nature communications

Article

https://doi.org/10.1038/s41467-024-54995-4

Multicopy subtelomeric genes underlie animal infectivity of divergent *Cryptosporidium hominis* subtypes

Received: 22 November 2023

Accepted: 27 November 2024

Published online: 30 December 2024

Check for updates

Wanyi Huang[1,2], Wei He[1,2], Yue Huang[1,2], Yongping Tang[1], Ming Chen[1], Lianbei Sun[1], Zuwei Yang[1], Tianyi Hou[1], Huimin Liu[1], Haoyu Chen[1], Tianpeng Wang[1], Na Li[1], Yaqiong Guo[1] ✉, Lihua Xiao[1] ✉ & Yaoyu Feng[1] ✉

The anthroponotic *Cryptosporidium hominis* differs from the zoonotic *C. parvum* in its lack of infectivity to animals, but several divergent subtypes have recently been found in nonhuman primates and equines. Here, we sequence 17 animal *C. hominis* isolates and generate a new IbA12G3 genome at the chromosome level. Comparative analysis with 222 human isolates shows significant genetic divergence of the animal isolates, with genetic recombination among them. They have additional subtelomeric insulinase and *MEDLE* genes. In interferon-γ knockout mice, three monkey isolates show differences in infectivity and induce higher and longer oocyst shedding than a reference *C. parvum* isolate. Deletion of the *MEDLE* genes significantly reduces the growth and pathogenicity of a virulent strain in mice. Co-infection of two fluorescence-tagged *C. hominis* subtypes produces bicolored oocysts, supporting the conclusion that mixed subtype infections can lead to genetic recombination. These data provide insight into potential determinants of host infectivity in *Cryptosporidium*, and a convenient animal model for biological studies of *C. hominis*.

Cryptosporidium spp. are the leading cause of diarrhea-related deaths in humans and various animals[1]. Of the 47 established *Cryptosporidium* species, *Cryptosporidium hominis* and *Cryptosporidium parvum* are the two most important species infecting humans[2]. The two species differ in genetic characteristics (such as the copy number of some subtelomeric genes) and in their ability to experimentally infect mice, rats, and calves[3]. *C. hominis* was once thought to infect only humans, but has been increasingly reported in animals[4]. Recent data indicate that nonhuman primates (NHPs), horses, and donkeys are natural hosts for *C. hominis*, with infection rates as high as 5.9% in farmed macaque monkeys and 13.6% in farmed donkeys[5–10].

To date, more than 10 *C. hominis* subtype families have been recognized based on sequence analysis of the 60-kDa glycoprotein (*gp60*) gene. Most of them show a preference for specific hosts[11]. Common human subtype families include Ia, Ib, Id, Ie, If, and Ig. They have been found in NHPs, but only occasionally in other animals[4]. Among others, Ik is a subtype family adapted to equine animals, while Ii, Im, and In are adapted to NHPs[12]. As NHPs are genetically related to humans, they have been considered as potential reservoirs for human-pathogenic *Cryptosporidium* species[13]. In recent years, both NHP-adapted subtype families (including IiA17, ImA18, ImA13, InA14, InA17, and InA26 subtypes) and human-pathogenic subtype families (including IaA13R7, IaA13R8, IaA14R7, IbA12G3, IdA20, IdA14, IeA11G3T3, IfA16G2 subtypes) have been detected in NHPs. Importantly, the equine-adapted Ik and NHP-adapted Ii subtypes have been found in humans in Europe, and the ImA13 subtype caused a human cryptosporidiosis outbreak in Kenya in 2022, suggesting a potential expansion of the host range of the equine- and NHP-adapted *C. hominis*

[1]State Key Laboratory for Animal Disease Control and Prevention, Center for Emerging and Zoonotic Diseases, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China. [2]These authors contributed equally: Wanyi Huang, Wei He, Yue Huang. ✉e-mail: guoyq@scau.edu.cn; lxiao1961@gmail.com; yyfeng@scau.edu.cn

to humans[14–16]. These animal-adapted subtype families of *C. hominis* have minor sequence differences from the human-pathogenic subtypes at the 18S rRNA locus[12]. However, sequence heterogeneity between different copies of the 18S rRNA gene is present in human *C. hominis* isolates and other *Cryptosporidium* species[17]. Therefore, the genetic and biological characteristics of these animal-adapted subtypes are poorly understood.

In this work, we obtain whole-genome sequences (WGS) from 17 NHP and equine-adapted isolates of *C. hominis*. We generate a chromosome-level genome of a monkey *C. hominis* IbA12G3 strain using long-read (PacBio) and short-read (Illumina) sequencing technologies and perform comparative genomic analyses of the WGS data along with those of 222 human isolates to understand the genetic characteristics of animal-adapted *C. hominis* subtypes. Because these isolates have additional subtelomeric *MEDLE* and insulinase (*INS*) genes that are normally found in *C. parvum*, we evaluate the susceptibility of interferon-γ knockout (GKO) mice to the NHP-adapted isolates. We knock out the two extra copies of the *MEDLE* genes to test their biological significance in *C. hominis* infectivity and pathogenicity. To demonstrate that these isolates are genetically tractable and to provide experimental support for the occurrence of genetic recombination in *C. hominis*, we also fluorescently tag two isolates with different infectivity and virulence and generate a genetic cross of them. Our results reveal the genomic characteristics of animal-adapted *C. hominis* isolates and the role of genetic recombination in their evolution. They provide insight into potential genetic determinants of host infectivity and virulence in *Cryptosporidium* spp. and a laboratory animal model for biological studies of *C. hominis*.

Results

A new chromosome-level reference genome of *C. hominis*

To gain insight into the genetic characteristics of *C. hominis*, we first generated a new *C. hominis* genome using the IbA12G3_46287 strain isolated from NHPs. The new reference genome had no gaps or ambiguous bases and contained 8 chromosomes with 13 telomeric ends with the AAACCT or AGGTTT repeats, except for the 5′ end of chromosome 1 and the 3′ end of chromosomes 7 and 8 (Supplementary Table 2). The chromosome-level IbA12G3_46287 genome has a genomic structure similar to the *C. parvum* reference genomes (IOWA-CDC and IOWA-ATCC)[18,19]. The unassembled chromosome ends were located upstream or downstream of the rRNA unit, which is more than 7000 bp in length. In long-read mapping, we observed an increase in read depth at the fully resolved 5′ end of chromosome 7, which contains one rRNA unit, with some reads shown as white blocks (i.e., can be aligned elsewhere), indicating the high sequence identity nature of the multicopy rRNA units (Supplementary Fig. 1a). We identified five rRNA units on chromosomes 1, 2, 7, and 8 in the new IbA12G3 genome. Between them, the 18S, 5.8S, and 28S rRNA genes were conserved in length and sequence between copies, except for the unit on chromosome 2, which had the B-type RNA sequence previously described in *C. parvum* IOWA-CDC strain[12]. This rRNA unit had very different *ITS1* and *ITS2* sequences (Supplementary Table 3).

Interestingly, a similar increase in read depth was observed in an upstream region of approximately 15 kb in length at the 5′ end of chromosome 7, which contains CHM_7g5511, CHM_7g5501, and CHM_7g4601 (orthologs of CPCDC_5g5511, CPCDC_5g5500, and CPCDC_7g4601 in *C. parvum*). Several types of reads were mapped to this region, likely indicating the collapse of some similar sequences (Supplementary Fig. 1a). Since the compressed regions contain both the rRNA unit and the subtelomeric *INS* gene (Supplementary Fig. 1a), these five sequence types could be associated with the five rRNA units. A T2T *C. parvum* genome (IOWA-BGF)[20] has recently been published, showing four 35 kb region duplications (sequence identity > 97%) that have been tentatively located at the 5′ ends of chromosomes 1 and 7 and the 3′ ends of chromosomes 7 and 8. Given the structural collinearity between the *C. parvum* and *C. hominis* genomes, this finding suggests that the collapsed regions in the *C. hominis* IbA12G3_46287 genome may contain sequences of the missing telomeric ends. Notably, the newly generated reference genome was comparable in completeness to the *C. parvum* reference genomes IOWA-CDC and IOWA-BGF (Supplementary Fig. 1b).

Genetic divergence of animal-adapted subtypes from human *C. hominis* subtypes

Using the new *C. hominis* reference genome, more than 371,000 core wgSNPs were detected between *C. hominis*, *C. parvum*, *C. tyzzeri*, *C. meleagridis*, and *C. cuniculus*. (the gnomonic regions for core wgSNPs are described in Supplementary Table 4). In contrast, approximately 15,500 core wgSNPs were present between the NHP-adapted and human-adapted *C. hominis* isolates, which was slightly less than the 16,900 core wgSNPs between the latter and equine-adapted isolates (Supplementary Data 2). This was higher than the maximum of 4260 core wgSNPs between human isolates of different geographical origin, but lower than the approximately 20,000 core wgSNPs between typical anthroponotic IIc and zoonotic IIa and IId subtypes of *C. parvum*. The clades of the maximum likelihood (ML) tree were segregated by *Cryptosporidium* species, with all *C. hominis* isolates clustering in a clade separate from *C. cuniculus*, *C. parvum*, *C. tyzzeri*, and *C. meleagridis* isolates (Supplementary Fig. 2a). These data suggest that animal-adapted and human-adapted *C. hominis* are genetically related, but there is host-source-associated genetic diversity in *C. hominis* genomes, and host-related differences are much greater than geographic differences in *C. hominis*.

We investigated the population structure of *C. hominis* using a total of 35,248 core wgSNPs in *C. hominis* isolates. They formed six major clusters in pairwise ANI comparisons. Among them, monkey and equine isolates formed their own clusters, with low nucleotide identities between the two groups and with human isolates (Fig. 1a). Consistent with these findings, the ML tree showed the formation of six clades that were largely segregated by the host. In particular, the monkey and equine *C. hominis* isolates had long branches, suggesting that they are divergent from the remaining isolates (Fig. 1b). Furthermore, the monkey and equine isolates were divergent from the human isolates in the PCA analysis, and there was no apparent gene flow between these groups in the phylogenetic network analysis (Supplementary Fig. 2b, c). In contrast, human *C. hominis* isolates showed high nucleotide identity (>99.9%) and frequent gene flow between them[21].

Population division in animal-adapted *C. hominis* isolates

In the PCA analysis, the genomes of the animal isolates formed four major clusters, with Ik, Ii, and most Im isolates forming their own clusters. However, isolates of several subtypes, including an IbA12G3 isolate, formed the fourth cluster (Fig. 1c). Pairwise *dxy* analysis showed that all IkA16G1 isolates from donkeys had high genetic similarity, while most of the IiA17 and ImA18 isolates from monkeys most showed high sequence identity within each subtype except for IiA17_45372 and ImA18_37833 (Fig. 1d). In contrast, InA14_45028, InA17_46269, ImA18_37833, ImA20_46260, and IbA12G3_46287 showed more sequence identity to each other than to other subtypes within the same subtype family. These data suggest that the animal isolates had different ancestral origins.

Gene flow between animal isolates

We performed identity-by-descent (IBD) analysis to further investigate the ancestral relationships among the animal isolates. Shared IBDs (with mean sharing fractions greater than 80%) were present in the following groups: (1) all three IkA16G1 isolates; (2) all five IiA17 isolates; (3) most ImA18 isolates; (4) ImA18_37833 and InA14_45028; (5) InA17_46269, ImA20_46260, and IbA12G3_46287 (Fig. 2a). This suggests that isolates within each group are genetically related and that there might be some

https://doi.org/10.1038/s41467-024-54995-4

gene flow between isolates in the group with multiple subtypes. In the phylogenetic network analysis, these animal isolates formed four major clusters with parallel edges between clusters of monkey isolates, suggesting the occurrence of gene flow (Fig. 2b). Results of the STRUCTURE analysis revealed the presence of genome admixture in IiA17_45372, InA14_45028, InA17_46269, ImA18_37833, ImA20_46260, and IbA12G3_46287, confirming the presence of genetic recombination among the isolates (Fig. 2c and Supplementary Fig. 2d).

Origin of IbA12G3 subtype in NHPs

The *gp60* subtype IbA12G3 is a member of the human-pathogenic Ib subtype family but has only recently been detected in NHPs. The monkey IbA12G3 isolate had 15,364 core SNPs compared to a human IbA12G3 isolate, which was much higher than the 2006–3905 SNPs between the latter and other human isolates (Fig. 3a and Supplementary Data 2). In contrast, InA17_46269, ImA20_46260, and monkey IbA12G3 had high IBD values and low *dxy* values in most regions of the eight chromosomes, indicating that these genomes were highly related (Fig. 2a and Supplementary Fig. 3). In addition, the results of the phylogenetic network analysis showed the presence of gene flow from InA17_46269 and ImA20_46260 to monkey IbA12G3 (Fig. 2b).

