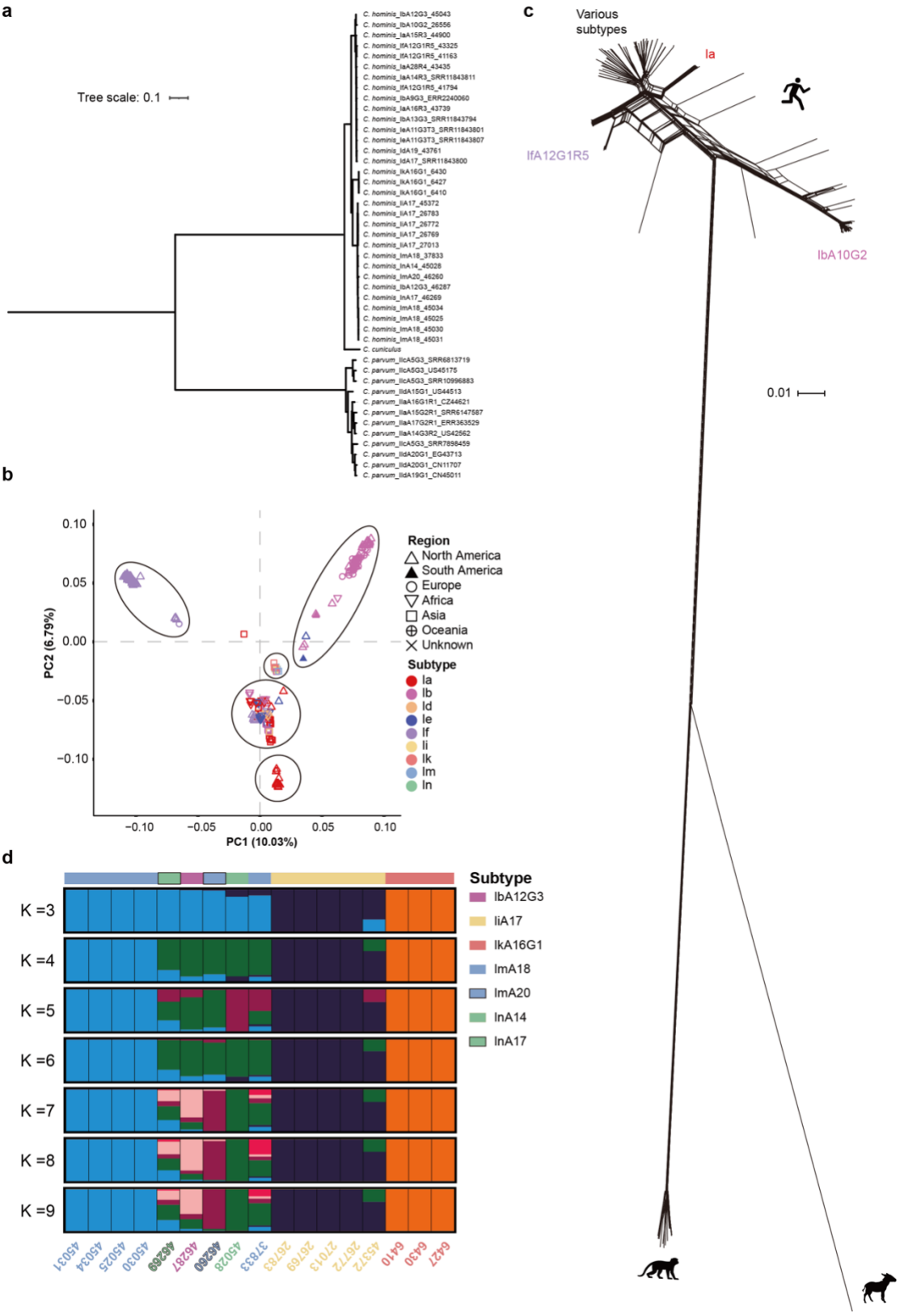
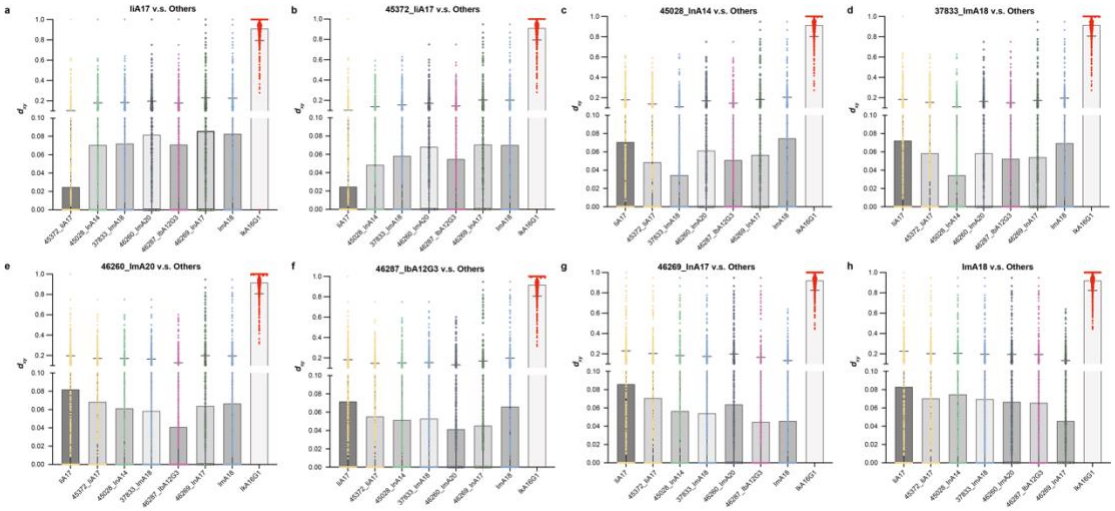
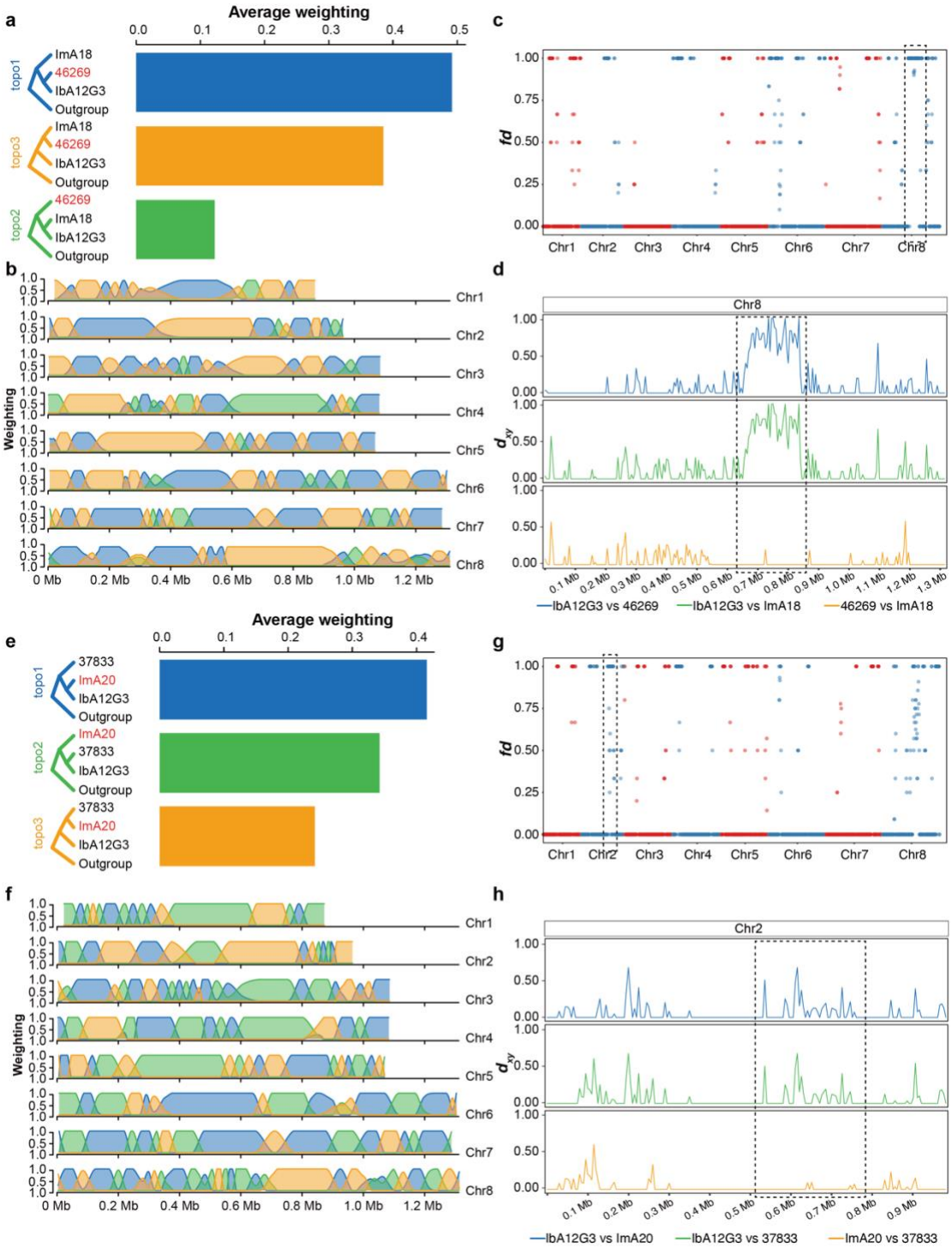
Extended data figures



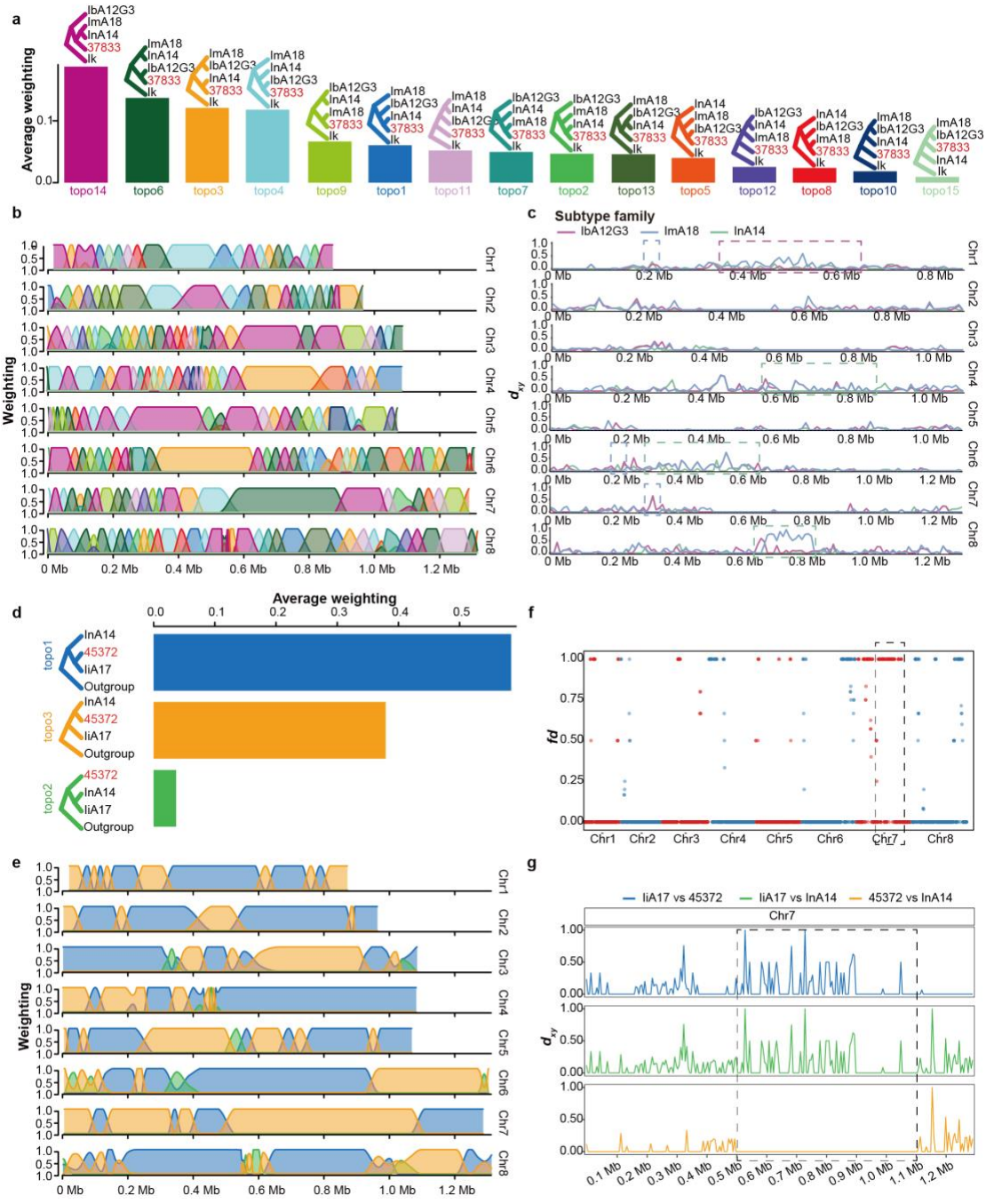
Extended Data Fig. 1 | Population subdivision and gene flow among *Cryptosporidium hominis*. a, Phylogenetic relationships of isolates *C. hominis, C. parvum*, and *C. cuniculus* inferred by maximum likelihood analysis of 291,371 wgSNPs. The ML tree shows the formation of species-segregated clades. **b**, Principal component analysis (PCA) of *C. hominis* divergent isolates based on SNPs, in which PC1 and PC2 account for variability among isolates. The colors of the symbols represent the *C. hominis* subtype families. **c**, Phylogenetic network of the *C. hominis* isolates based on 36,459 wgSNPs. The parallel edges in the network are suggestive of gene flows among isolates. **d**, STRUCTURE plots constructed using K = 3 to 9 indicate the representing the percentage of shared ancestry among the *C. hominis* animal isolates.



