synulox and Halocur, though within farm treatment was not always consistent. To mitigate their effect, these covariates were included as fixed effects in the MaAsLin2 analysis so that any significant differences between the control and Cryptosporidium-positive groups were a consequence of prospective infection rather than antimicrobial treatment. Regardless of taking treatments into account, the data showed that the composition and metabolic potential of the microbiome were significantly affected by these routine prophylaxes (Supplementary Files 7 and 10).

No significant differences were observed between controls and Cryptosporidium-positive calves for the other destratified functional categories that were profiled in HUMAnN3, including biological processes, cellular components, and molecular functions (Supplementary Files 12-14).
Discussion

Understanding the features of the calf faecal microbiome that contribute to Cryptosporidium susceptibility could inform the development of new therapies and preventive strategies against infection in cattle. Here we conducted a retrospective case-control study in which shotgun metagenomic sequencing was used to determine the taxonomy and functional potential of the faecal microbiome in control and Cryptosporidium-positive neonatal calves, prior to infection.

Due to the unstable nature of the early microbiome, there is a lot of microbial variability between calves in the first week of life [48, 49]. This trend was observed in our data, as no taxon was significantly more abundant between the calf faecal microbiomes of the control and Cryptosporidium-positive groups. Despite no significant differences in diversity or taxa relative abundance between the control and Cryptosporidium-positive groups, the general composition of the calf faecal microbiota followed the patterns seen in other studies investigating the early calf microbiome. For example, the predominant phyla in the first week of life were Firmicutes and Proteobacteria which were observed in other studies of the neonatal calf microbiome [48, 49].

Though the faecal microbiota did not directly predict the susceptibility of calves towards cryptosporidiosis, the multivariate analysis revealed that specific pathways were associated with the Cryptosporidium-positive group. These pathways were related to isoprenoid precursor biosynthesis, haem biosynthesis and purine salvage. The majority of these pathways were attributed to Escherichia coli, Klebsiella pneumoniae and Proteus mirabilis; all species belonging to the Enterobacteriaceae family. Though this family was not significantly more abundant in the Cryptosporidium-positive group, Enterobacteriaceae has been shown to be associated with diarrheal disease in calves [50]. It is likely that this trend was not observed in our data as the samples were collected before onset of infection.

Though the effect size (coefficient) of the significant pathways was small, it is striking that all of the pathways are also absent in Cryptosporidium parasites due to the lack of an apicoplast and traditional mitochondria. With this in mind, components of these microbial pathways could potentially be exploited as targets in the development of novel therapies or preventatives against bovine cryptosporidiosis.

This study showed that the control group had a lower relative abundance of isoprenoid precursor biosynthesis-related pathways in comparison to the Cryptosporidium-positive group suggesting a higher abundance of microbial isoprenoid precursors may lead to increased susceptibility to C. parvum infection. There are two isoprenoid precursors; isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP), that make up a wide variety of biological molecules that are essential for cellular growth in all living organisms. The majority of the significant pathways relate to the MEP pathway which is one of two pathways responsible for the production of isoprenoid precursors. The MEP pathway is the method by which most Gram-negative bacteria, and some Gram-positive bacteria and eukaryotes produce isoprenoid precursors [51]. These compounds are used in the biosynthesis of 2-methyl-1,3-butadiene, also known as isoprene. Isoprene is found in myriad isoprenoid compounds including sterols like cholesterol, vitamins A and D, carotenoids, and haem A [52, 53]. In addition, the teichoic acid (glycerol) biosynthesis pathway includes an interaction with isoprenoids in order to synthesise teichoic acid, a structural component of Gram-positive bacteria cell walls [54].

The MEP pathway takes place in the apicoplast of apicomplexans such as Plasmodium and Toxoplasma [55]. However, Cryptosporidium lacks an apicoplast and as a result is void of the MEP pathway [56]. Though the MEP pathway is absent in C. parvum parasites, it has been shown that the parasite encodes enzymes connected to the use of isoprenoid precursors, indicating that Cryptosporidium must scavenge the isoprenoid precursors from an external source [57]. Some have suggested that Cryptosporidium exploits the production of isoprenoid precursors from the mammalian host cells which are generated via the MVA pathway [55]. The MVA pathway is the method by which mammals and most Gram-positive bacteria produce isoprenoid precursors. Indeed, a component of the bacterial MVA pathway was also significantly lower in the control group versus the Cryptosporidium-positive group in our study. This suggests that C. parvum may exploit both MEP and MVA pathways of the bacterial microbiota. But whether Cryptosporidium scavenges isoprenoid precursors from the host, the microbiome or both is unknown. In fact, the inhibition of the MVA pathway of host cells in vitro has been shown to reduce growth of C. parvum infection in HCT-8 cells using the statin, Itavastatin [58]. This outcome in conjunction with the results of our study would suggest...
that *C. parvum* may use a combination of host MVA and microbial MEP and MVA pathways in order to scavenge sufficient supplies of IPP. If this were the case, this may imply that the difficulties of culturing *Cryptosporidium in vitro* are due to a lack of bacterial isoprenoid precursors to scavenge and thus introducing these isoprenoid precursors could improve *in vitro* infection rates for experimental research.