Phylogenetic topology weighting was performed on the genomes of InA17_46269, ImA20_46260, monkey IbA12G3, and human IbA12G3. This confirmed the genetic relatedness of InA17_46269 and monkey

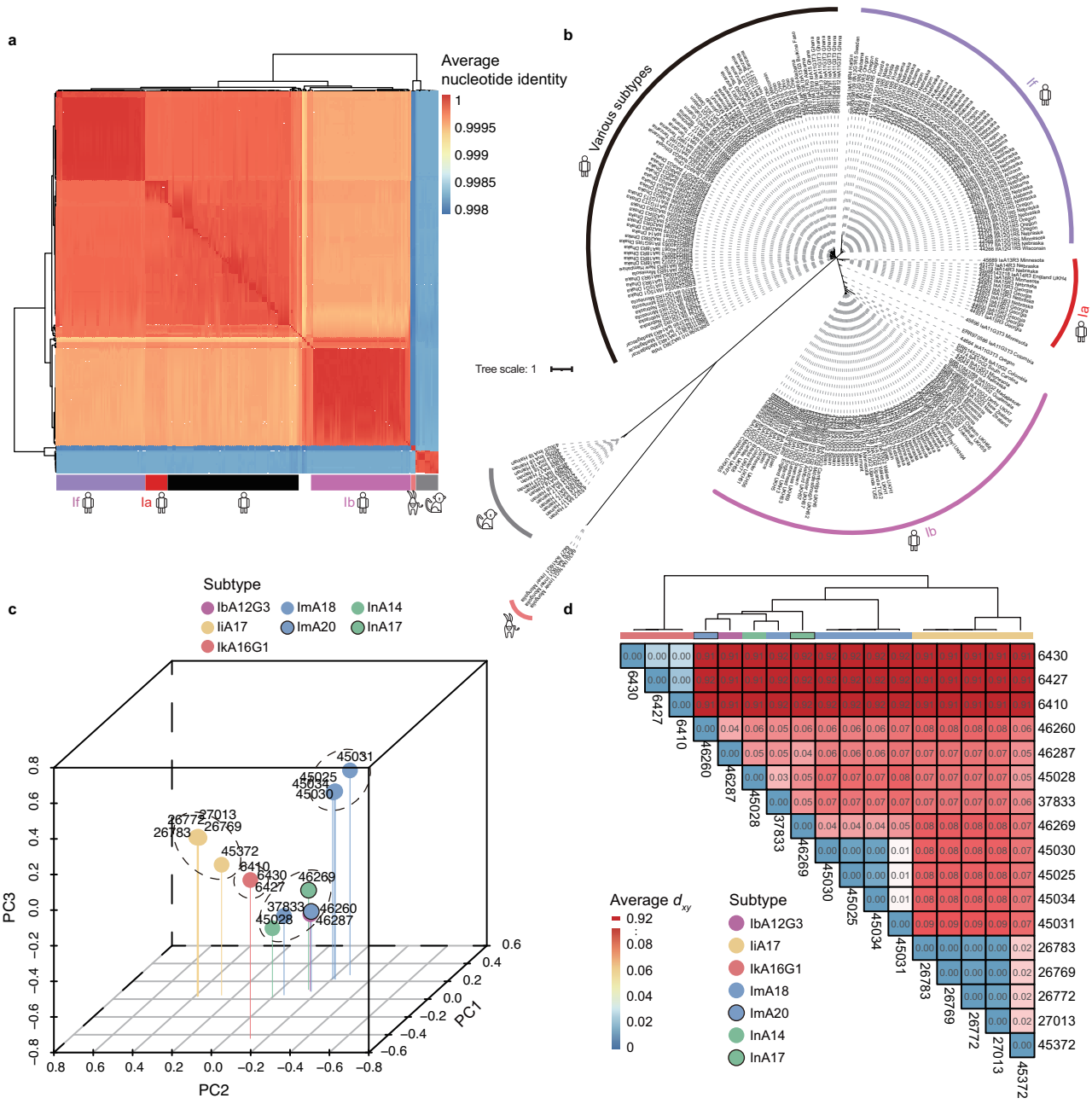


Fig. 1 | Genetic divergence of *Cryptosporidium hominis* isolates from monkeys and donkeys. a Pairwise comparisons of average nucleotide identity (ANI) of *C. hominis* genomes. **b** Phylogenetic relationships of *C. hominis* isolates inferred by maximum likelihood analysis of 35,248 core wgSNPs. These *C. hominis* isolates (colored by subtype family) formed 6 major clades. Among them, two divergent clades contain the monkey and equine isolates. **a**, **b** Hosts and subtype families are marked with different colors (see figure key). **c** Principal component analysis (PCA) of animal isolates based on pruned SNPs, in which PC1, PC2, and PC3 account for variability among isolates. **d** Mean absolute divergence (*dxy*) between pairs of *C. hominis* genomes. Low *dxy* values indicate that the genomes are closely related. **c**,**d** Subtypes are marked with different colors (see figure key). Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-54995-4

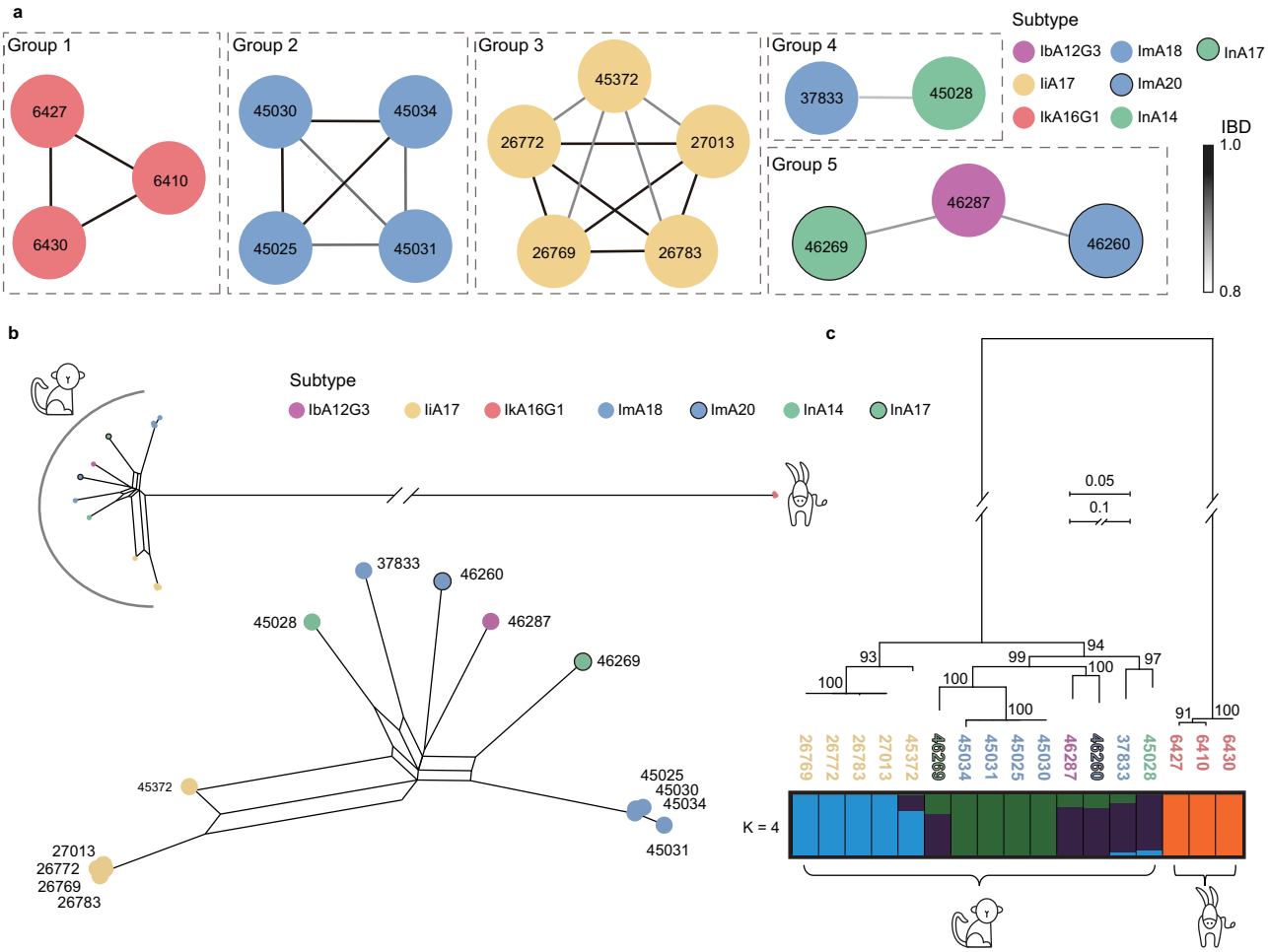


Fig. 2 | Gene flow among animal-adapted *Cryptosporidium hominis* isolates. **a** Relatedness network for pairs of isolates inferred using identity-by-descent (IBD) analysis. Nodes represent isolates (colored by subtype family), while edges between nodes indicate IBD sharing. All edges with IBD higher than the threshold (IBD = 0.8) are shown. **b** A phylogenetic network of animal isolates based on 15,886 core wgSNPs. The parallel edges in the network indicate gene flow between isolates. **c** Phylogeny and population structure of the divergent *C. hominis* isolates with $K = 4$. The rooted ML tree was constructed using the TVM + G model implemented using 15,886 core wgSNPs. The major hosts of the clades are marked at the bottom panel and the isolates are colored by subtypes as in panel A. Each isolate is represented by a single bar in 4 colored ancestral components depending on the K values used. Other results with K values of 2 to 7 are shown in Figure S2d. **a**–**c** Subtypes are marked with different colors (see figure key). Source data are provided as a Source Data file.

IbA12G3, with the average weighting of the two as sister populations (topo4) accounting for 60% of the genome (Fig. 3b). The presence of other topologies indicated the occurrence of sequence introgression from other sources (Fig. 3b, c). Consistent with these findings, low *dxy* values were observed between monkey IbA12G3 and InA17_46269 in large regions of chromosomes 1 and 5, between monkey IbA12G3 and ImA20_46260 on chromosomes 4 and 8, and between monkey and human IbA12G3 at the *gp60* locus (Fig. 3d, colored boxes). In addition, the monkey IbA12G3 showed substantial *dxy* values in some areas of the genome (such as the 650–750 kb region on chromosome 1), suggesting the presence of additional introgression of other exogenous sequences into this isolate.

The two IbA12G3_human isolates were collected in North America, while the monkey IbA12G3 isolate was found in China. Therefore, we also tested whether the Asian human isolates played a role in shaping the monkey IbA12G3 genome. The ML tree showed a very long branch for the isolate from a Chinese human and the formation of a cluster with the If subtype family, while the Dhaka isolates formed its subclade located in a mixed population with other human isolates from around the world (Fig. 1b). The data indicated that the Chinese isolate shared sequences with the US isolates, while the Dhaka isolates were segregated by geographic origin[21]. Phylogenetic topology weighting was performed on the monkey IbA12G3, human IbA12G3, and the Chinese IaA18R4 genomes, as well as on the monkey IbA12G3, human IbA12G3, and the Dhaka Ib genomes. The data showed that the genetic relatedness of human and monkey IbA12G3 was higher than that of monkey IbA12G3 and Asian isolates (Supplementary Fig. 4). The fewer SNPs and lower *dxy* values between the human and monkey IbA12G3 than between the latter and the Asian isolates further supported this finding (Supplementary Data 2 and Supplementary Fig. 3). These data suggest that the current Asian human isolates are divergent from the monkey isolates, despite being from the same geographic region.

Genetic recombination in other monkey *C. hominis* isolates

The IBD data suggested that InA17_46269 and ImA20_46260 were under the influence of IbA12G3_46287 (Fig. 2a). Average weighting, *fd*, and *dxy* analyses confirmed the occurrence of ImA18 sequence introgression in the InA17_46269 genome, and ImA18_37833 sequence introgression in the ImA20_46260 genome (Supplementary Figs. 3 and 5). Specifically, the former involved nucleotides 550 k–1000 k on chromosome 8, while the latter involved nucleotides 500 k–750 k on chromosome 2 (Supplementary Fig. 5).

Isolate 37833 was an ImA18 subtype at the *gp60* locus, but was genetically related to InA14_45028 based on the IBD analysis (Fig. 2a).

https://doi.org/10.1038/s41467-024-54995-4

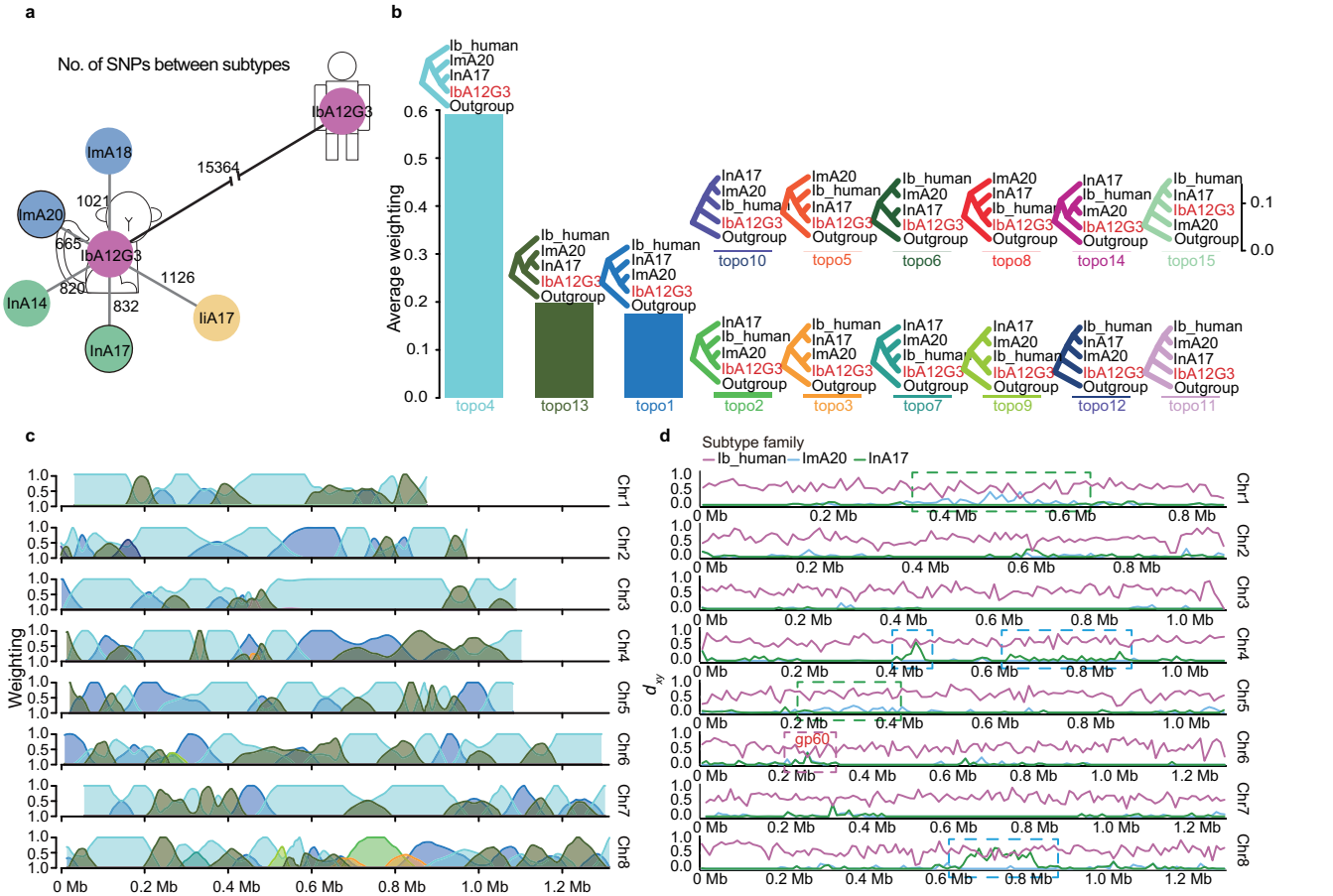


Fig. 3 | Origin of *Cryptosporidium hominis* IbA12G3 subtype from monkeys. **a** Overall genetic differences between monkey and human IbA12G3 isolates based on the number of core wgSNPs shared between isolates. **b** Genome-wide distribution of phylogenetic relationships among the monkey IbA12G3, ImA20_46260, InA17_46269, and human IbA12G3 using a 50-SNP sliding window with IkA16G1 as the outgroup. The top panel shows all possible topologies, while the bottom panel shows the genome-wide average weighting of each topology. **c** Distribution of topology weightings (colors as in B) across eight chromosomes. **d** Distribution of *dxy* values between the monkey IbA12G3 genome and the ImA20_46260, InA17_46269, and human IbA12G3 genomes across the eight chromosomes. The green and blue boxes show regions with high sequence identity between monkey IbA12G3 and InA17_46269 and ImA20_46260, respectively, while the purple box shows the introgression of the human IbA12G3 sequence into monkey IbA12G3. Source data are provided as a Source Data file.