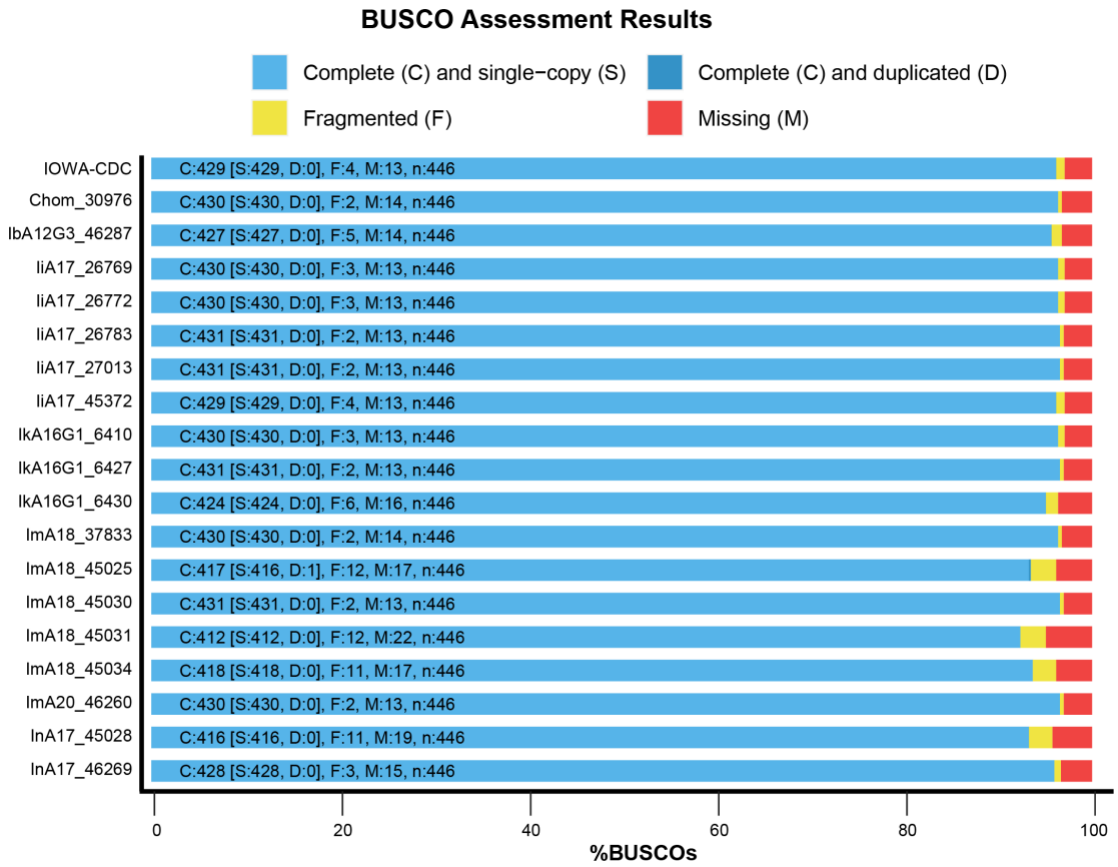
Extended Data Fig. 2 | Absolute divergence (*dxy*) between each pair of *C. hominis* genomes. a-h, Points with different colors indicate the *dxy* values of each 1,000 region across chromosomes. The top lines of the bars represent the mean *dxy* values between each pair of *C. hominis* genomes, while the error bars show the SD values. Low *dxy* values indicate genomes are closely related.


