Supplementary Online Materials

for

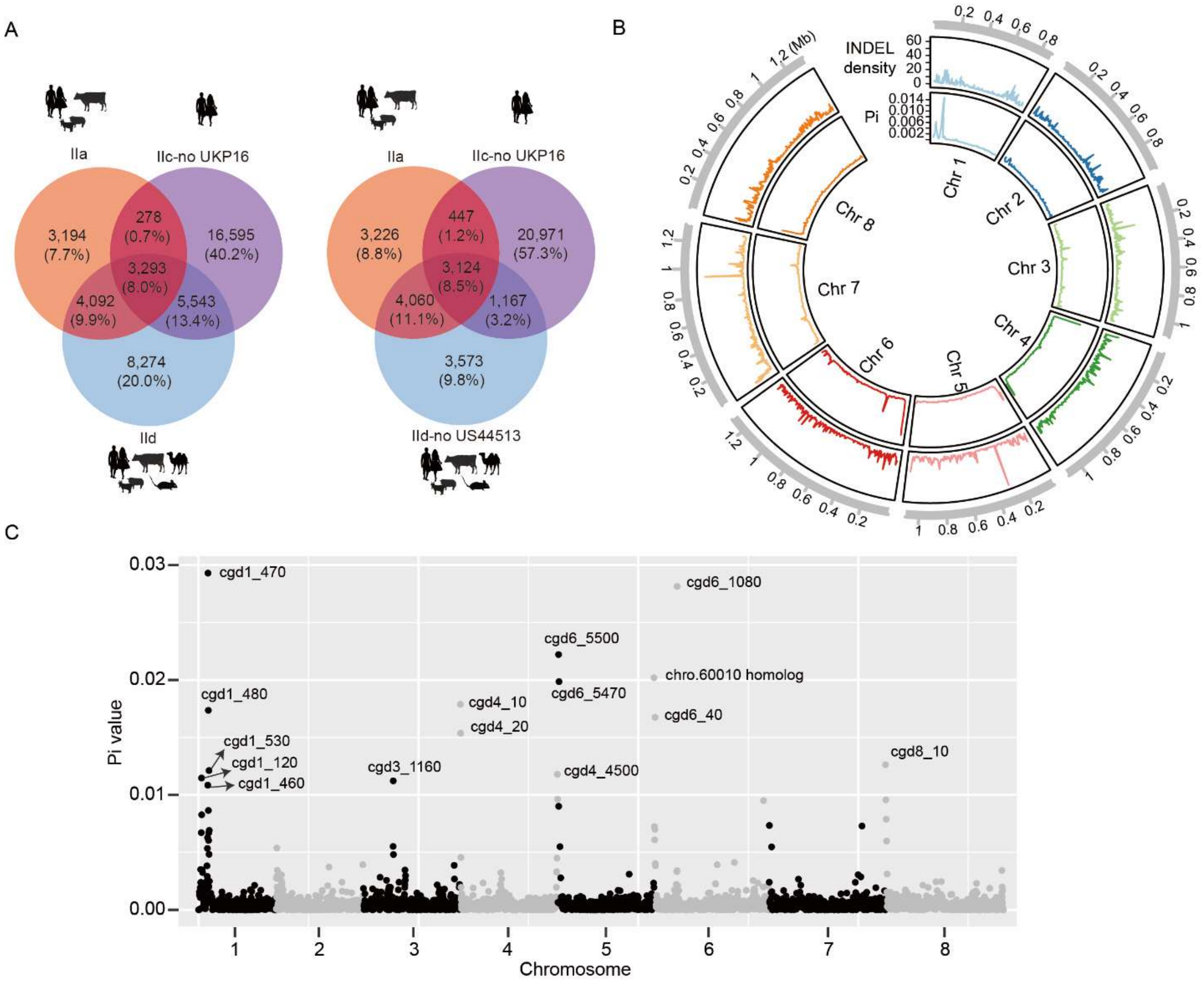
Sympatric Recombination in Zoonotic *Cryptosporidium* Leads to Emergence of Populations with Modified Host Preference

Tianpeng Wang, Yaqiong Guo, Dawn M. Roellig, Na Li, Mónica Santín, Jason Lombard, Martin Kváč, Doaa Naguib, Ziding Zhang*, Yaoyu Feng*, and Lihua Xiao*

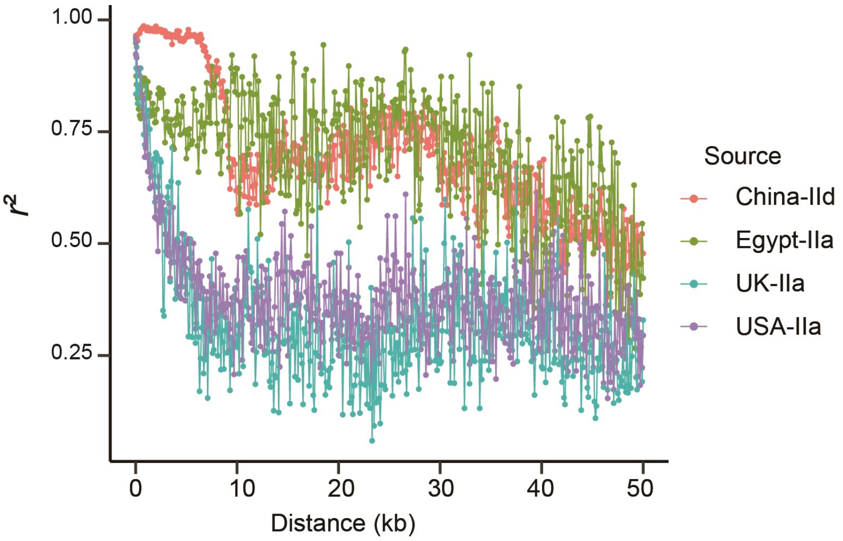
*Corresponding authors: E-mails: zidingzhang@cau.edu.cn; yyfeng@scau.edu.cn, lxiao1961@gmail.com.