The generated phylogenetic network and the mean *dxy* showed that ImA18_37833 was under the influence of InA14_45028 and IbA12G3_46287 (Fig. 1d and Supplementary Fig. 3). Phylogenetic topology weighting of these isolates and the ImA18 population showed no dominance of any specific topology across the eight chromosomes, indicating the occurrence of multiple sequence introgression events in ImA18_37833. The results of the *dxy* analysis suggested that IbA12G3_46287 contributed a large region of chromosome 1, ImA18 contributed regions of chromosomes 1, 6, and 7, while InA14_45028 contributed regions of chromosomes 4 and 6 (Supplementary Fig. 6a–c). Similarly, the IiA17 isolates analyzed in this study showed high identity to each other, except for IiA17_45372, which appeared to have gene flow from InA14_45028. A large region of the InA14_45028 with an *fd* value of 1 and low *dxy* values was present on chromosome 7, confirming the occurrence of InA14_45028 sequence introgression into IiA17_45372 (Supplementary Fig. 6d–g).

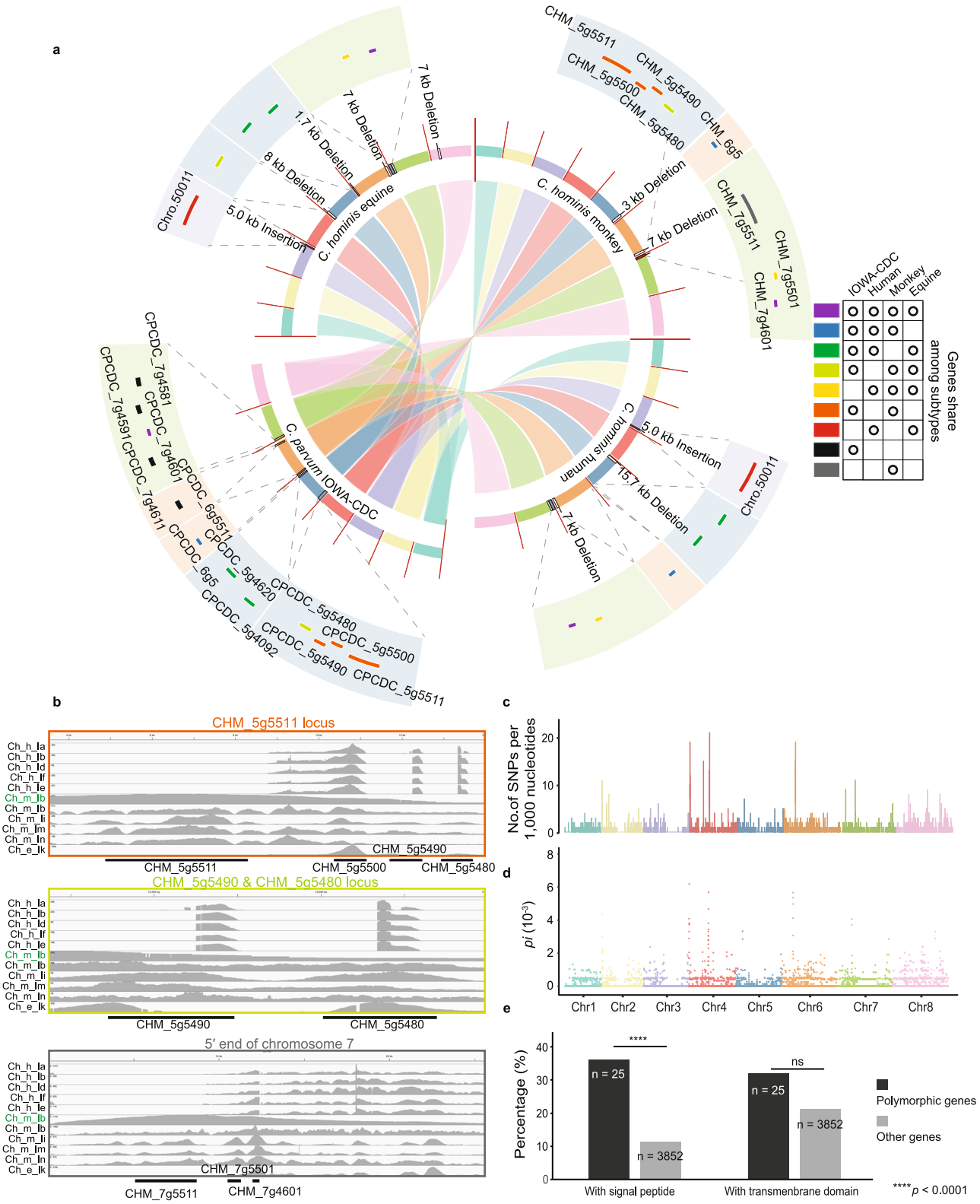
Differences in gene content among genomes of NHP, equine and human *C. hominis* subtypes

We obtained short-read assemblies of other animal *C. hominis* isolates, with genome sizes ranging from 8.97 to 9.12 Mb in 62 to 825 contigs. The majority of them were comparable in completeness to the new *C. hominis* IbA12G3_46287 genome (Supplementary Fig. 1). The data of whole genome alignment showed that the genomes from human, equine, and monkey *C. hominis* isolates were largely collinear with only a few minor structural differences (Fig. 4a and Supplementary Fig. 7).

Among the *C. hominis* genomes, all monkey isolates had an approximately 16 kb insertion in the 5′ region of chromosome 5. This region contained the orthologs of the CPCDC_5g5511, CPCDC_5g5500, CPCDC_5g5490, and CPCDC_5g5480 genes, which encode an INS-like protease, an NFDQ secretory protein, and two MEDLE secretory proteins (Supplementary Fig. 7a). The CPCDC_5g5480 gene was also present in the equine isolates. In read-mapping, reads from monkey or equine isolates were mapped to these loci, but no reads were mapped from human isolates (Fig. 4b). These data and the PCR results supported gene insertions and deletions (Supplementary Fig. 8).

Although we cannot determine the exact copy numbers of the subtelomeric *INS* and *MEDLE* genes in the genomes of *C. hominis* isolates from NHP due to the collapse of highly identical sequences at the 5′ end of chromosome 1 and the 3′ ends of chromosomes 7 and 8, the absence of such genes in the *C. hominis* genomes of human isolates was determined not only by aligning all assemblies to the new reference genome but also by blast analyses, read mapping, and PCR. For example, no reads from human *C. hominis* isolates were mapped to the orthologs of genes CPCDC_5g5511 (an *INS* gene), CPCDC_5g5490 (a *MEDLE* gene), and CPCDC_5g5480 (another *MEDLE* gene) at the 5′ end of chromosome 5 of the *C. hominis* IbA12G3_46287 (Fig. 4b). The results of the PCR analysis supported the presence of several *C.*

https://doi.org/10.1038/s41467-024-54995-4



parvum-associated genes in the NHP isolates but not in the human isolates (Supplementary Fig. 8). Blastn analysis of genome assemblies from *C. hominis* human isolates did not find any sequences of the *INS* gene, while only one sequence of the *MEDLE* gene was found and identified as an ortholog of CPCDC_7g4601. In tblastn analysis of databases constructed with short reads from Illumina sequencing, the protein sequence of the *INS* gene (CHM_5g5511) pulled out some reads from *C. hominis* monkey and *C. parvum* isolates with average identities of 100% and > 90%, respectively, but with the average identity of <50% from *C. hominis* human isolates. The latter were assembled to an ortholog of CPCDC_3g4270 (an *INS*-like gene with no CNV between isolates). The protein sequences of the *MEDLE* genes (CHM_5g5490 and CHM_5g5480) pulled out reads in *C. hominis* NHP and human isolates with maximum identities of 100% and 80%. These reads from the NHP and human isolates were assembled to three sequences (representing the orthologs of CPCDC_5g5490, CPCDC_5g5480, and

Fig. 4 | Genomic differences among *Cryptosporidium hominis* genomes. **a** Syntenic relationship in gene organization among genomes of human, monkey, and equine *C. hominis* isolates using the *C. parvum* IOWA-CDC and the *C. hominis* IbA12G3 genomes as the references. The genomes are largely colinear. Copy number variations (CNVs) of genes between *C. hominis* genomes were detected by genome alignment and verified by read-mapping, blast analyses, and nested PCR. The CNVs are shown in blocks of different colors (see figure key). Among the genomes analyzed, the orthologs of CPCDC_5g5511 and CPCDC_5g5490 (the orange block) are absent in human and equine *C. hominis* isolates, while the ortholog of CPCDC_5g5480 (the light green block) is absent in human *C. hominis*; orthologs of CPCDC_5g4092 and CPCDC_5g4620 (the green block) are absent in monkey *C. hominis*; the ortholog of CPCDC_6g5 (the blue block) is absent in equine *C. hominis*; the ortholog of Chro.50011 (the red block) is absent in *C. parvum* and monkey *C. hominis* isolates. Monkey *C. hominis* genomes have an additional copy of the ortholog of CPCDC_5g5511 (the gray block) on chromosome 7, compared with other genomes analyzed. All *C. hominis* genomes do not have orthologs of CPCDC_6g5511, CPCDC_7g4611, CPCDC_7g4591, and CPCDC_7g4581 (the black block). **b** Confirmation of the subteleomeric genes in animal-adapted *C. hominis* subtypes using read-mapping. All reads were mapped to the chromosome-level reference genome of *C. hominis*. Data from PacBio are highlighted in green. **c** Distribution of core SNPs in monkey *C. hominis* genomes by chromosome (with a 1000 bp sliding window and 500 bp steps). **d** Nucleotide diversity (*Pi*) among monkey *C. hominis* genomes using a 1000 bp sliding window. **e** Differences in the distribution of genes encoding proteins with signal peptides and transmembrane domains between highly polymorphic genes and other genes. Highly polymorphic genes are more likely to encode proteins with signal peptides ($p = 7.5 \times 10^{-15}$). Statistical analysis was performed using the one-tailed chi-squared test. Source data are provided as a Source Data file.

CPCDC_7g4601) and one sequence (representing the ortholog of CPCDC_7g4601), respectively. These results support the conclusion that *C. hominis* human isolates lacked the one *INS* and the two *MELDEL* genes detected in *C. hominis* NHP isolates (Supplementary Table 5). Since increased read coverage and multiple read types were observed in a 10 kb region at the 5′ end of chromosome 7 in the monkey isolate IbA12G3_46287, which contained the other copy of the ortholog of CPCDC_5g5511, *C. hominis* NHP isolates may have several additional copies of the CPCDC_5g5511 gene (Fig. 4b, Supplementary Fig. 1a, and Supplementary Table 5).

It is also difficult to determine the copy number of the CPCDC_5g5500 gene in *C. hominis* isolates. The gene encodes a secretory *NFDQ* protein and is located at the 5′ end of chromosome 5 in *C. parvum*. A paralog of it is located at the 3′ end of chromosome 6 in *C. parvum*. Both copies are next of the two copies of the subtelomeric *INS* gene (CPCDC_5g5511 and CPCDC_6g5511). In the *C. hominis* IbA12G3_46287 genome, these two copies of the *NFDQ* gene are located at the 5′ ends of chromosome 5 and chromosome 7. The increase in read depth in a telematic region at the 5′ end of chromosome 7 containing the CHM_7g5501 (ortholog of CPCDC_5g5500 in *C. parvum*) and the appearance of several read types suggest that the monkey isolate may have additional copies of this gene (Fig. 4b and Supplementary Fig. 1a). This is likely part of the sequence compression at the chromosome ends. In addition, this increase in read coverage of the *NFDQ* gene occurred in other *C. hominis* isolates (Fig. 4b).