a
Average weighting
0.0 0.1 0.2 0.3 0.4 0.5
topo1
ImA18
46269
IbA12G3
Outgroup
topo3
ImA18
46269
IbA12G3
Outgroup
topo2
46269
ImA18
IbA12G3
Outgroup
b
Weighting
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8
0 Mb 0.2 Mb 0.4 Mb 0.6 Mb 0.8 Mb 1.0 Mb 1.2 Mb
c
fd
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8
d
Chr8
dxy
IbA12G3 vs 46269
IbA12G3 vs ImA18
46269 vs ImA18
e
Average weighting
0.0 0.1 0.2 0.3 0.4
topo1
37833
ImA20
IbA12G3
Outgroup
topo2
ImA20
37833
IbA12G3
Outgroup
topo3
37833
ImA20
IbA12G3
Outgroup
f
Weighting
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8
0 Mb 0.2 Mb 0.4 Mb 0.6 Mb 0.8 Mb 1.0 Mb 1.2 Mb
g
fd
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8
h
Chr2
dxy
IbA12G3 vs ImA20
IbA12G3 vs 37833
ImA20 vs 37833

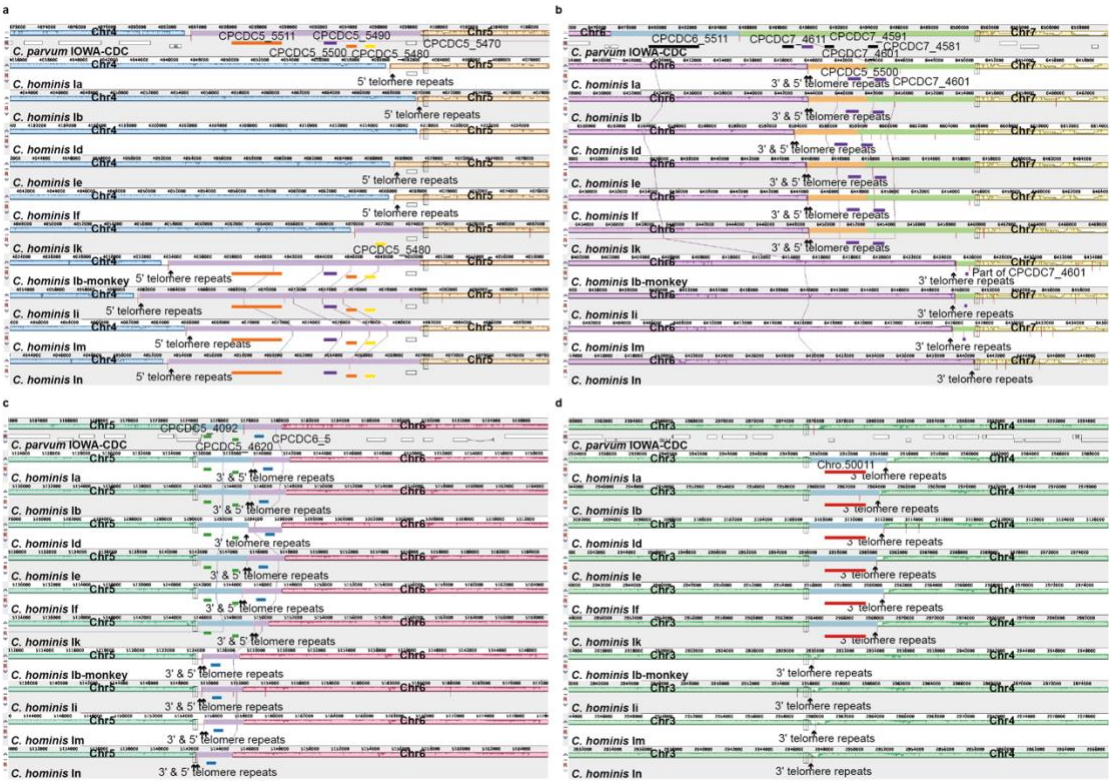
Extended Data Fig. 3 | Origin of *Cryptosporidium hominis* ImA18_37833 and IiA17_45372. **a**, Genome-wide distribution of phylogenetic relationships among the ImA18_37833, InA14_45028, IbA12G3_46287, and ImA18 population using a 50-SNP sliding window with IkA16G1 as the outgroup. The bar graph shows the genome-wide average weighting of each topology. **b**, Distribution of topology weightings (colors as in **a**) across eight chromosomes. **c**, Absolute divergence (*dxy*) between the ImA18_37833 genome and InA14_45028, IbA12G3_46287, and ImA18 genomes across the eight chromosomes. The purple, blue, and green boxes show the introgression from IbA12G3_46287, ImA18 population, and InA14_45028 sequences into the ImA18_37833 genome, respectively. **d**, Genome-wide distribution of phylogenetic relationships among the IiA17_45372, InA14_45028, and IiA17 isolates using a 50-SNP sliding window with IkA16G1 as the outgroup. The left panel shows all possible topologies, while the right panel shows the genome-wide average weighting of each topology. **e**, Distribution of topology weightings (colors as in **d**) across eight chromosomes. **f**, Values of *fd* statistics of IiA17_45372, InA14_45028, and IiA17 population calculated with a 20-kb window with 5-kb steps. The *fd* value of 1 indicate sequence introgression from InA14_45028 to IiA17_45372. The box shows a large region of such introgression. **g**, Distribution of *dxy* values between the IiA17_45372, InA14_45028, and IiA17 population in chromosome 7. The box shows a region with high sequence identity between IiA17_45372 and InA14_45028.