This PDF file includes:

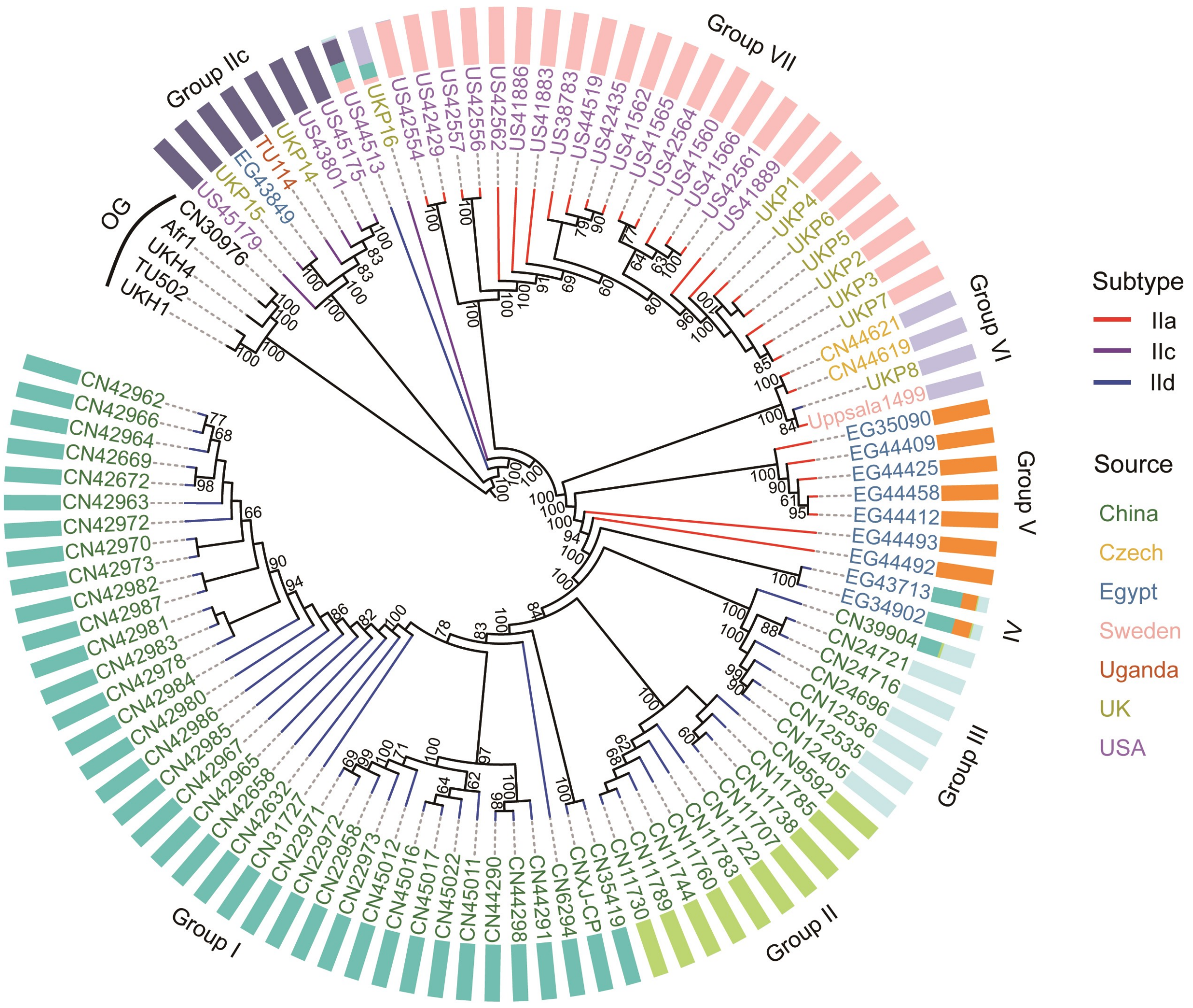
Supplementary figs. S1 to S15
Supplementary tables S1 to S8