The superpathway of haem biosynthesis was significantly lower in abundance in the control group compared to the *Cryptosporidium*-positive group. This suggests that calves with a higher relative abundance of microbial haem pathways are more susceptible to *C. parvum* infection. Similarly to the apicoplast, a traditional mitochondrion is also lacking in *Cryptosporidium*, along with the ability to synthesise haem. Unlike other apicomplexan parasites which exhibit multiple cytochromes, *Cryptosporidium* expresses one haem-containing enzyme of unknown function, suggesting that *Cryptosporidium* has some requirement for haem though it may be minimal [59]. A possible function of this singular enzyme could be sterol manufacture as this is the only process that is utterly haem-dependent and found in most eukaryotes [60]. As previously mentioned, sterol production requires isoprenoid precursors, meaning that haem interacts indirectly with the MEP/MVA pathway. Indeed, haem B, may be converted to other haem derivatives such as haem A and O by transfer of farnesyl groups, a product of the MEP pathway, illustrating another pathway that interacts with isoprenoid metabolites [52]. Though haem requirement of *Cryptosporidium* may be minimal, this singular enzyme could be inhibited to reduce *Cryptosporidium* infection in calves.

The purine salvage pathway was found to be significantly lower in abundance in the control group, whereas the IMP degradation pathway was significantly higher in the control group compared to the *Cryptosporidium*-positive group. Purine nucleotides are essential for the survival of *C. parvum* as like any living organism, they require an energy source and the constituents to assemble DNA and RNA. However, *C. parvum* is not able to synthesise de novo purines and does not possess its own mechanism of purine salvage. It has been proposed that *C. parvum* is able to take advantage of host cell purine salvage pathways in order to take purine nucleotides from the cytoplasm of the host cell in which it resides [61]. Though *C. parvum* may manipulate the host to exploit its purine salvage pathways, our data suggests that calves that have a higher relative abundance of purine salvage pathways within the microbiota are at higher risk of becoming infected with *C. parvum*. This implies that *Cryptosporidium* may use the purine salvage pathways of the microbiome in conjunction with host salvage mechanisms. Therefore, we postulate that the bacterial purine salvage pathway could be targeted to inhibit *C. parvum* infection.

A study investigating the effect of purine nucleosides on *in vitro* *C. parvum* infection showed that inosine improved the growth of the parasite in THP-1 cells, particularly the trophic stages. This shows the importance of purine metabolism in *C. parvum* [62]. It has been suggested that inhibition of activities in the pathway between adenosine and guanosine monophosphate (GMP) in *Cryptosporidium* would lead to killing of the parasite as it relies upon this single pathway to produce GMP [63]. Comparable to haem, it appears that *Cryptosporidium* has simplified its mechanism for purine procurement. The enzyme, IMP dehydrogenase (IMPDH), catalyses the rate-limiting step that converts exogenous purines such as adenosine into GMP [64]. Therefore, it has been proposed that IMPDH could be a potential candidate for drug development against *C. parvum* infection.

These findings lead us to suggest potential therapeutic or preventative strategies against *Cryptosporidium* infection such as compounds that directly inhibit these microbial pathways or probiotics/FMT therapies that reduce the proportion of microbes contributing to them. If effective, the main dilemma of inhibiting microbial pathways or manipulating the microbiome in any way is the potential negative impact on the microbiome, and in turn the host. However, further research is required to explore these recommendations.

**Limitations**

The study design inherently has limitations. We were only able to show associations between susceptibility to infection and features of the microbiome since the study was observational. Consequently, any conclusions drawn from this study will require further investigation. *In vitro* research in a *C. parvum*-bacteria-host cell co-culture system with pathway inhibitors could be a possible approach to ascertaining the importance of bacterial pathways in *C. parvum* infection.
The provision of antibiotics and other treatments at birth to several enrolled calves was unavoidable as the study was conducted on commercial dairies. The compositional and functional differences of the microbiome between control and *Cryptosporidium*-positive calves may have been more pronounced without these routine treatments as their ability to alter the microbiome may have had a masking effect. Though it is highly likely that the results of the study may have differed had the calves not received treatment, this aspect of the study is also a strength as it improves the external validity of the study findings.

**Conclusion**

In summary, we conclude that *C. parvum* may be able to harness the isoprenoid precursor biosynthesis, haem biosynthesis and purine salvage pathways of the host microbiota in order to survive and calves that are more abundant in these microbiota-associated pathways may be more susceptible to *Cryptosporidium* infection. This could be important for development of novel treatments or preventative strategies against bovine cryptosporidiosis as components of these pathways could be exploited as potential drug targets.