All monkey isolates had a 3 kb deletion in the 3′ region of chromosome 5. This resulted in the loss of two orthologs of CPCDC_5g4620 and CPCDC_5g4092 (Supplementary Fig. 7c). The equine isolates had a 1.7 kb deletion in the 5′ region of chromosome 6, resulting in the loss of the ortholog of CPCDC_6g5, while they had a 5 kb fragment in the 3′ region of chromosome 3 containing a *C. hominis* specific gene, Chro.50011, which was absent in the monkey isolates (Supplementary Fig. 7d). Read mapping and blast results supported the validity of these gene insertions and deletions.

Sequence differences among monkey *C. hominis* isolates

There were 2314 SNPs in the core genome (without telemetric regions) among the monkey *C. hominis* isolates. Altogether, 25 genes displayed substitution rates above the threshold (average + 3 SD) and contained more than two SNPs (Fig. 4c and Supplementary Table 6). These findings were supported by the *Pi* values among these genomes (Fig. 4d). Most of the polymorphic genes were located in the subtelomeric regions of the chromosomes and encoded mostly proteins with the signal peptides (Fig. 4e). They included four genes encode mucin proteins (the orthologs of CPCDC_2g420, CPCDC_2g450, CPCDC_4g1300, and CPCDC_6g1080) (Supplementary Table 6).

Mouse infectivity of monkey *C. hominis* isolates

The presence of several subtelomeric *INS* and *MEDLE* genes that are associated with broad host range of *C. parvum* prompted us to evaluate the infectivity of monkey *C. hominis* isolates to mice. GKO mice were inoculated with 10^4 oocysts of ImA20, InA17, and the monkey IbA12G3. Results of qPCR analysis showed that most mice were positive 4 days post infection (DPI) (Fig. 5a). DNA sequence analysis of the *gp60* gene confirmed that the subtypes of oocysts produced in the mice at the peak of infection (DPI 9) and at other times matched those of the inoculated oocysts (Supplementary Fig. 9a, b), confirming that the mice were infected with these *C. hominis* subtypes.

Oocyst shedding in mice infected with the IOWA-Waterborne isolate of *C. parvum* peaked at DPI 9 (oocysts per gram of feces, OPG = 4 logs) and began to decline, eventually becoming negative after DPI 60 (Fig. 5a). In contrast, the mice infected with monkey *C. hominis* isolates showed a rapid increase in oocyst shedding intensity, with OPG reaching 6 logs on DPI 9. Thereafter, OPG levels decreased slowly but remained positive for over 80 days. Peak OPGs and mean OPGs over 80 days were significantly higher in mice infected monkey isolates than in mice infected with the IOWA-Waterborne isolate.

Histological examination at DPI 12 revealed that the parasite load in the ileum was significantly higher in mice infected with monkey *C. hominis* isolates than in those infected with the IOWA-Waterborne isolate (Supplementary Fig. 9c). In addition, the intestinal villus length/crypt depth (v/c) ratio was significantly lower in mice infected with monkey isolates, but not in those infected with IOWA-Waterborne (Fig. 5c, d). Scanning electron microscopy (SEM) analysis of the ileal tissue confirmed the presence of abundant parasites on the intestinal epithelium of mice infected with monkey isolates, particularly IbA12G3 (Fig. 5e). No parasites were observed in the intestinal tissues of uninfected mice.

Differences in virulence and infectivity between monkey *C. hominis* isolates

OPG levels of mice infected with InA17 and ImA20 subtypes reached almost 7 logs but decreased to 4 logs within 1 month. In contrast, OPG levels of mice infected with IbA12G3 reached over 7 logs and remained at high levels (6 logs) for two months (Fig. 5a). In addition, the mice infected with IbA12G3 and InA17 had slightly reduced body weight gain (Fig. 5b). The higher virulence of IbA12G3 was also supported by the measurement of parasite burden and v/c ratio in the ileal tissue (Fig. 5c,d and Supplementary Fig. 9c).

In experimental infections of GKO mice with varying numbers of oocysts, a single IbA12G3 oocyst resulted in infection in three of eight mice, with a calculated ID50 of seven oocysts. On the other hand, five InA17 oocysts were able to induce infection in five of eight mice, with ID50 values of eight oocysts. In contrast, even the use of 1000 ImA20 oocysts induced infections in only two of four mice, with a calculated ID50 of >1000 oocysts (Supplementary Table 7). The prepatent period and the time to peak oocyst shedding were longer in mice infected with lower oocyst doses (Fig. 5f–h).

https://doi.org/10.1038/s41467-024-54995-4

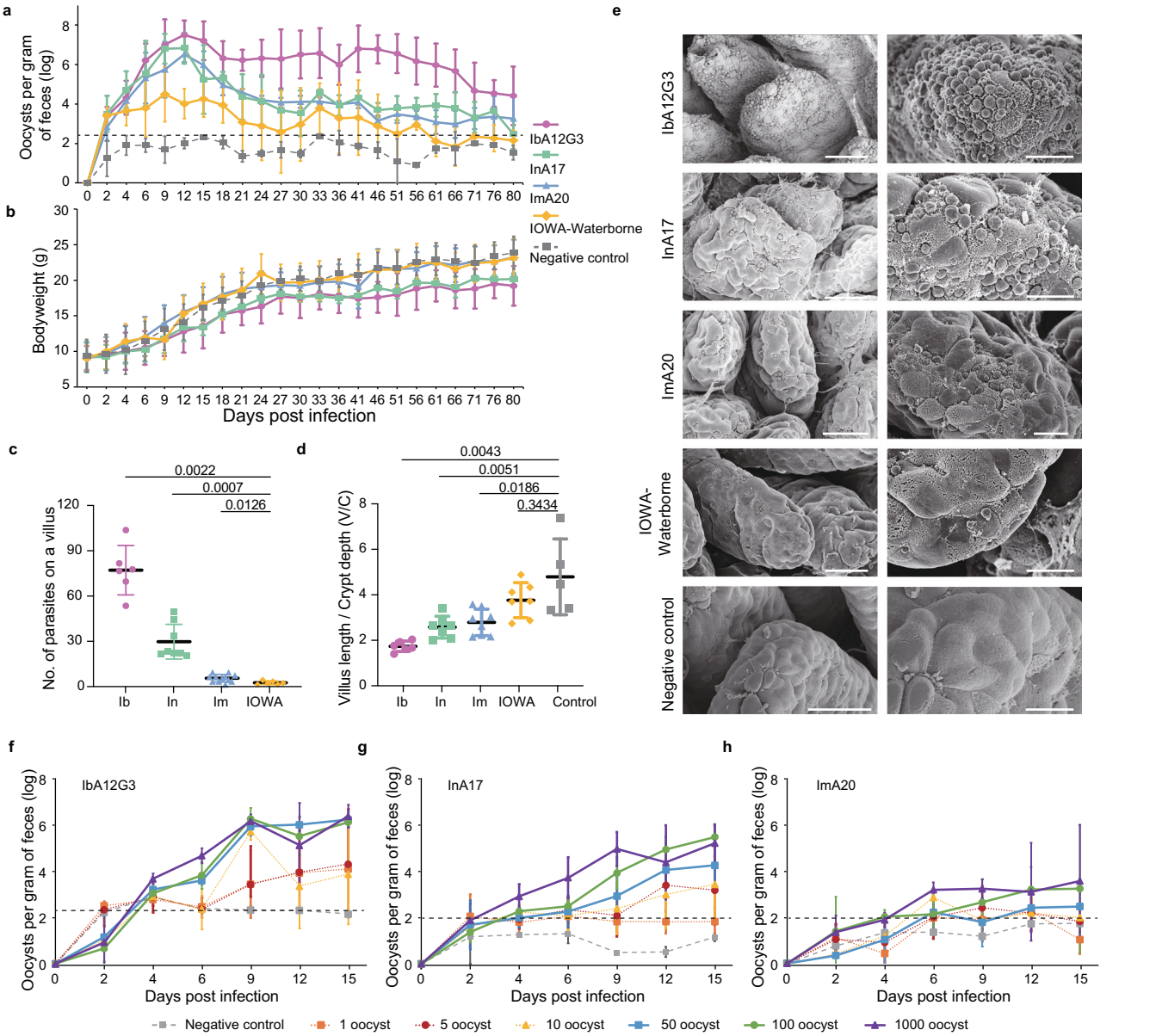


Fig. 5 | Differences in infection patterns and virulence among three *Cryptosporidium hominis* isolates from monkeys. a Oocyst shedding patterns in interferon-γ knockout mice ($n = 6$ per group, *C. hominis* IbA12G3, InA17, and ImA20, *C. parvum* IIaA17G2R1, and uninfected). The detection threshold is indicated by the dotted line. The peak values of oocysts per gram of feces (OPG) of IbA12G3, ImA20, and InA17 are significantly higher than those of IIaA17G2R1 ($p = 0.0022$, 0.0022, and 0.0095), and the mean OPGs of IbA12G3, ImA20, and InA17 over 80 days are significantly higher than those of IIaA17G2R1 ($p < 0.0001$, $p < 0.0001$, and $p < 0.0001$). The mean OPG of IbA12G3 over 80 days is significantly higher than that of ImA20 and InA17 ($p < 0.0001$ and $p < 0.0001$). Statistical analysis was performed using the two-tailed Wilcoxon signed-rank test. **b** Differences in body weight among the five groups of mice. **c** Mean number of parasites per villus in the small intestine of infected mice ($n = 3$ per group). The number of parasites after *C. hominis* infection was significantly higher than that of *C. parvum* IIaA17G2R1 subtype ($p = 0.0022$, 0.0007, and 0.0126 for IbA12G3, InA17, and ImA20). **d** Differences in the villus/crypt height ratio (v/c) of the ileum of infected mice. The v/c values of mice infected with monkey *C. hominis* were significantly lower than those of healthy mice ($p = 0.0043$, 0.0051, and 0.0186 for IbA12G3, InA17, and ImA20), while there was no significant difference between *C. parvum* IIaA17G2R1-infected and uninfected control mice ($p = 0.3434$). **c**, **d** Statistical analyses were performed using the two-tailed Mann–Whitney *U* test. **e** Scanning electron microscopy images of the ileum and colon of GKO mice of the five groups ($n = 3$). Scale bars = 30 μm (left) and = 10 μm (right). **f**–**h** Oocyst shedding patterns of GKO mice infected with different doses (1000 ($n = 4$), 100 ($n = 4$), 50 ($n = 6$), 10 ($n = 6$), 5 ($n = 8$), 1 ($n = 8$), and 0 ($n = 4$)) of IbA12G3, InA17, and ImA20 isolates. **a,b,c,d,f,g,h** Data are presented as mean values +/− SD. Source data are provided as a Source Data file.

Contribution of subtelomeric *MEDLE* and *INS* genes to *C. hominis* infectivity in mice

To assess the biological significance of the *MEDLE* and *INS* genes, we first attempted to replace the additional CHM_5g5480, CHM_5g5490 and CHM_5g5511 genes in the IbA12G3 genome (the orthologs of the CPCDC_5g5480, CPCDC_5g5490, and CPCDC_5g5511 genes in *C. parvum*) with the tdTomato sequence (Fig. 6a, b). Among them, CHM_5g5480 and CHM_5g5490 were dispensable because we were able to generate a mutant expressing tdTomato but lacking the CHM_5g5480 and CHM_5g5490 (Fig. 6c). The complete deletion of the genes was confirmed by the read mapping results, which showed that the locus could not be fully covered by the reads from the IbA12G3-ΔCHM_5g5480/5490 line (Fig. 6d). The PCR analysis also confirmed the complete deletion of the genes (Supplementary Fig. 10a).

After several failed attempts to knock out the full length of the CHM_5g5511 (*INS*) gene, an alternative approach was taken to remove