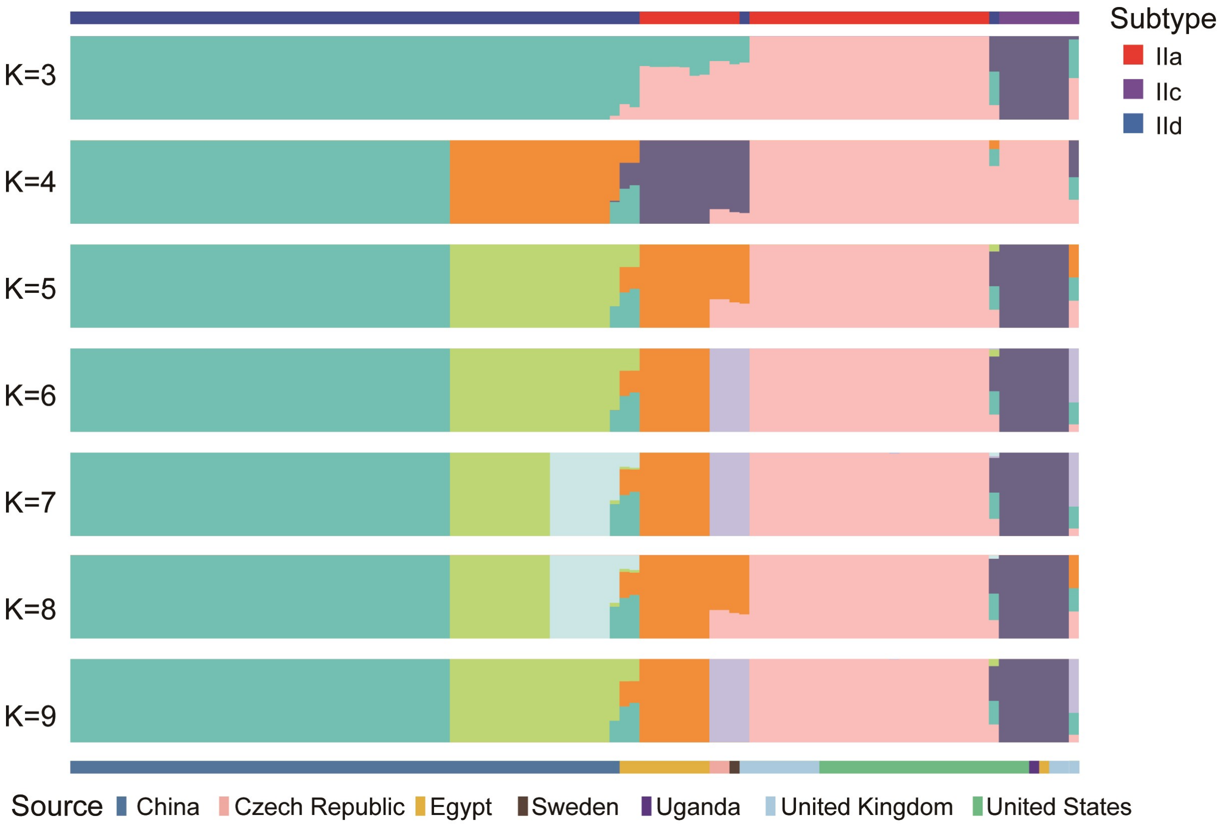
Supplementary fig. S1. Overall genomic variation in *Cryptosporidium parvum*. (*A*) Distribution of shared and subtype family-specific SNPs among three gp60 subtype families, including IIa (red), IIc (purple) and IId (blue). The SNP distribution is displayed with the zoonotic IIcA5G3j isolate UKP16 being removed from IIc (top) and the divergent isolate US44513 removed from IId (bottom). (***B***) Distribution of sequence variations along the genome. INDELs density (outer) was calculated using a 5-kb sliding window, while nucleotide diversity (Pi, inner) was calculated using window size of 10 kb with step size of 5 kb. (***C***) Nucleotide diversity (Pi) of genes within the genomes of *C. parvum* isolates analyzed in the study. Each gene is represented as a dot and colored with grey and black alternatively corresponding to the chromosomes. The 15 genes with Pi values over 0.01 are marked with their gene IDs.



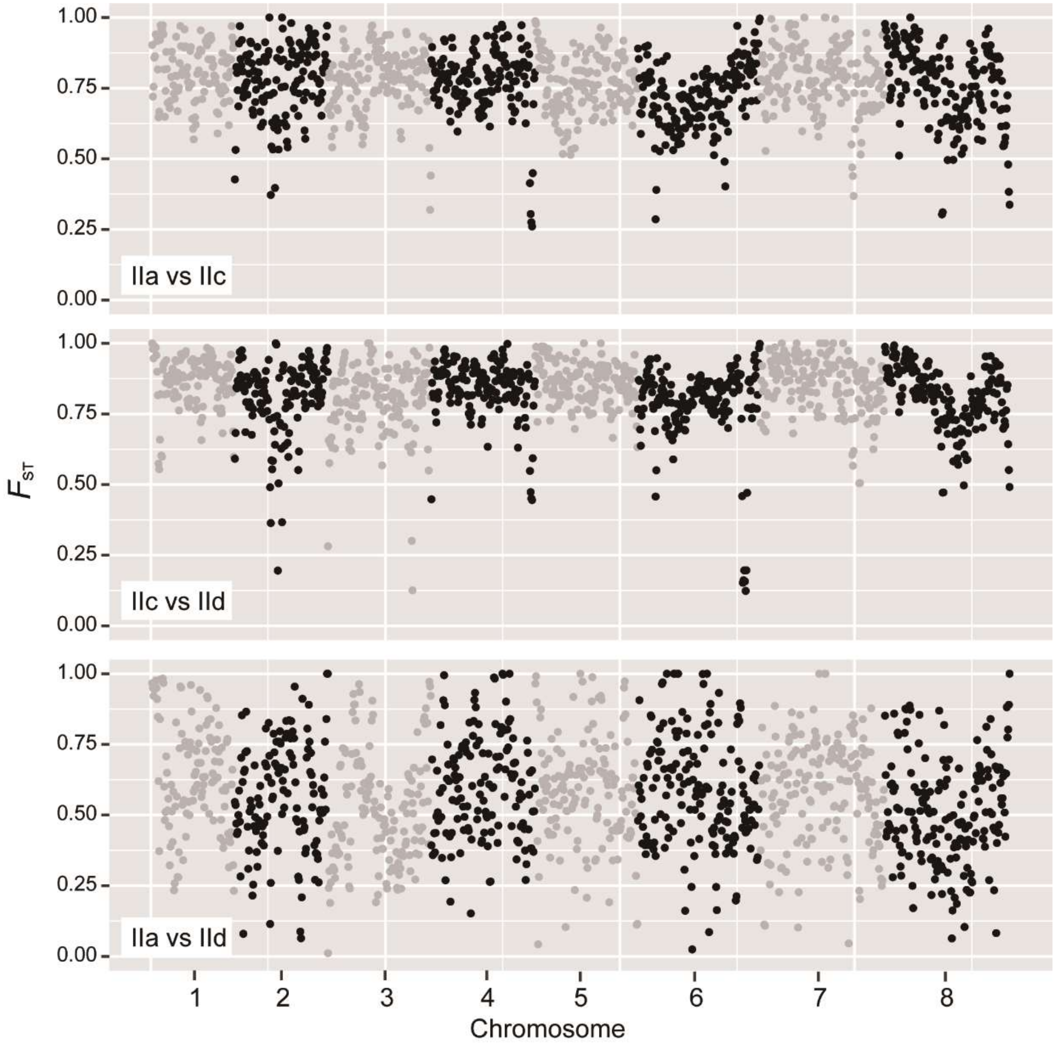
Supplementary fig. S2. Decay of genome-wide linkage disequilibrium (LD) by genetic distance among isolates from four major sources, including China (IId only), Egypt (IIa only), UK (IIa only) and USA (IIa only). An equal number of samples were randomly selected from each group for LD calculation to reduce the potential influence of the sample size.