a
Average weighting
topo14
topo6
topo3
topo4
topo9
topo1
topo11
topo7
topo2
topo13
topo5
topo12
topo8
topo10
topo15
b
Weighting
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
0 Mb
0.2 Mb
0.4 Mb
0.6 Mb
0.8 Mb
1.0 Mb
1.2 Mb
c
Subtype family
lbA12G3
lmA18
lnA14
d
Average weighting
topo1
topo3
topo2
lnA14
45372
liA17
Outgroup
e
Weighting
f
fd
g
liA17 vs 45372
liA17 vs lnA14
45372 vs lnA14
Chr7

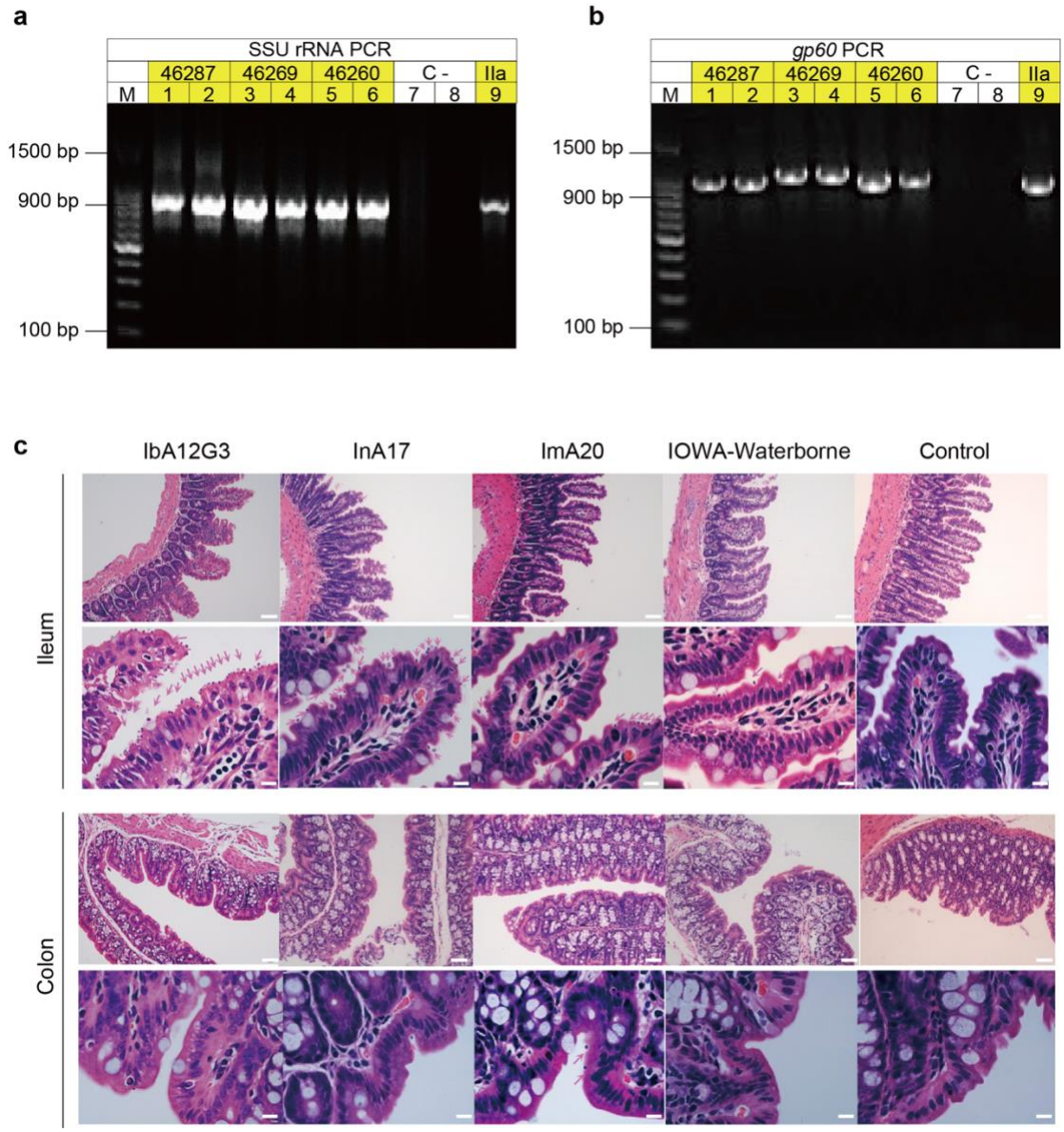
Extended Data Fig. 4 | Origin of *Cryptosporidium hominis* InA17_46269 and ImA20_46260. a,**e**, The left panel shows all possible topologies, while the right panel shows the genome-wide average weighting of each topology. **a**, Genome-wide distribution of phylogenetic relationships among the InA17_46269, monkey IbA12G3, and the ImA18 population using a 50-SNP sliding window with IkA16G1 as the outgroup. **b**, Distribution of topology weightings across eight chromosomes (colors as in A). **c**, The *fd* statistics of InA17_46269, IbA12G3, and the ImA18 population using a 20-kb window with 5-kb steps. The *fd* values equal to 1 indicate the introgression from ImA18 population to InA17_46269. **d**, The *dxy* between the InA17_46269, monkey IbA12G3, and the ImA18 population in chromosome 8. **c**,**d**, The boxes show a region from the ImA18 population to InA17_46269. **e**, Genome-wide distribution of phylogenetic relationships among the ImA18_37833, ImA20_46260, and monkey IbA12G3 using a 50-SNP sliding window with IkA16G1 as the outgroup. **f**, Distribution of topology weightings across eight chromosomes (colors as in E). **g**, The *fd* statistics of ImA18_37833, ImA20_46260, and monkey IbA12G3 isolates calculated using a 20-kb window with 5-kb steps. The *fd* values equal to 1 indicate the introgression from ImA18_37833 sequences to ImA20_46260. **h**, The *dxy* between ImA18_37833, ImA20_46260, and monkey IbA12G3 in chromosome 2. **g**,**h**, The box shows a large introgression region from ImA18_37833 to ImA20_46260.



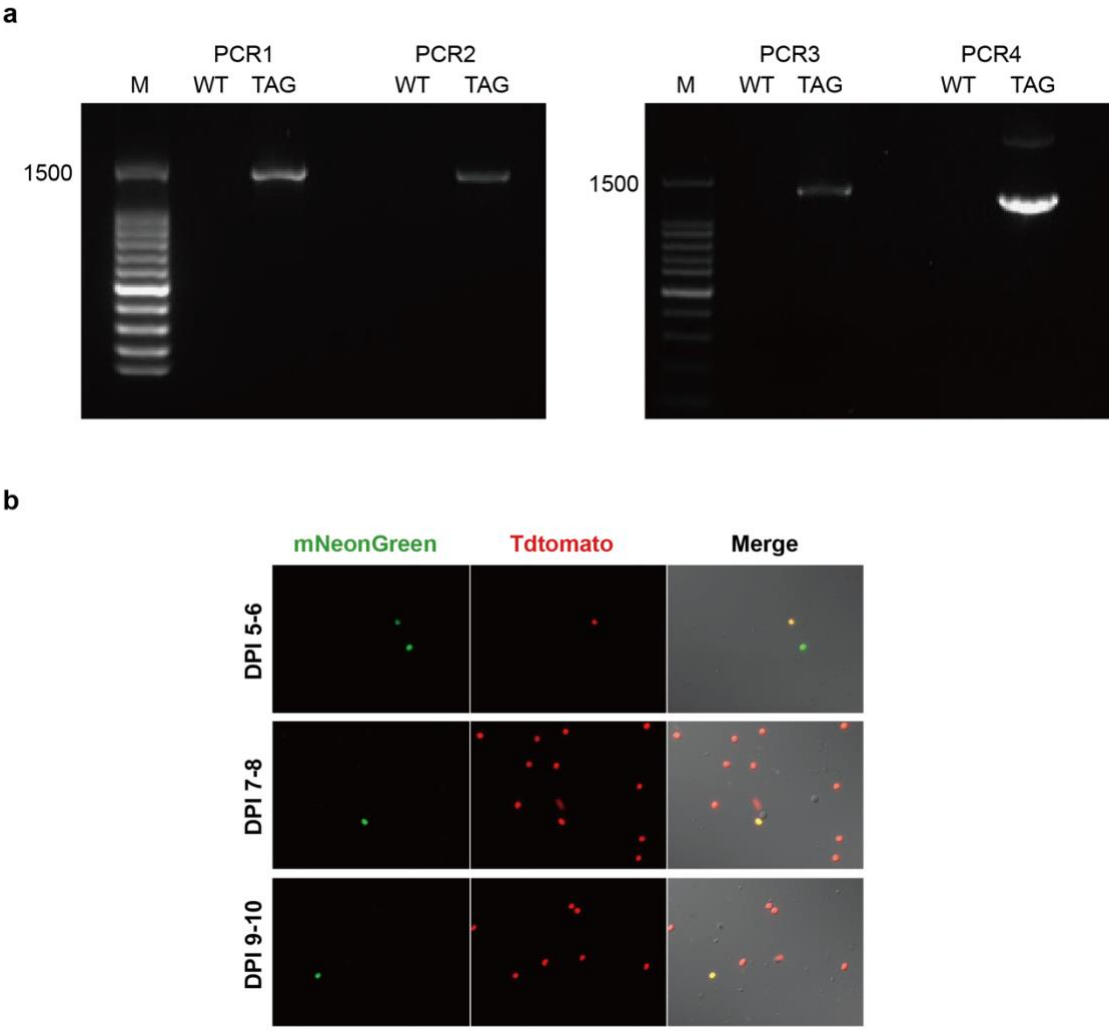
Extended Data Fig. 5 | The completeness of *Cryptosporidium* genomes using the BUSCO software. The percentages in the blue rectangle represent the completeness of the genomes.



Extended Data Fig. 6 | Deletion of genes in *Cryptosporidium hominis* human and equine isolates in comparison with *C. parvum* IOWA-CDC and *C. hominis* monkey isolates. The color blocks (known as Locally Collinear Blocks) are conserved segments of sequences internally free from genome rearrangements, whereas the inverted white peaks within each block are sequence divergence between the reference genome and the other genomes. Assembled chromosomes or contigs are bordered by vertical red lines. **a**, Deletion at the 5′ end of chromosome 5 indicate by purple blocks. The subtype-specific orthologs of CPCDC5_5511 and CPCDC5_5490 are absent in *C. hominis* human and equine isolates (orange blocks), while orthologs of and CPCDC5_5480 are absent in *C. hominis* human (yellow blocks). **b**, Deletion at the 3′ end of chromosome 6 in *C. hominis* genomes indicate by blue blocks, while deletion at the 5’ end of chromosome 7 in *C. hominis* monkey isolates indicate by green blocks. Species-specific genes are absent in *C. hominis* genomes (orthologs of CPCDC6_5511, CPCDC8_4611, CPCDC7_4591, and CPCDC7_4581), which are indicated by black blocks, while the orthologs of CPCDC7_4601 were incomplete in *C. hominis* monkey isolates (purple