https://doi.org/10.1038/s41467-024-54995-4

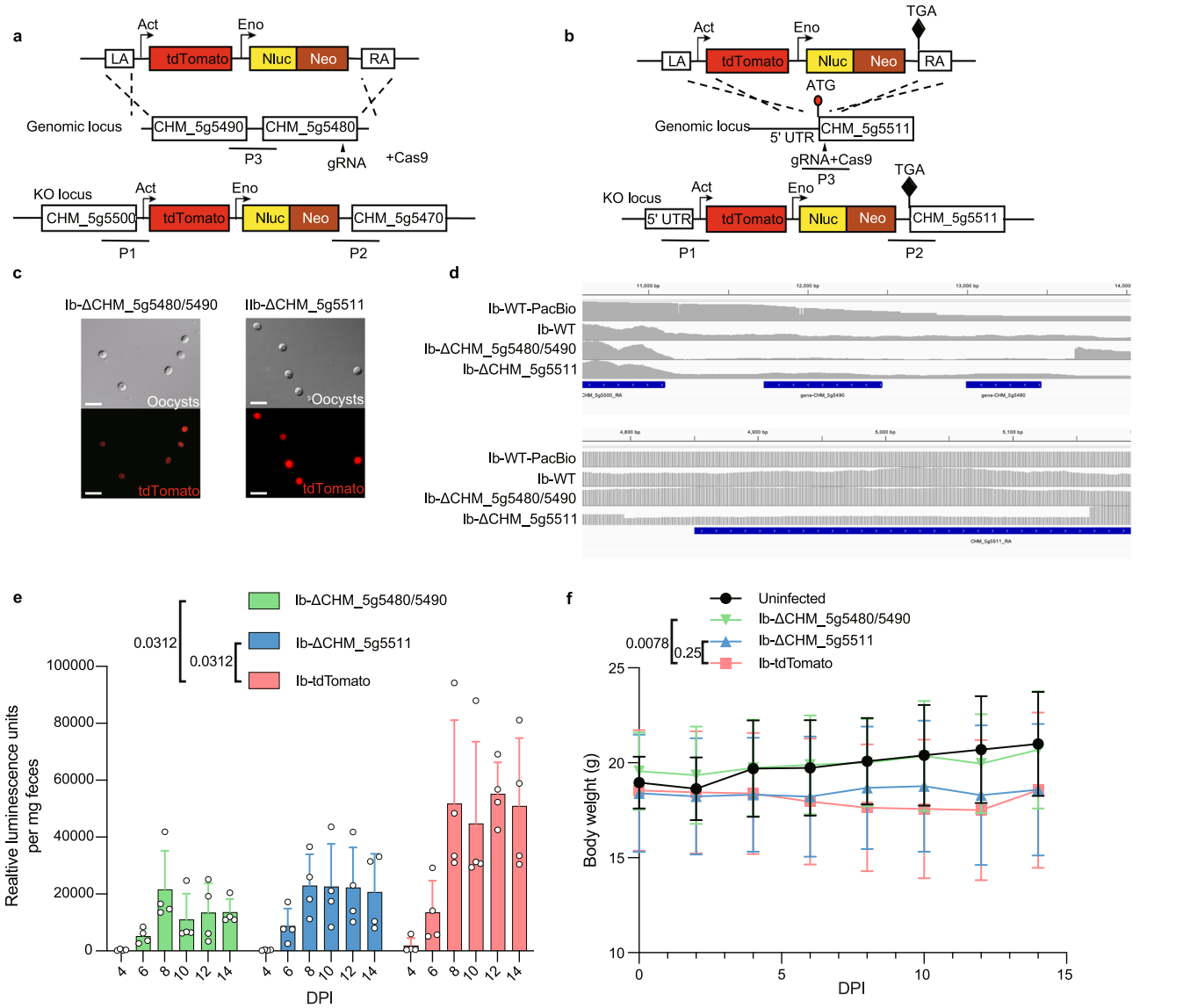


Fig. 6 | Subtelomeric *MEDLE* and *INS* genes play important roles in *C. hominis* infectivity in mice. a, **b** Strategy used to replace the target region with sequences encoding tdTomato, NanoLuciferase (Nluc), and neomycin-resistance gene (Neo). **a** The region containing CHM_5g5490 and CHM_5g5480 was deleted. **b** The 310 bp from the 5′ end of the Ib-CHM_5g5511 gene was deleted, and a stop codon was inserted downstream of the deletion site. **c** Microscopy of oocysts from mice infected with the transgenic parasites, with all oocysts being red. Scale bars in the figures are 10 μm. **d** Confirmation of the knockout genes in *C. hominis* IbA12G3 strain using read mapping. All reads were mapped to the chromosome-level reference genome of *C. hominis*. **e** Oocyst shedding patterns in uninfected interferon-γ knockout (GKO) mice and mice experimentally infected with *C. hominis* IbA12G3 CHM_5g5490/CHM_5g5480 deletion (Ib_ΔCHM_5g5480/5490), partial deletion of CHM_5g5511 (Ib_ΔCHM_5g5511), and tdTomato tagged strains (Ib-tdTomato). GKO mice ($n = 4$ per group) were infected with 10^4 oocysts. The mean OPGs of Ib_ΔCHM_5g5480/5490 and Ib_ΔCHM_5g5511 over 14 days are significantly higher than those of Ib-tdTomato ($p = 0.0312$ and 0.0312, respectively). **f** Differences in body weight among the four groups of mice ($n = 4$ per group). Infection with the IbA12G3-ΔCHM_5g5480/5490 improved the weight gain of mice compared to the Ib-tdTomato group ($p = 0.0078$), while no difference in weight gain was observed between the mice infected with IbA12G3-ΔCHM_5g5511 and Ib-tdTomato ($p = 0.25$). **e**, **f** Error bars represent SD values. Statistical analyses were performed using the two-tailed Wilcoxon signed-rank test. Source data are provided as a Source Data file.

310 bp from the 5′ end of the Ib-*INS* gene and insert a stop codon downstream of the deletion site to disrupt its expression (Fig. 6b). The sgRNA and the homology repair fragments were designed based on the identical regions of the different copies of CHM_5g5511. The IFA results showed that all the mutants expressed tdTomato, and the PCR results showed that the mutants carried the selective tags (Fig. 6c and Supplementary Fig. 10b). These data indicated that all the parasites were transgenic. However, in the WGS analysis of the mutant line, the 5′ end of the *INS* gene had mapped reads with approximately 50% reduced coverage, suggesting that only parts of the multicopy *INS* gene were knocked out. This finding was supported by the PCR results, as the mutants had both the insertion and the KO region (Supplementary Fig. 10b). We also tried three times to knock out the whole genome region on chromosome 5 with two *MEDLE* and one *INS* genes, but were unable to obtain transgenic parasites after paromomycin selection.

In GKO mice, deletion of CHM_5g5480/CHM_5g5490 or partial deletion of CHM_5g5511 resulted in significantly lower infection intensity than the tagged wild-type (Ib-tdTomato), as indicated by the measurement of fecal luciferase activity (Fig. 6e). Infection with the IbA12G3-ΔCHM_5g5480/5490 resulted in significantly higher weight gain of mice than infection with Ib-tdTomato, with the former

https://doi.org/10.1038/s41467-024-54995-4

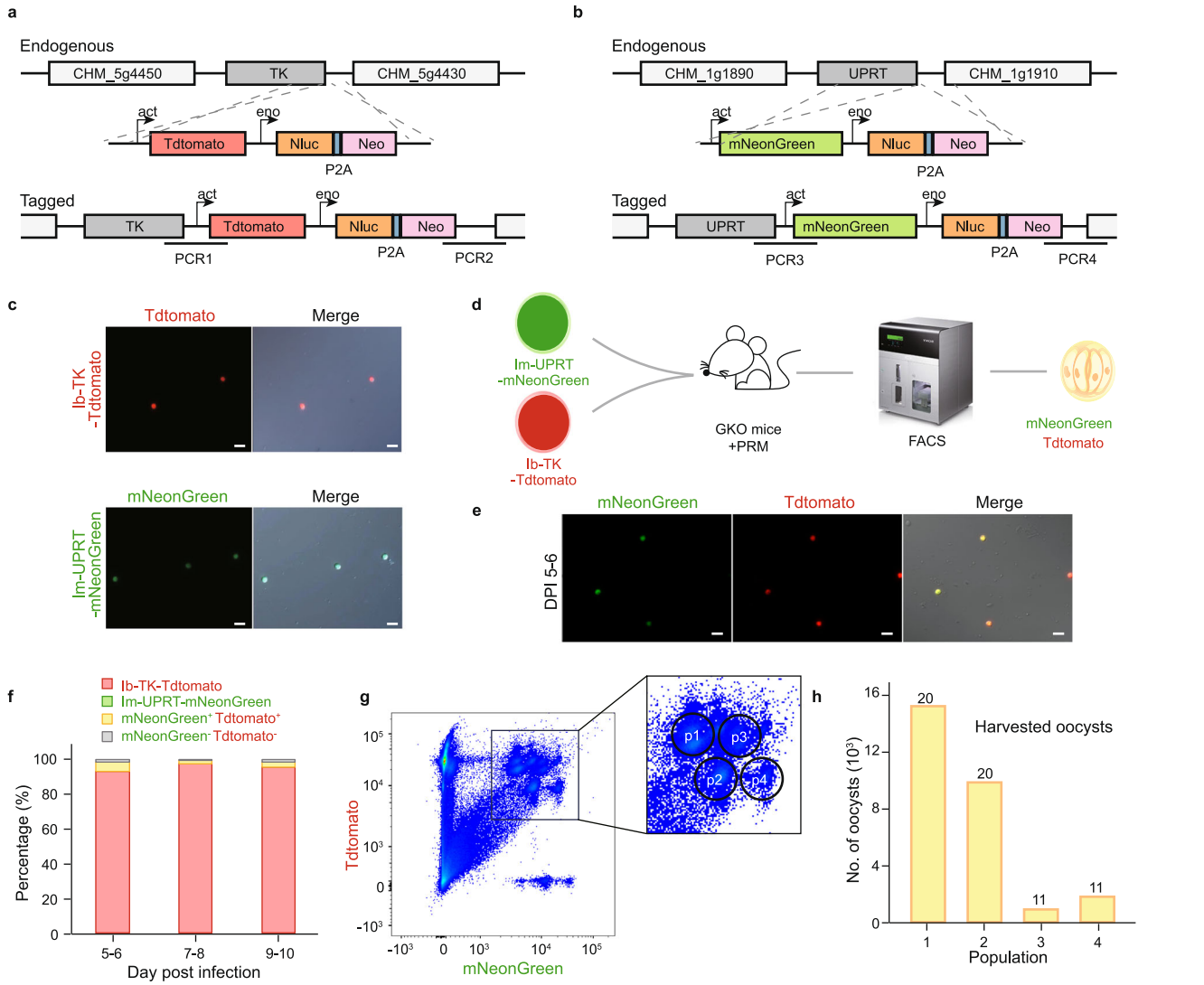


Fig. 7 | Genetic tagging and crosses of two monkey isolates of *Cryptosporidium hominis* with different virulence. a Illustration of the strategy used in endogenous tagging of the IbA12G3 isolate downstream of the TK (CHM_5g4440, ortholog of CPCDC_5g4440 in *C. parvum*) locus with a cassette of tdTomato, NanoLuciferase (Nluc), and neomycin-resistance gene (Neo). **b** Illustration of the strategy used in endogenous tagging of the ImA20 isolate downstream of the s UPRT (CHM_1g1900, ortholog of CPCDC_1g1900 in *C. parvum*) locus with a mNeonGreen-Nluc-P2A-Neo cassette. **c** Image of oocysts purified from fecal samples of mice infected with endogenously tagged *C. hominis* IbA12G3. and ImA20 isolates. **d** Diagram of the genetic cross of endogenously tagged IbA12G3. and ImA20 isolates. 10^4 oocysts of each Ib-TK-Tdtomato oocysts (red) and Im-UPRT-mNeonGreen oocysts (green) were used to infect the same GKO mouse ($n = 3$). Oocysts expressing both fluorescent reporters (yellow) were identified by fluorescence microscopy and collected by flow cytometric analysis of purified oocysts. **e** Fluorescence microscopy images of purified oocysts at DPI 5. Recombinant progeny expressed both with Tdtomato and mNeonGreen oocysts and were yellow in the merged images. **c**, **e** Scale bars in the figures are 10 μm. **f** Distribution of oocysts of different colors in the progeny of the genetic cross at different time points after infection. **g** Results of flow cytometry of oocysts purified at DPI 5 showing the oocyst population with double positive fluorescence. **h** Number of yellow oocysts harvested from the four regions in I by flow cytometric sorting. Numbers on the bars are numbers of oocysts confirmed as yellow by fluorescence microscopy. All harvested oocysts examined by fluorescence microscopy were yellow. Source data are provided as a Source Data file.

behaving like the uninfected group (Fig. 6f). However, no difference in weight gain was observed between mice infected with IbA12G3-ΔCHM_5g5511 and Ib-tdTomato.

Tractability and utility of monkey *C. hominis* isolates

To facilitate biological studies of *C. hominis* infection, we generated transgenic lines by introducing the Tdtomato sequence into the IbA12G3 isolate and the mNeonGreen sequence into the ImA20 isolate using CRISPR/Cas9 (Fig. 7a, b). The transgenic lines also carried the nanoluciferase sequence for easy measurement of parasite load. Immunofluorescence staining showed that oocysts harvested from the fecal samples of GKO mice inoculated with the transgenic lines expressed Tdtomato or mNeonGreen, supporting the correct tagging of the two lines (Fig. 7c). PCR and sequence analyses confirmed that both transgenic lines had the correct insertion into the genomic loci (Supplementary Fig. 11a).

We co-infected GKO mice with equal numbers of IbA12G3-TK-Tdtomato and ImA20-UPRT-mNeonGreen oocysts and selected the transgenic lines by paromomycin treatment (Fig. 7d). Some oocysts of the genetic cross harvested at DPI 5 had both red and green colors (yellow in merged images) (Fig. 7e and Supplementary Fig. 11b). Red oocysts had the absolute advantage (>90%), consistent with the higher infectivity of IbA12G3 than ImA20 (Fig. 7f). Adjusting the infection dose (increasing the infection dose of ImA20 tenfold) increased the yield of yellow progeny (Supplementary Fig. 11c). Flow cytometric analysis of the progeny showed that the yellow oocysts clustered into four

populations at the late time point of infection (DPI 5 and DPI 6) (Fig. 7g and Supplementary Fig. 11d, e). Fluorescence microscopy revealed that all oocysts from these populations were yellow (Fig. 7h and Supplementary Fig. 11f). They provide experimental support for the conclusion that co-infection of two *C. hominis* subtypes can produce genetic recombinants, which was observed in the evolutionary analysis of the monkey isolates.

Discussion

The results of this study have confirmed the biological classification and the genetic uniqueness of *C. hominis* naturally circulating in several animal species at the whole genome level. While several unique *gp60* subtype families have been reported in equine animals (Ik) and Old-World monkeys (Ii, Im, and In) in recent years, very few sequence data are available from these isolates. Sequence analysis of the conserved 18S rRNA gene suggests that these subtypes may be host-adapted *C. hominis* variants with minor sequence differences from the common *C. hominis* subtype families in humans[12]. This was supported by partial WGS data from an IkA18G1 isolate from a human patient in Sweden[22]. In the present study based on WGS data, *C. hominis* from humans, equine animals, and macaques formed a distinct clade separate from *C. hominis* related species. However, their genomes are divergent from each other and there is limited gene flow between the three groups. These results support the animal isolates as *C. hominis* and the existence of host adaptation within this species[4,12].

Comparative genomic data from the study suggest that copy number variation in subtelomeric genes may contribute to the difference in host range between *C. hominis* animal and human isolates. *C. hominis* isolates from humans, equine animals, and macaques differ in the copy numbers of the *MEDLE* and *INS* genes. Proteins encoded by these two gene families have been associated with host specificity in *Cryptosporidium* spp[23,24]. Therefore, these differences may lead to differences in host infectivity among the human, monkey, and equine subtypes of *C. hominis*. Surprisingly, the monkey *C. hominis* subtype ImA13 recently caused an outbreak of diarrhea in humans, suggesting an underestimated host range of the NHP-adapted *C. hominis*[14].