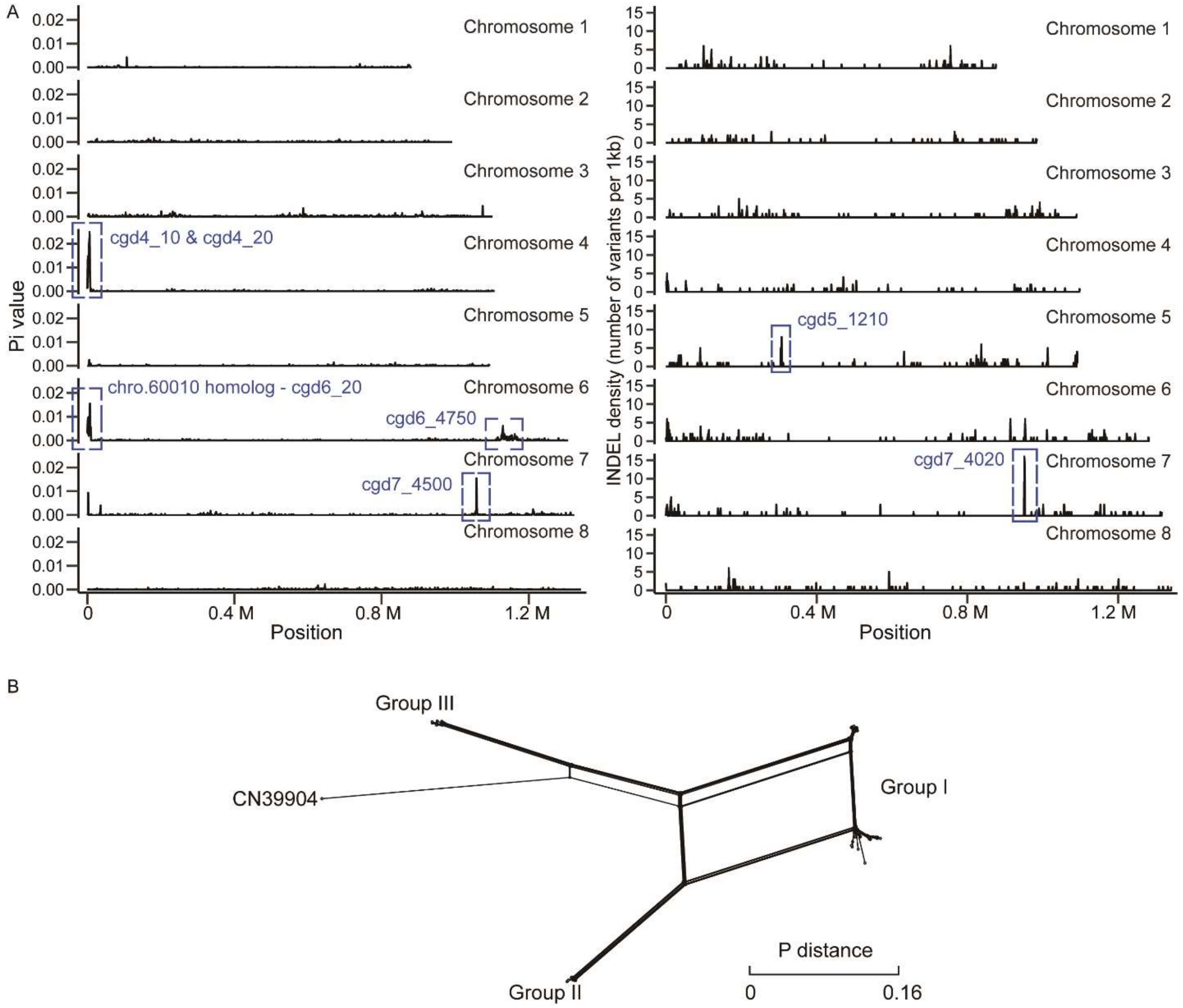
Supplementary fig. S3. Rooted whole genomic SNPs-based maximum likelihood tree of *Cryptosporidium parvum*. A total of 241,392 SNPs were used to build the consensus alignment. *C. hominis* was used as the outgroup (OG). The branch of each *C. parvum* isolate is colored by its *gp60* subtype, including IIa (red), IIc (purple) and IId (blue). The name of each isolate is colored by its source. Histograms outside the tree is the population structure estimated at K = 7. Subpopulations are labeled with groups I to VII and IIc.