**Abbreviations**

- BCS – Body Condition Score
- CGR – Centre of Genomic Research
- MEP - Methylerythritol phosphate
- MVA – Mevalonate
- IMP – Inosine 5-monophosphate
- IPP - Isopentenyl diphosphate
- DMAPP – Dimethylallyl diphosphate
- GMP - Guanosine monophosphate
- IMPDH - Inosine 5-monophosphate dehydrogenase

**Declarations**

**Ethics approval**

The study was conducted following ethical approval by the University of Liverpool Research Ethics Committee (VREC927), and procedures regulated by the Animals (Scientific Procedures) Act were conducted under a United Kingdom (UK) Home Office Licence (P191F589B).

**Consent for publication**

Not applicable

**Availability of data and material**

The raw sequence data can be found in the NCBI repository in BioProject: PRJNA935534.

**Competing interests**

Not applicable
**Funding**

The sequencing was funded by the Wellcome Trust Institutional Strategic Support Fund: Veterinary Research Career Development Clinical Leave Fellowships. All other work for this study was funded by the BBSRC NLD DTP (BB/M011186/1) and BBSRC Tools and Resources Development Fund Award (BB/M019071/1).

Dr Stewart is supported by the Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 221745/Z/20/Z), a Newcastle University Academic career Track (NUAcT) Fellowship, and the 2021 Lister Institute Prize Fellow Award. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

**Authors’ contributions**

MFH was involved with the study design and managed the sample processing and QC, as well as diversity, taxonomy, and MaAsLin2 analyses. MFH also prepared figures 1-4, supplementary figures 3-9, table 1, supplementary table 1, and wrote the main manuscript text. BEG conducted the faecal sample collection, storage, and transport and provided all relevant metadata. FJ generated the sequencing libraries. CN sequenced the libraries on the Illumina® NovaSeq 6000 platform (Illumina®, San Diego, USA). SH performed the sequence processing and taxonomic/functional profiling as well as MaAsLin2 analysis. SH provided custom scripts and generated supplementary figures 1-2. JSD provided the Wellcome Trust funding for the sequencing. CJS, JSD, JLC, and GO all contributed to the study design and provided guidance on the sample processing, data analysis, and writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

Not applicable

**References**


47. Graphpad Software: GraphPad Prism. 9.3.1 edition.


Figures
**Figure 1**

**Experimental study design.** Faecal swab samples were collected from calves during week 1 of life from 3 farms. Calves were observed for signs of diarrhoeal disease amongst other health monitoring checks. Calves that exhibited diarrhoea had a lateral flow test to determine the cause of infection and another swab was taken at the point of scour. Healthy control calves (n=33) and calves that tested positive for *C. parvum* (n=32) were selected matching for age, sex, breed, and prior treatment where possible. Week 1 faecal swab samples from the selected control and *Cryptosporidium*-positive calves were extracted and any DNA that did not meet CGR QC requirements was excluded from the study. The final 60 DNA samples underwent shotgun metagenomic sequencing, processing, and analysis.
Figure 2

Alpha and beta diversity of control versus *Cryptosporidium*-positive and Day 1-3 versus Day 4-7 sampling groups. **A)** Species richness of control (n=30) and *Cryptosporidium*-positive (n=30) groups. **B)** Shannon index of control (n=30) and *Cryptosporidium*-positive (n=30) groups. **C)** Bray Curtis PCoA ordination plot of the control (n=30) and *Cryptosporidium*-positive (n=30) groups. **D)** Species richness of calves sampled on Day 1-3 (n=20) versus Day 4-7 (n=40). **E)** Shannon index of calves sampled on Day 1-3 (n=20) versus Day 4-7 (n=40). **F)** Bray Curtis PCoA ordination plot of calves sampled on Day 1-3 (n=20) versus Day 4-7 (n=40). Plotted in RStudio.
Figure 3

Mean relative abundance of the microbial composition among the control and Cryptosporidium-positive groups. **A)** Phyla relative abundance (≥1%) in the control group (n=30) versus the Cryptosporidium-positive group (n=30). **B)** Genera relative abundance (≥1%) in the control group (n=30) versus the Cryptosporidium-positive group (n=30). Plotted in RStudio.
Figure 4


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1TaxonomicandFunctionalProfilingScript.txt
Supplementary File 2: Taxonomic Rank Parse Script.txt
Supplementary File 3: Bracken Species Relative Abundance Table.txt
Supplementary File 4: Metaphlan2 Species Relative Abundance Table.txt
Supplementary File 5: Bracken Phyla Maaslin2 Results Table.txt
Supplementary File 6: Bracken Genus Maaslin2 Results Table.txt
Supplementary File 7: Bracken Species Maaslin2 Results Table.txt
Supplementary File 8: Metaphlan2 Species Maaslin2 Results Table.txt
Supplementary File 9: De-stratified Pathway Abundance Table.txt
Supplementary File 10: Pathway Abundance Maaslin2 Results Table.txt
Supplementary File 11: Stratified Pathway Abundance Species Assignment Table.txt
Supplementary File 12: Biological Processes Maaslin2 Results Table.txt
Supplementary File 13: Cellular Components Maaslin2 Results Table.txt
Supplementary File 14: Molecular Functions Maaslin2 Results Table.txt
Supplementary File 15: Metadata File.txt
Supplementary File 16: R Analysis Code.html
Supplementary Tables and Figures.pptx