Multiple recombination underlies the genetic diversity and population genetic structures of *C. hominis* NHP-adapted subtypes. We found strong evidence for sequence introgression among IbA12G3, IiA17, ImA18, ImA20, InA14, and InA17. ImA18 has been reported in China as a the dominant subtype for *C. hominis* infection in crab-eating macaques on a breeding farm in Hainan[5]. The IiA17 subtype was first reported in a rhesus monkey in the United States and has been detected several times in macaque monkeys and humans in China and Thailand since 2014[5,6,8,15,25,26]. This subtype is likely to be a native *C. hominis* subtype in NHPs in China and Southeast Asia. In addition, the InA14 subtype is also common in macaque monkeys on the same farm[5]. We have further provided experimental support for the occurrence of genetic recombination during co-infection of two *C. hominis* subtypes in mice, demonstrating a high frequency of genetic recombination in *C. hominis* and the formation of progenies with recombined genomes. Therefore, the co-occurrence of these subtypes along with several other subtypes on the farm allows for genetic recombination and underlies the genetic diversity. Genetic recombination also plays a key role in shaping the population structure of *C. hominis* in humans and *C. parvum* in both humans and animals[21,27–29].

Genetic recombination between the NHP-adapted subtypes and the human IbA12G3 subtype is likely responsible for the emergence of the divergent subtype in monkeys. The monkey IbA12G3 isolate has the common Ib-type *gp60* sequence seen in human IbA12G3 isolates, but has sequences similar to the ImA20 and InA17 subtypes in NHP in most genomic regions. The IbA12G3 subtype is a common human-pathogenic subtype that has been reported in many countries since 2009 and has been identified as the cause of cryptosporidiosis outbreaks in the United Kingdom and Australia[30–37]. It has been found in crab-eating monkeys in China since 2014[8]. Because the IbA12G3 subtype has a global distribution in humans but is only occasionally detected in NHPs in Bangladesh and China[5,38], we considered the gene flow from humans to NHPs more likely for this unique IbA12G3 genome that we sequenced. However, we have no other more robust evidence for the direction of *C. hominis* flow between humans and NHPs.

Taking advantage of the presence of several *C. parvum*-specific genes that have been associated with a broad host range, we have established a mouse infection model of *C. hominis* using these monkey isolates. Previous studies have shown that common subtypes of human *C. hominis* are not infectious to rodents[3]. Because the NHP-adapted *C. hominis* subtypes have several subtelomeric *INS* and *MEDLE* genes that have been associated with animal infectivity in *C. parvum*, we evaluated their infectivity to GKO mice. The three *C. hominis* isolates from NHPs not only infected these mice but also induced significantly higher and more prolonged oocyst shedding than the IIa-Waterborne reference of *C. parvum*. Thus, the host range of monkey *C. hominis* isolates is indeed broader than that of human isolates.

In the present study, we provide experimental evidence for the potential involvement of the subtelomeric *INS* and *MEDLE* genes in *C. hominis* infectivity and pathogenicity. These genes have been implicated in the invasion and development of *Cryptosporidium* spp[39–41]. Deletion of the subtelomeric MEDLE genes (orthologs of the CPCDC_5g5490 and CPCDC_5g5480 genes) significantly reduced *C. hominis* growth in GKO mice and prevented body weight loss in infected animals. The data suggest that *MEDLE* genes are involved in host infectivity, but other genes may also contribute to *C. hominis* infectivity in mice. In contrast, the *INS* gene was partially knocked out. This may be due to the presence of multiple copies of the subtelomeric *INS* gene with high sequence identity in this isolate, and the current CRISPR procedure cannot simultaneously manipulate multiple identical loci on different chromosomes. Interestingly, the partial deletion of the *INS* genes also significantly reduced *C. hominis* growth but not as much as the *MEDLE* genes. In addition, we attempted to knock out the entire *INS* and *MEDLE* region with double sgRNAs covering the entire region. The failure to obtain transgenic parasites after paromomycin selection may support the importance of these genes, but we could not exclude that the region was too large to be knocked out. Studies of multiple genes in *Cryptosporidium* remain challenging, although the combined function of *INS* and *MEDLE* could be further explored in future studies using the recently developed second selection markers[42,43]. Overall, the data support the previous suggestion that copy number variation in subtelomeric genes encoding the INS and MEDLE secretory proteins may contribute to the differences in host infectivity between different *Cryptosporidium* species and subtypes[24,27,44].

The three monkey isolates examined in this study, each with different *gp60* subtypes, exhibit different infectivity and virulence. In particular, IbA12G3 induced a significantly higher intensity of oocyst shedding in mice than ImA20 and InA17. In addition, both IbA12G3 and InA17 infections resulted in significantly higher parasite loads in the mouse intestine than ImA20, with the former two also causing reductions in body weight gain, consistent with their much lower ID_{50} values. Previously, differences in infectivity and virulence have been observed between different subtype families of *C. parvum* and *C. tyzzeri*[45–47], and between different isolates of the same *gp60* subtype[48–50]. Results from comparative genomic analysis indicate that sequence differences in genes encoding mucin glycoproteins and other secretory proteins are associated with different phenotypic traits of isolates in mice[45,46]. This is consistent with the data obtained in the present study. Among the highly polymorphic genes between monkey *C. hominis* isolates, there are four genes encoding mucin proteins including the GP60. Mucin glycoproteins play a critical role in sporozoite invasion[51], especially GP60, which also has multiple functions

during *C. parvum* infection and is an important factor in host infectivity[52].

The newly developed mouse model of *C. hominis* infection provides an accessible mechanism for advanced biological characterization of this important human pathogen. In the present study, we were able to endogenously tag two isolates with different infectivity and virulence to mice with fluorescent reporters and a nanoluciferase, allowing for easy detection and tracking of them. More importantly, we were able to genetically cross the two isolates and enrich the progeny of the genetic cross using flow cytometry. Previously, genetic crosses between *C. parvum* isolates and between *C. parvum* and *C. tyzzeri* have been made in vivo and in vitro[53–56]. However, the genetic crosses reported here are intended to demonstrate that these *C. hominis* isolates are genetically tractable, and together with the newly developed *C. hominis* mouse model they provide an accessible mechanism for advanced biological characterization of this important human pathogen. We anticipate using this approach for future discovery of genetic determinants of host infectivity and virulence and comparative studies of *C. parvum* and *C. hominis*.

Prior to this study, our understanding of the emergence of the novel *C. hominis* subtypes was limited by the lack of WGS data. In this study, we acquired WGS data from four animal-adapted *C. hominis* subtype families and demonstrated their genomic uniqueness. Previous evolutionary genetic and multilocus sequence analyses had indicated the presence of geographic segregation within *C. hominis*[21,57], but we did not detect obvious sequence introgression from Asian human isolates into the Chinese monkey isolates. Therefore, a more systematic analysis of human isolates from China, where the novel subtypes are emerging[12], and Kenya, where one of the novel subtypes caused a recent human cryptosporidiosis outbreak[14], is needed to improve the understanding of the transmission of the novel *C. hominis* subtypes. While the newly generated reference *C. hominis* genome provides insights into the copy number variation of subtelomeric genes, three chromosome ends with high sequence identity in *C. parvum*[20] are missing from the assembly due to PacBio's read length limitation. The increased read depth in another chromosome end during read mapping suggests the presence of collapsed sequences[58]. Therefore, chromosome end replication in *C. hominis* may have occurred in the same chromosomes as in *C. parvum*. This replication may have also occurred in other species that are closely related to *C. parvum*. More long read data from other *C. hominis* isolates are needed to confirm this observation[20].

In conclusion, the genomes of the NHP-adapted and equine-adapted subtypes differ significantly from those of the human-pathogenic subtype families. These differences are mainly in the CNVs of the subtelomeric *MEDLE* and *INS* genes. These differences contribute to their infectivity in GKO mice. The results of this study also indicate that the co-occurrence of different subtypes has led to frequent recombination between subtypes, shaping the population structure of *C. hominis* in NHPs and the emergence of new subtypes. These observations improve our understanding of the evolutionary genetics of *C. hominis* and provide rich targets for advanced studies of genetic determinants of host infectivity and virulence of isolates. In particular, the *C. hominis* mouse model and genetically tractable isolates of varying infectivity and virulence developed in this study will be very useful for comparative studies of *C. hominis* and *C. parvum* and advanced characterizations of genetic determinants of phenotypic traits of these important pathogens.

Methods

Cryptosporidium isolates and mice

A total of 17 *C. hominis*-positive samples, collocated from crab-eating macaques (*Macaca fascicularis*) and donkeys (*Equus asinus*) in China between 2017 and 2022, were used in the study (Supplementary Table 1). They were obtained from previous molecular epidemiological studies of *Cryptosporidium* spp[5,9]. Fecal samples containing oocysts were stored in 2.5% potassium dichromate solution at 4 °C prior to the oocyst isolation for WSS analysis and infection studies. The *C. parvum* "IOWA" isolate (IIaA17G2R1 subtype) was purchased from Waterborne, Inc. (New Orleans, LA, United States). The species and subtype identity of the isolates were determined by PCR and sequence analysis of the 18S rRNA and *gp60* genes[59,60]. Information on the sequenced isolates is provided in Supplementary Data 1.

GKO mice on a C57BL/6 background were obtained from the Institute of Laboratory Animals Science, Chinse Academy of Medical Sciences. 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 Institutional Animal Care and Use Committee, South China Agricultural University (No. 2022c011).

Whole-genome sequencing

Oocysts were purified from *C. hominis*-positive samples using discontinuous sucrose and cesium chloride gradients as described[61]. Seventeen DNA samples were extracted from the purified oocysts after five freeze-thaw cycles using the QIAamp DNA minikit (Qiagen, United States, Cat. 51304). Whole-genome amplification (WGA) was performed on the extracted DNA using a REPLI-g Midi kit (Qiagen, Cat. 150043) following the manufacturer-recommended procedures. The WGA products were sequenced on an Illumina HiSeq 2500 (Illumina, San Diego, CA, United States) using the 150-bp or 250-bp paired-end approach. DNA was also extracted from IbA12G3-monkey-46287 oocysts using the traditional phenol-chloroform method. After WGA, the product was sequenced using standard PacBio ccs procedures.

A total of 240 whole-genome sequence (WGS) datasets of *C. hominis*, *C. parvum*, *C. cuniculus*, *C. tyzzeri*, *C. ubiquitum*, and *C. meleagridis* were downloaded from the Sequence Read Archive database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/sra/) using prefetch v2.11.3 (https://github.com/ncbi/sra-tools). They were generated in published studies[21–23,27,62–69]. Some basic information about the WGS data is provided in Supplementary Data 1.

Genome assembly and gene annotation for the reference *C. hominis* genome

Sequence reads obtained from PacBio were filtered for chimeras using 3rd-ChimeraMiner[70]. Sequence reads obtained from Illumina were trimmed for adapter sequences and poor sequence quality (phred score <30 and trimmed length <60) using fastp v0.23.4. A hybrid de novo assembly was performed with the PacBio and Illumina reads of IbA12G3-monkey-46287 using wengan v0.2 with M model, a genome size of 9.5 Mb, haploid, and a Kmer of 63. The PacBio reads were also assembled de novo using the hifiasm v0.19.8 with a Kmer of 63. The final assembly was polished using racon v1.5.0 with default settings and manual inspection of the long PacBio reads mapped to the contigs. The assemblies were aligned and sorted against the reference genomes of *C. parvum* IOWA-CDC (of the dominant IIaA15G2R1 subtype)[18] and IOWA-BGF (of the IIaA17G2R1 subtype)[20] using the "move contig" tool of Mauve v2.3.1. To assess the quality of the new assembly, the PacBio long reads were mapped back to the new assembly using minimap2 v2.28 with the map-hifi mode (https://github.com/lh3/minimap2) and visualized with igv v2.17.1.

Gene annotation for the *C. hominis* IbA12G3 genome was performed using AUGUSTUS v3.5.0 (https://github.com/Gaius-Augustus/Augustus) and GeMoMa[71]. Annotations from the *C. parvum* reference genomes (IOWA-CDC and IOWA-BGF) were used to train AUGUSTUS for the prediction, while the annotations from the reference genomes were transferred to the *C. hominis* IbA12G3 based on conserved

synteny. The uncharacteristic genes were annotated by searching the NCBI non-redundant protein database using Blastp v2.10.1 with an e-value of 1×10^{-6}.

Genome assembly of other isolates and comparative genomic analyses

All trimmed Illumina reads were assembled de novo using CLC Genomics Workbench with a word size of 63 and bubble size of 400. Genomes were also assembled using SPAdes 3.1 (http://cab.spbu.ru/software/spades/) with a kmer of 63 and the careful mode. The assemblies were aligned and sorted against the chromosome-level reference genome of *C. hominis* IbA12G3-monkey using Mauve's "move contig" tool to obtain the size of the final genome assemblies of the isolates. The genomes with two or more types of *gp60* or 18S rRNA sequences and/or an assembly larger than 9.1 Mb were considered as mixed infections and excluded from further analysis.

The trimmed reads were mapped to the new reference *C. hominis* 46287 genome and the other isolates using the BWA-MEM v0.7.17. Duplicate reads were marked and removed in the bam files using the MarkDuplicates algorithm of picard v3.1.1 (https://broadinstitute.github.io/picard/) with REMOVE_DUPLICATES=true, CREATE_INDEX=true, and ASSUME_SORTED=true. Genome coverage and sequence depth were estimated using the mpileup algorithm of SAMtools v1.7 (http://samtools.sourceforge.net/). The mpileup and call algorithms of BCFtools v1.12 (https://samtools.github.io/bcftools/) were used to generate a VCF file of sequence variants, with parameters -c 50, -d 500, and -ploidy 1. Low quality SNPs (QUAL < 30, FORMAT/DP < 3 and AVG FORMAT/DP > 25) were filtered out using the BCFtools filter algorithm. The bam file of 46287 isolates from the read mapping was visualized in IGV, to look for pile-ups representing the repetitive sequence regions in the genome[58]. To avoid the introduction of artificial whole-genome SNPs (wgSNPs) estimated by short reads from different family members randomly mapped to a locus[72], the wgSNPs were divided into core wgSNPs and secondary wgSNPs using the view algorithm of BCFtools. The core wgSNPs cover most of the genomic regions (Supplementary Table 4), except for the subtelomeric sequences and repetitive sequences, such as the CPCDC_3g4270/CPCDC_3g4260 locus and the rRNA units[18].