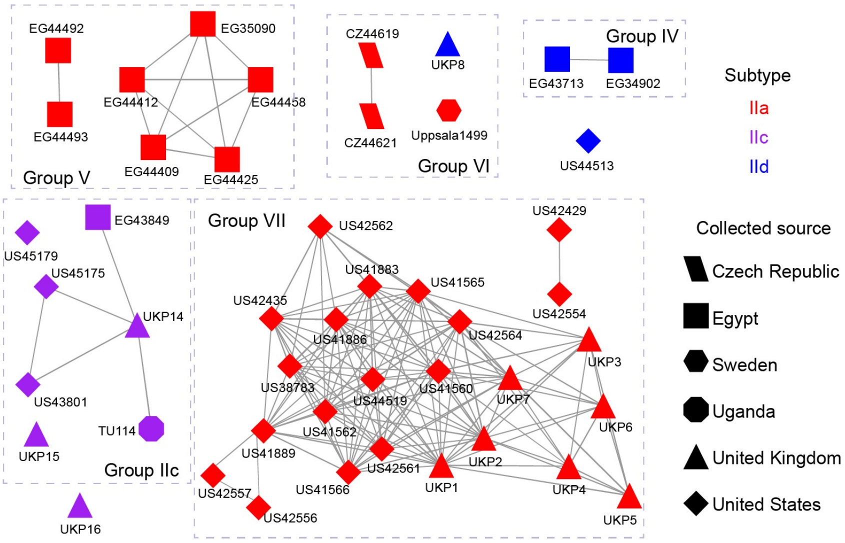
Supplementary fig. S4. Comparisons of population structures of *Cryptosporidium parvum* inferred with different K values. Population structures were inferred with K values of 3 to 9 using fastStructure. Each isolate is represented by a single bar in colored segments. The mosaic color bar above the histogram represents *gp60* subtypes and the one below the histogram represents the geographic origin of isolates.



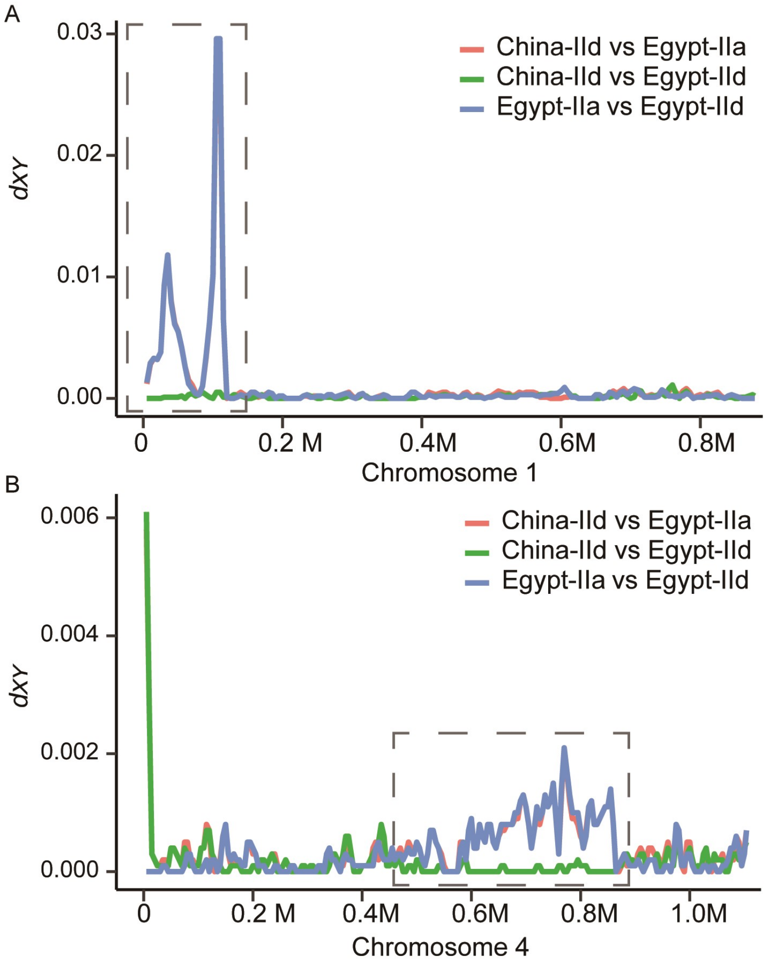
Supplementary fig. S5. Population differentiation between groups of *Cryptosporidium parvum*. Pair-wise *F*ST values were calculated along the eight chromosomes (colored with grey and black alternatively) for groups of *C. parvum*. Three representative groups were selected to represent the corresponding subtype family.