blocks). **c**, Deletion at the 3′ end of chromosome 5 in *C. hominis* monkey isolates indicate by blue blocks, while deletion at the 5′ end of chromosome 6 in *C. hominis* equine isolates indicate by purple blocks. Subtype-specific genes (orthologs of CPCDC5_4092 and CPCDC5_4620) are absent in *C. hominis* monkey isolates (green blocks), while the orthologs of CPCDC6_5 are absent in *C. hominis* equine isolates (blue blocks). **d**, Insertions at the 3′ end of chromosome 3 indicate by blue blocks, and the subtype-specific gene (Chro.50011) is absent in *C. parvum* and *C. hominis* monkey isolates (red blocks).



Extended Data Fig. 7 | *Cryptosporidium hominis* and *C. parvum* infections in GKO mice. a,**b**, M means the 100 bp DNA ladder. C- are negative controls from primary and secondary PCR. SSU rRNA positive samples are market in yellow. **a**, Nested PCR base on SSU rRNA. **b**, Nested PCR base on *gp60* gene. **c**, Hematoxylin and eosin microscopy images of the ileum and colon of GKO mice infected with *C. hominis* IbA12G3, InA14, ImA18, and *C. parvum* IIaA17G2R1 in comparison with the uninfected control. Mice were infected with 10,000 oocysts, killed 12 days post infection and histologically scored for pathological changes over the course of infection. The parasites are pointed with red arrows.



Extended Data Fig. 8 | Confirmation of transgenic and recombinant *C. hominis*. **a**, PCR confirmation that TK-Tdtomato-Nluc-P2A-Neo and UPRT-mNeonGreen-Nluc-P2A-Neo oocysts amplified in mice have the correct insert. **b**, Image of fluorescence microscopy of purified oocysts at different days. Recombinant progeny expressed both with Tdtomato and mNeonGreen oocysts and were yellow in merged images.