Population genetic analyses

Relationships among *C. hominis* isolates were examined using principal component analysis (PCA), STRUCTURE, phylogenetic, and IBD[73] analyses. Core wgSNPs identified across all *C. hominis* genomes were pruned using PLINK v2.00 (http://zzz.bwh.harvard.edu/plink/) with -mac 3 and -geno of 0.2 to generate a dataset of high-quality core wgSNPs. The dataset was subjected to PCA analysis using LDAK v5.2[74]. Clustering among the isolates was visualized using the R package 'ggplot2' (https://www.rdocumentation.org/packages/ggplot2). The core wgSNPs among the animal *C. hominis* isolates were extracted from the previous high-quality core wgSNPs dataset with --max-missing 1 using Vcftools v0.1.16 (https://vcftools.github.io/index.html). These wgSNPs were used for PCA analysis using LDAK. A set of unlinked sites was generated with 10-kb sliding windows using PLINK with --indep-pairwise of 10 5 0.4. These wgSNPs were also subjected to STRUCTURE analysis using Structure v.2.3.4 (https://web.stanford.edu/group/pritchardlab/structure.html), with the best number of subpopulations (K value) being calculated using Pophelper v2.3.1 (http://www.royfrancis.com/pophelper/ articles/index.html). The results of the STRUCTURE analysis were visualized using the CLUMPAK server (https://tau.evolseq.net/clumpak/). The wgSNPs among other isolates were extracted from the wgSNPs using Vcftools. ML trees were generated from the wgSNPs using RAxM-NG with a substitution model and 1000 replicates of bootstrapping. The substitution model was estimated using jModelTest v2.1.10 based on values from Akaike Information Criterion[75]. The IBD analysis was used to identify isolates with shared ancestry using hmmIBD v2.0.4. Isolates with IBD sharing greater than the mean + SD of their genomes were considered as related ones. Relatedness networks for pairs of isolates were generated using Cytoscape v3.9.1 (https://cytoscape.org/).

Gene flow between populations was assessed using phylogenetic network analysis and topology weighting. The neighbor-net algorithm of SplitsTree6[76] (https://github.com/husonlab/splitstree6) was used to generate phylogenetic networks. Twist.py (https://github.com/simonhmartin/twisst) was used to estimate topology weighting with a 50 bp window across eight chromosomes to examine phylogenetic relationships across the genome among three populations of *C. hominis* isolates and an outgroup. The genome-wide average weighting of each topology and the distribution of topology weightings were visualized using the R package 'PlotTwist' (https://github.com/JoachimGoedhart/PlotTwist). To identify the introgressed genomic regions across the whole genome, absolute divergence (*dxy*) was estimated across the genome using a set of SNPs. The *dxy* values between populations or isolates were calculated in a 10-kb sliding window using popgenWindows.py in Genomics_general (https://github.com/simonhmartin/genomics_general), and visualized in line plots using the R package 'ggplot2'. In addition, the modified *fd* test with 100-kb sliding windows and 10-kb steps was performed as described using ABBABABAwindows.py in Genomics_general. Three populations and one outgroup were used with the relationship (((P1, P2), P3), O), where P1 is closer to P2 than P3. Positive *fd* values were considered as introgression signals.

Structural differences between *C. hominis* genomes

All short-read genomes were aligned against the new *C. hominis* IbA12G3 genome and *C. parvum* reference genomes (IOWA-CDC and IOWA-BGF) using the "move contig" tool in Mauve v2.3.1. Structural variations (SVs) and copy number variations (CNVs) between the genomes were then manually detected. The CNVs between the genomes of animal and human isolates were verified by read mapping, blast analyses, and PCR. Reads were mapped to the reference genomes containing the genomic sequences of the SNVs using minimap2 for long reads and BWA-MEM for short reads (see above). These genomic sequences were used to construct databases using makeblastdb in Blast+ v2.16.0, and the assemblies were compared to the databases using blastn. In addition, the reads from the sequencing were used to build nucleotide databases, and the protein sequences of the CNVs were compared to them using tblastn. Primers for PCR amplification are showed in Supplementary Table 8. Nucleotide diversity (*Pi*) was used to measure the degree of variability of genomic regions in a group. It was calculated among *C. hominis* isolates using the Vcftools v0.1.16 with a 10-kb sliding window.

Experimental infection of mice with *C. hominis*

Prior to experimental infection, all GKO mice were verified to be *Cryptosporidium*-free by nested PCR analysis of the 18S rRNA gene. All cages, water bottles, and bedding were autoclaved before use. All infection studies were performed on mice between 4 and 5 weeks of age, with equal numbers of male and female mice. Mice in infected groups received *Cryptosporidium* oocysts by oral gavage, while mice in uninfected control groups received PBS. All mice in these infection studies were housed individually in separate cages (one mouse per cage). Oocyst shedding was measured by qPCR analysis of the 18S rRNA gene as described[45]. Sequence analysis of the *gp60* gene was used to verify the identity of *C. hominis* periodically throughout the course of infection.

To isolate *C. hominis* from monkeys, three GKO mice were used for infection with oocysts from each clinical sample. All mice were inoculated with oocysts that were freshly purified from clinical samples. Infected mice were housed separately from each other in individual cages and the oocyst shedding of each mouse was monitored for

60 days. All comparative infection studies and other assays used oocysts freshly purified from experimentally infected mice. Due to the high cost of GKO mice and the variable freshness and oocyst load of clinical samples, only three monkey isolates were maintained in our laboratory, including ImA20, InA17, and IbA12G3. These isolates represent three closely related subtype families. We also successfully infected GKO mice with another ImA18 isolate, but had to make the decision to maintain only three isolates due to cost and labor. The equine isolates were not tested for mouse infectivity because they have fewer subtelomeric *INS* and *MEDLE* genes than the monkey isolates (less likely to infect mice) and because of the small number of oocysts purified from clinical samples.

To evaluate the virulence and infectivity of *C. hominis* isolates from monkeys, 30 GKO mice were divided into five groups ($n = 6$ for each group), including the IbA12G3 infection group, the InA17 infection group, the ImA20 infection group, the *C. parvum* IIa-Waterborne infection group, and the uninfected control group. All mice except those in the control group were inoculated with 10^4 oocysts that were freshly purified from maintenance mice. The oocyst shedding pattern and body weight of each mouse were monitored for 80 days.

To determine the median infectious dose (ID_{50}) of *C. hominis* isolates from monkeys, 120 GKO mice were randomly assigned to seven groups for each of the three subtypes described above, each receiving different doses of oocysts, including 1000 ($n = 4$), 100 ($n = 4$), 50 ($n = 6$), 10 ($n = 6$), 5 ($n = 8$), 1 ($n = 8$), and 0 ($n = 4$). Oocyst shedding in each mouse was monitored for 15 days using qPCR as described above, and the ID50 of each isolate was calculated using logistic regression.

Study of pathological changes induced by the *C. hominis* isolates

Fifteen 4-week-old GKO mice were randomly divided into five groups ($n = 3$ for each group): the *C. hominis* IbA12G3 group, the InA17 group, the ImA20 group, the *C. parvum* IIa-Waterborne group, and the uninfected control group. Mice in the infected groups were inoculated with 10^4 oocysts freshly purified from maintenance mice. Mice were sacrificed at DPI 12 for the assessment of parasite burden and pathological changes in the ileum and colon by histology and scanning electron microscopy (SEM) as described[45].

Construction of CRISPR/Cas9 plasmids for genetic modification of *C. hominis*

The CRISPR/Cas9 technology was used to generate transgenic *C. hominis* lines with fluorescent and nanoluciferase reporters as described[77]. CRISPR/Cas9 plasmids were generated by adding a single-guide RNA (sgRNA) targeting the gene of interest to the linear Cas9 plasmid using the Gibson assembly kit. The sgRNAs were designed to target the *TK* gene in the IbA12G3 genome and the *UPRT* gene in the ImA20 genome, while the tagging plasmids contained the TK-Tdtomato-Nluc-P2A-neo and UPRT-mNeonGreen-Nluc-P2A-neo repair cassettes and DNA sequences homologous to the regions targeted by the sgRNAs.

A sgRNA targeting the 3′ end of the CHM_5g5480 was used to knock out the CHM_5g5480 and CHM_5g5490 genes. The homology repair fragments were PCR amplified from the IbA12G3 genome, consisting of 1202 bp of the CHM_5g5490 5′ UTR and 1358 bp of the CHM_5g5480 3′ UTR. The fragments were assembled with a reporter-drug selection cassette amplified from TK-Tdtomato-Nluc-P2A-neo. In addition, to knock out all passible copies of the CHM_5g5511, the sgRNA and the homology repair fragments were designed based on the identical regions of the different copies of CHM_5g5511. A gRNA targeting the 5′ end of the genes was used. The 783 bp sequence upstream and 559 bp sequence downstream of the 310-bp 5′ end of the CHM_5g5511 were amplified from the IbA12G3 genome and used to insert the dual reporter-drug selection cassette (Tdtomato-Nluc-P2A-neo). The sgRNA and repair fragments targeted conserved sequences based on visual inspection of the PacBio read mapping to the reference *C. hominis* genome.

Generation of transgenic parasites

Fresh oocysts were treated with 10% Clorox (0.5% sodium hypochlorite) on ice for 10 min, followed by treatment with sodium taurodeoxycholate at 37 °C for 60 min. The excysted sporozoites of IbA12G3 and ImA20 isolates were electroporated with TK-Tdtomato-Nluc-P2A-neo and UPRT-mNeonGreen-Nluc-P2A-neo plasmids, respectively, while former strain also was electroporated with ΔCHM_5g5511- Tdtomato-Nluc-P2A-neo or ΔCHM_5g5480/5490-Tdtomato-Nluc-P2A-neo plasmids, together with the corresponding sgRNAs. The electroporated parasites were orally gavaged to three GKO mice that were previously treated orally with sodium bicarbonate to neutralize gastric acid. Transgenic parasites were selected by treating GKO mice with 16 g/L paromomycin in the drinking water starting 18 h after the oocyst inoculation. They were verified by measurement of nanoluciferase activity using Cytation (BioTek, Synergy H1), fluorescence microscopic analysis of oocysts purified from fecal pellets of infected mice under a microscope (Olympus, BX53), PCR amplification, and WGS of the purified oocysts.

Experimental infection of mice with transgenic *C. hominis*

To evaluate the infectivity of transgenic *C. hominis* isolates to GKO mice, 16 four-week-old GKO mice were divided into four groups ($n = 4$ for each group), including the tagged IbA12G3 infection group (IbA12G3-Tdtomato), the CHM_5g5511 KO group (IbA12G3-ΔCHM_5g5511), the CHM_5g5480 and CHM_5g5490 KO group (IbA12G3-ΔCHM_5g5480/5490), and the uninfected control group. All mice except those in the control group were inoculated with 10^4 oocysts. Mice in the infection study were housed in individual cages, and oocyst shedding patterns and body weight of each mouse were monitored for 15 days.

Genetic cross between *C. hominis* isolates

Three 4-week-old GKO mice were simultaneously infected with 10^4 oocysts of each tdTomato-tagged IbA12G3 and mNeonGreen-tagged ImA20 lines. Fecal samples were collected daily starting at DPI 5. Oocysts were purified from feces by sucrose and cesium chloride gradient centrifugation and examined by immunofluorescence microscopy. Yellow (both tdTomato and mNeonGreen positive) oocysts were sorted and harvested by flow cytometry using FACSAria III (BD Biosciences, USA), and the data were analyzed using BD FACSDiva v8.0.3. To increase the yield of yellow progeny, three 4-week-old GKO mice were infected with 10^5 oocysts of mNeonGreen-tagged ImA20 lines and 10^4 oocysts of tdTomato-tagged IbA12G3.

Statistical analysis

All statistical analyses were performed using the Prism v10 unless otherwise noted. The Mann–Whitney *U* test was used to compare the means of two groups, and the Wilcoxon signed-rank test was used to compare the growth curves of two groups.

Reporting summary

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

Data availability

All sequence data have been deposited in the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/sra/) under the BioProject accession number PRJNA1033585. The final genome assembly of *C. hominis* IbA12G3_46287 has been submitted to GenBank under accession numbers JBGCUF000000000. Accession codes for the genomic sequence datasets analyzed in this study are provided in Supplementary Data 1. Source data are provided with this paper. The.vcf file used

in this study is available in the Figshare database[78]. Source data are provided with this paper.

Code availability

All analyses are described and cited in the Methods. Custom codes used in this work have been deposited on GitHub (https://github.com/Wanyi-Huang/PopGene_Plot).