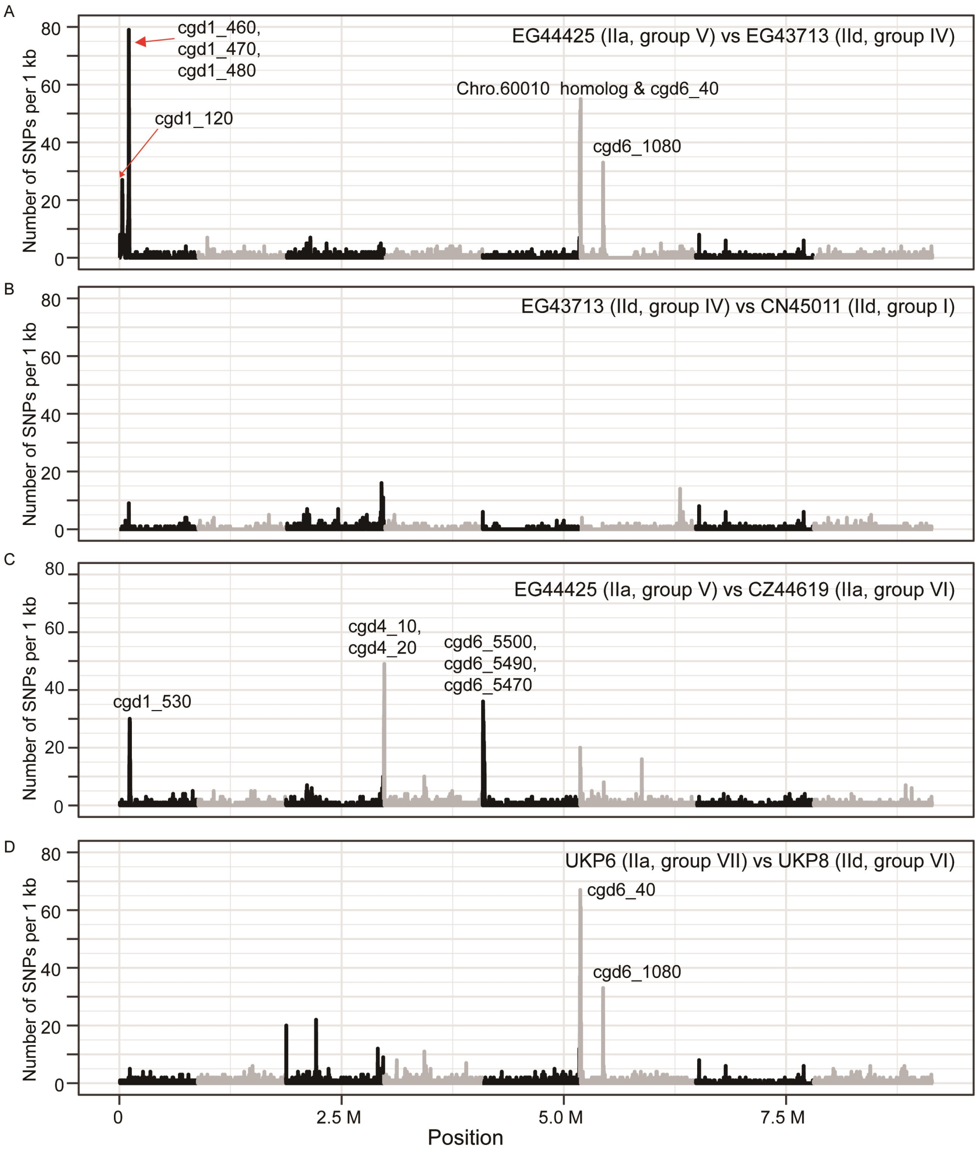
Supplementary fig. S6. Genome diversity of *Cryptosporidium parvum* in China. (***A***) Distribution of nucleotide diversity (Pi) and INDEL density along the eight chromosomes in the genome. Statistics were calculated with a 1-kb sliding window. Highly polymorphic genes are labeled. (***B***) Phylogenetic network of IId isolates in China. Concatenated SNPs were used to build the network with SplitsTree. The group labels are the same as in the neighbor-joining tree of all isolates (fig. 3*A*). Among the isolates, group I includes those from Shanghai, Guangdong, Henan and Jiangsu, group II includes those from Heilongjiang and Inner Mongolia, group III includes those from Hebei, while CN39904 is from Xinjiang.