References

1. 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).
2. Ryan, U. M., Feng, Y., Fayer, R. & Xiao, L. Taxonomy and molecular epidemiology of *Cryptosporidium* and *Giardia* - a 50 year perspective (1971-2021). *Int. J. Parasitol.* **51**, 1099–1119 (2021).
3. Morgan-Ryan, U. M. et al. *Cryptosporidium hominis* n. sp. (Apicomplexa: Cryptosporidiidae) from Homo sapiens. *J. Eukaryot. Microbiol.* **49**, 433–440 (2002).
4. Widmer, G., Koster, P. C. & Carmena, D. *Cryptosporidium hominis* infections in non-human animal species: revisiting the concept of host specificity. *Int. J. Parasitol.* **50**, 253–262 (2020).
5. Chen, L. et al. *Cryptosporidium parvum* and *Cryptosporidium hominis* subtypes in crab-eating macaques. *Parasit. Vectors* **12**, 350 (2019).
6. Zhao, W. et al. Molecular prevalence and subtyping of *Cryptosporidium hominis* among captive long-tailed macaques (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*) from Hainan Island, southern China. *Parasit. Vectors* **12**, 192 (2019).
7. Jian, F. et al. Common occurrence of *Cryptosporidium hominis* in horses and donkeys. *Infect. Genet. Evol.* **43**, 261–266 (2016).
8. Karim, M. R. et al. Multilocus typing of *Cryptosporidium* spp. and *Giardia* duodenalis from non-human primates in China. *Int. J. Parasitol.* **44**, 1039–1047 (2014).
9. Li, F. et al. Different distribution of *Cryptosporidium* species between horses and donkeys. *Infect. Genet. Evol.* **75**, 103954–103954 (2019).
10. Wang, W. et al. Prevalence and genotypic identification of *Cryptosporidium* in free-ranging and farm-raised donkeys (Equus asinus asinus) in Xinjiang, China. *Parasite* **27**, 45 (2020).
11. Xiao, L. & Feng, Y. Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. *Food Waterborne Parasitol.* **8-9**, 14–32 (2017).
12. Feng, Y., Ryan, U. M. & Xiao, L. Genetic diversity and population structure of *Cryptosporidium*. *Trends Parasitol.* **34**, 997–1011 (2018).
13. Li, J. et al. An investigation of parasitic infections and review of molecular characterization of the intestinal protozoa in nonhuman primates in China from 2009 to 2015. *Int. J. Parasitol. Parasites Wildl.* **6**, 8–15 (2017).
14. Toriro, R. et al. Outbreak of diarrhea caused by a novel *Cryptosporidium hominis* subtype during british military training in Kenya. *Open Forum Infect. Dis.* **11**, ofae001 (2024).
15. Lebbad, M., Winiecka-Krusnell, J., Insulander, M. & Beser, J. Molecular characterization and epidemiological investigation of *Cryptosporidium hominis* IkA18G1 and *C. hominis* monkey genotype IiA17, two unusual subtypes diagnosed in Swedish patients. *Exp. Parasitol.* **188**, 50–57 (2018).
16. Lebbad, M., Winiecka-Krusnell, J., Stensvold, C. R. & Beser, J. High diversity of *Cryptosporidium* species and subtypes identified in cryptosporidiosis acquired in Sweden and abroad. *Pathogens* **10**, 523 (2021).
17. Xiao, L. et al. Presence of heterogeneous copies of the small subunit rRNA gene in *Cryptosporidium parvum* human and marsupial genotypes and *Cryptosporidium* felis. *J. Eukaryot. Microbiol.* **46**, 44s–45s (1999).
18. Huang, W. et al. Sequence introgression from exogenous lineages underlies genomic and biological differences among *Cryptosporidium parvum* IOWA lines. *Water Res.* **254**, 121333 (2024).
19. Baptista, R. P. et al. Long-read assembly and comparative evidence-based reanalysis of *Cryptosporidium* genome sequences reveals expanded transporter repertoire and duplication of entire chromosome ends including subtelomeric regions. *Genome Res.* **32**, 203–213 (2021).
20. Baptista, R. P., Xiao, R., Li, Y., Glenn, T. C. & Kissinger, J. C. New T2T assembly of *Cryptosporidium parvum* IOWA annotated with reference genome gene identifiers. Preprint at *bioRxiv* (2023).
21. Huang, W. et al. Multiple introductions and recombination events underlie the emergence of a hyper-transmissible *Cryptosporidium hominis* subtype in the USA. *Cell Host Microbe* **31**, 112–123 e114 (2023).
22. Sikora, P. et al. Genomic variation in IbA10G2 and other patient-derived *Cryptosporidium hominis* subtypes. *J. Clin. Microbiol.* **55**, 844–858 (2017).
23. 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).
24. Feng, Y. et al. Comparative genomic analysis of the IId subtype family of *Cryptosporidium parvum*. *Int. J. Parasitol.* **47**, 281–290 (2017).
25. Mallon, M. E., MacLeod, A., Wastling, J. M., Smith, H. & Tait, A. Multilocus genotyping of *Cryptosporidium parvum* Type 2: population genetics and sub-structuring. *Infect. Genet. Evol.* **3**, 207–218 (2003).
26. Feng, Y., Lal, A. A., Li, N. & Xiao, L. Subtypes of *Cryptosporidium* spp. in mice and other small mammals. *Exp. Parasitol.* **127**, 238–242 (2011).
27. Wang, T. et al. Sympatric recombination in zoonotic *Cryptosporidium* leads to emergence of populations with modified host preference. *Mol. Biol. Evol.* **39**, msac150 (2022).
28. Corsi, G. I. et al. Recent genetic exchanges and admixture shape the genome and population structure of the zoonotic pathogen *Cryptosporidium parvum*. *Mol. Ecol.* **0**, 1–13 (2022).
29. Nader, J. L. et al. Evolutionary genomics of anthroponosis in *Cryptosporidium*. *Nat. Microbiol.* **4**, 826–836 (2019).
30. 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).
31. Braima, K. et al. Molecular analysis of cryptosporidiosis cases in Western Australia in 2019 and 2020 supports the occurrence of two swimming pool associated outbreaks and reveals the emergence of a rare IbA12G3 subtype. *Infect. Genet. Evol.* **92**, 104859 (2021).
32. Millán, R., Köster, P. C., Fuentes, I. & Carmena, D. IbA12G3: First report of a rare sub-genotype in Spain. *Enferm. Infecc. Y Microbiol. Clin.* **37**, 279–281 (2019).
33. Urrea-Quezada, A. et al. Clinical Manifestations of Cryptosporidiosis and Identification of a new *Cryptosporidium* Subtype in Patients From Sonora, Mexico. *Pediatr. Infect. Dis. J.* **37**, E136–E138 (2018).
34. Guy, R. A. et al. Molecular characterization of *Cryptosporidium* isolates from humans in Ontario, Canada. *Parasit. Vectors* **14**, 69 (2021).
35. Bujila, I. et al. *Cryptosporidium* species and subtypes identified in human domestic cases through the national microbiological surveillance programme in Sweden from 2018 to 2022. *BMC Infect. Dis.* **24**, 146 (2024).
36. Bacchetti, R., Connelly, L., Browning, L. & Alexander, C. L. Changing molecular profiles of human cryptosporidiosis cases in Scotland as a result of the coronavirus disease, COVID-19 pandemic. *Br. J. Biomed. Sci.* **80**, 11462 (2023).

37. Insulander, M. et al. Molecular epidemiology and clinical manifestations of human cryptosporidiosis in Sweden. *Epidemiol. Infect.* **141**, 1009–1020 (2013).
38. Karim, M. R. et al. Occurrence and molecular characterization of *Cryptosporidium* spp. and *Giardia duodenalis* among captive mammals in the Bangladesh National Zoo. *Parasitol. Int.* **84**, 102414 (2021).
39. Su, J. et al. Differential expression of three *Cryptosporidium* species-specific MEDLE proteins. *Front. Microbiol.* **10**, 1177 (2019).
40. Fei, J. et al. Characterization of MEDLE-1, a protein in early development of *Cryptosporidium parvum*. *Parasit. Vectors* **11**, 312 (2018).
41. Zhang, S. et al. Characterization of a species-specific insulinase-like protease in *Cryptosporidium parvum*. *Front. Microbiol.* **10**, 354 (2019).
42. Vinayak, S. et al. Bicyclic azetidines kill the diarrheal pathogen *Cryptosporidium* in mice by inhibiting parasite phenylalanyl-tRNA synthetase. *Sci. Transl. Med.* **12**, eaba8412 (2020).
43. 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).
44. Liu, S. et al. Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. *BMC Genomics* **17**, 1006 (2016).
45. He, X. et al. A productive immunocompetent mouse model of cryptosporidiosis with long oocyst shedding duration for immunological studies. *J. Infect.* **84**, 710–721 (2022).
46. Jia, R. et al. High infectivity and unique genomic sequence characteristics of *Cryptosporidium parvum* in China. *PLoS Negl. Trop. Dis.* **16**, e0010714 (2022).
47. Sateriale, A. et al. A genetically tractable, natural mouse model of cryptosporidiosis offers insights into host protective immunity. *Cell Host Microbe* **26**, 135–146 e135 (2019).
48. Gaber, M. et al. Evidences of brain and lung invasion of a local water *Cryptosporidium parvum* isolate in comparison to Iowa strain: serological and immunohistochemical cytokine evaluation. *Ann. Parasitol.* **66**, 311–318 (2020).
49. Sayed, F. G., Hamza, A. I., Galal, L. A., Sayed, D. M. & Gaber, M. Virulence of geographically different *Cryptosporidium parvum* isolates in experimental animal model. *Ann. Parasitol.* **62**, 221–232 (2016).
50. Sonzogni-Desautels, K., Mead, J. R. & Ndao, M. Mouse models for use in *Cryptosporidium* infection studies and quantification of parasite burden using flow cytometry, qPCR, and histopathology. *Methods Mol. Biol.* **2052**, 229–251 (2020).
51. Ludington, J. G. & Ward, H. D. The *Cryptosporidium parvum* c-type lectin CpClec mediates infection of intestinal epithelial cells via interactions with sulfated proteoglycans. *Infect. Immun.* **84**, 1593–1602 (2016).
52. Li, M. et al. Variant surface protein GP60 contributes to host infectivity of *Cryptosporidium parvum*. *Commun. Biol.* **7**, 1175 (2024).
53. Wilke, G. et al. A stem-cell-derived platform enables complete *Cryptosporidium* development in vitro and genetic tractability. *Cell Host Microbe* **26**, 123–134 e128 (2019).
54. 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).
55. 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).
56. Shaw, S. et al. Genetic crosses within and between species of *Cryptosporidium*. *Proc. Natl. Acad. Sci. USA* **121**, e2313210120 (2024).
57. Tanriverdi, S. et al. Inferences about the global population structures of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *Appl. Environ. Microbiol.* **74**, 7227–7234 (2008).
58. Baptista, R. P. & Kissinger, J. C. Is reliance on an inaccurate genome sequence sabotaging your experiments? *PLoS Pathog.* **15**, e1007901 (2019).
59. Xiao, L. et al. Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater. *Appl. Environ. Microbiol.* **67**, 1097–1101 (2001).
60. Alves, M. et al. Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. *J. Clin. Microbiol.* **41**, 2744–2747 (2003).
61. Arrowood, M. J. & Donaldson, K. Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium chloride gradients. *J. Eukaryot. Microbiol.* **43**, 89S (1996).
62. Isaza, J. P. et al. Revisiting the reference genomes of human pathogenic *Cryptosporidium* species: reannotation of *C. parvum* Iowa and a new *C. hominis* reference. *Sci. Rep.* **5**, 16324 (2015).
63. Hadfield, S. J. et al. Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples. *BMC Genomics* **16**, 650 (2015).
64. Gilchrist, C. A. et al. Genetic diversity of *Cryptosporidium hominis* in a Bangladeshi community as revealed by whole-genome sequencing. *J. Infect. Dis.* **218**, 259–264 (2018).
65. Amid, C. et al. The COMPARE data hubs. *Database* **2019**, 1–14 (2019).
66. Arias-Agudelo, L. M., Garcia-Montoya, G., Cabarcas, F., Galvan-Diaz, A. L. & Alzate, J. F. Comparative genomic analysis of the principal *Cryptosporidium* species that infect humans. *PeerJ* **8**, e10478 (2020).
67. Tichkule, S. et al. Comparative genomics revealed adaptive admixture in *Cryptosporidium hominis* in Africa. *Microb. Genom.* **7**, 000493 (2020).
68. Knox, M. A., Garcia, R. J. & Hayman, D. T. S. Draft genome assemblies of two *Cryptosporidium hominis* isolates from New Zealand. *Microbiol. Resour. Announc.* **10**, e0036321 (2021).
69. Cabarcas, F. et al. *Cryptosporidium hominis* phylogenomic analysis reveals separate lineages with continental segregation. *Front. Genet.* **12**, 740940 (2021).
70. Lu, N. et al. Exploration of whole genome amplification generated chimeric sequences in long-read sequencing data. *Brief. Bioinform.* **24**, bbad275 (2023).
71. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
72. Agyabeng-Dadzie, F., Xiao, R. & Kissinger, J. C. *Cryptosporidium* genomics - current understanding, advances, and applications. *Curr. Trop. Med. Rep.* **11**, 92–103 (2024).
73. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 196 (2018).
74. Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nat. Genet.* **52**, 458–462 (2020).
75. Posada, D. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
76. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
77. Vinayak, S. et al. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* **523**, 477–480 (2015).
78. Huang, W., Xiao, L. & Feng, Y. Multicopy subtelomeric genes underlie animal infectivity of divergent *Cryptosporidium hominis* subtypes Data sets. *figshare* https://doi.org/10.6084/m9.figshare.24566845 (2024).

Acknowledgements
This work was supported in part by the National Key Research and Development Program of China (2022YFD1800200 to L.X.), Guangdong Major Project of Basic and Applied Basic Research (2020B0301030007 to L.X.), National Natural Science Foundation of China (32030109 to L.X.), 111 Project (D20008 to L.X.), and Double First-Class Discipline Promotion Project (2023B10564003 to L.X.). We thank the Instrumental Analysis & Research Center, South China Agricultural University for electron microscopy sample processing and image acquisition.

Author contributions
Conceptualization: Y.F. and L.X. Methodology and investigation: Wanyi, W.H., Y.H., Y.T., L.S., Z.Y., H.L., H.C., Y.G., and N.L. Comparative genomic and evolutionary genetic analyses: Wanyi, M.C., T.H., and T.W. Genetic Engineering: W.H. and Y.H. Supervision: Y.G., Y.F., and L.X. Writing—original draft: Wanyi, Y.G., 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/s41467-024-54995-4.

Correspondence and requests for materials should be addressed to Yaqiong Guo, Lihua Xiao or Yaoyu Feng.

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