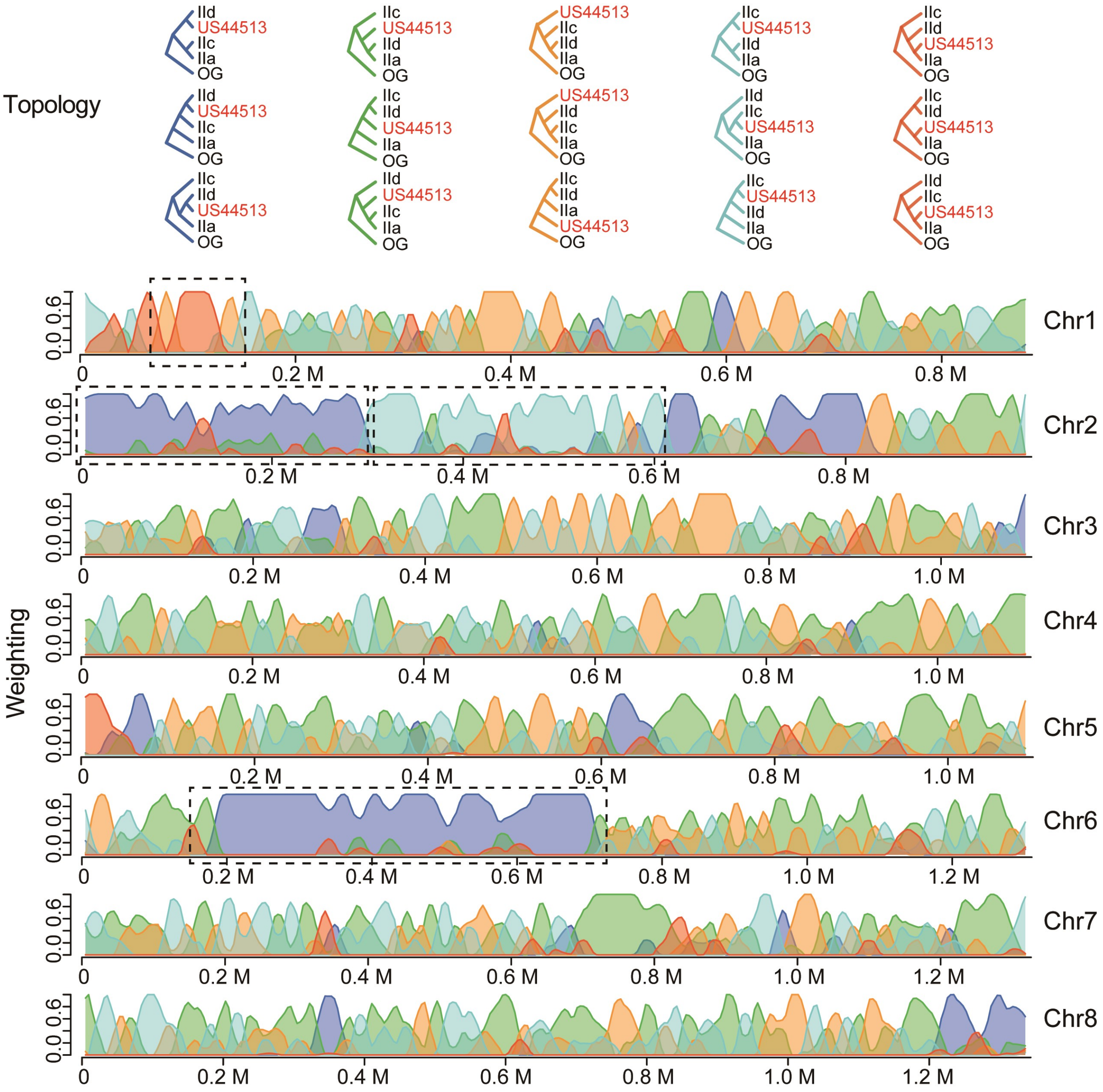
Supplementary fig. S7. Relatedness network for pairs of *Cryptosporidium parvum* isolates from countries other than China based on Identity By Descent (IBD) sharing. The remaining 46 genomes were used in the construction of the network with 155 edges. The edge between two nodes is linked if the IBD sharing fraction is over 90% between the genomes, indicating high genetic homology.



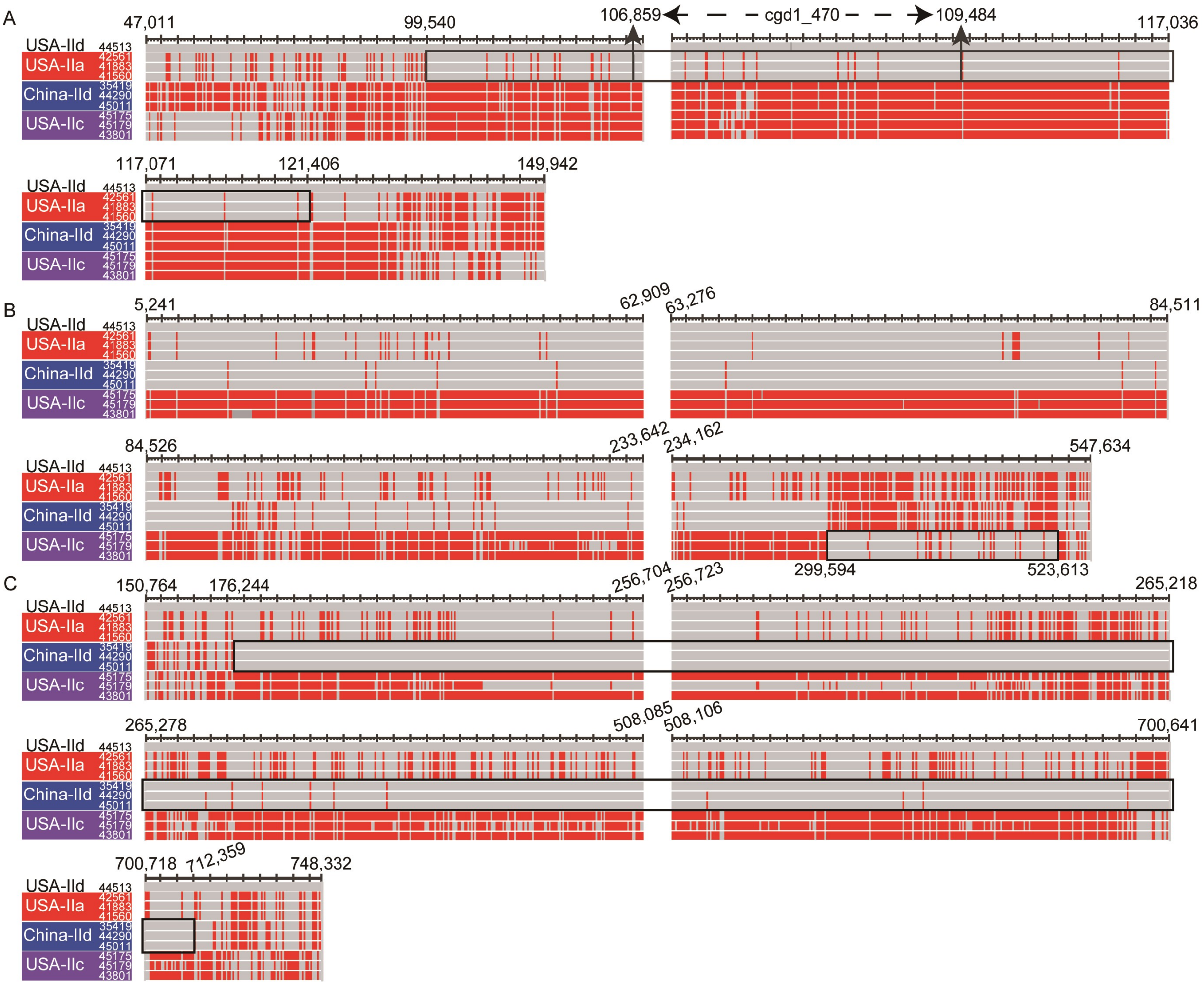
Supplementary fig. S8. Sequence divergence in two regions in chromosomes 1 and 4 among three subpopulations of *Cryptosporidium parvum*. (***A***) The IId genomes from Egypt and China are virtually identical to each other in chromosome 1 and differ from IIa genomes in Egypt in the 5′ subtelomeric region. High sequence differences between IIa and IId in Egypt are also present around cgd1_470, which encodes mucin 8. (***B***) Sequence similarity in a large region (600 k-870 k) in chromosome 4 between IId genomes from Egypt and China, with substantial sequence divergence from IIa genomes in Egypt. A 10-kb sliding window with 5-kb steps was used in the calculation of genetic divergence (d_{XY}).



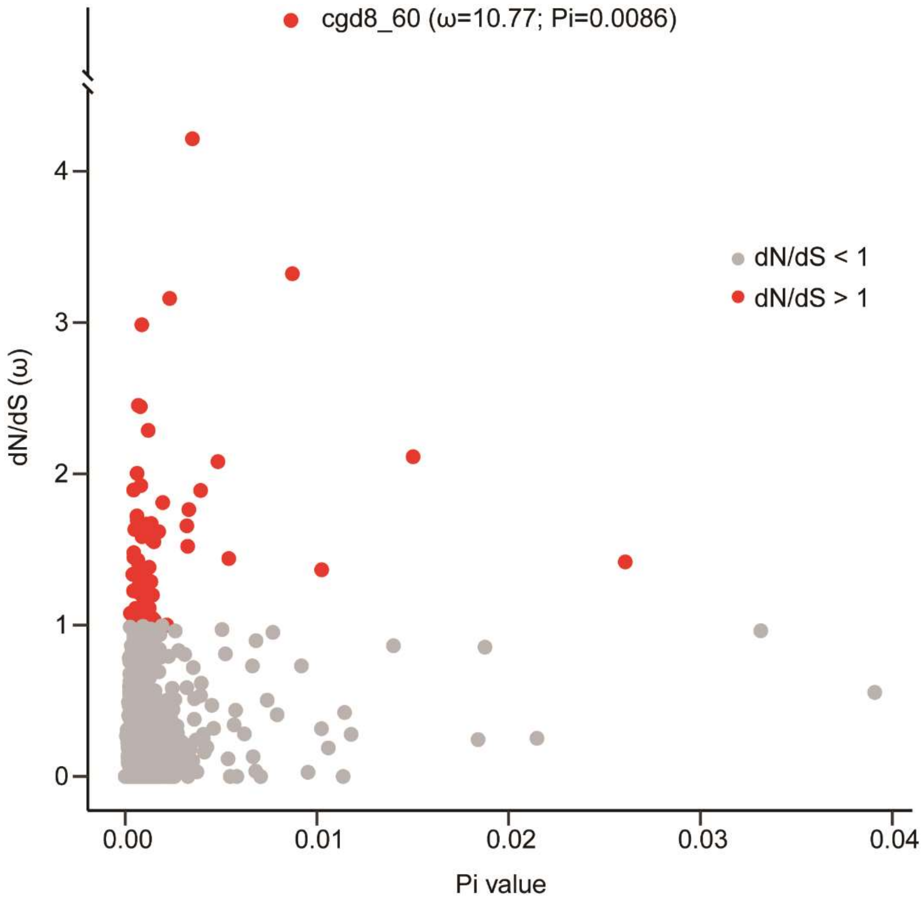
Supplementary fig. S9. Distribution of SNPs along the concatenated genome of *Cryptosporidium parvum* between IId and IIa genomes of different groups. Representative genomes from the different groups were used in the variant analysis, including (***A***) IIa genome from group V and IId genome from IV, (***B***) IId genomes from groups I and IV, (***C***) IIa genomes from groups V and VI, (***D***) IIa genome from group VII and IId genome from VI. SNPs between each pair were extracted using in-house python scripts from the overall dataset. The graphs were drawn with a 1-kb sliding window and the highly polymorphic genes (number of SNVs > 20/1,000 nucleotides) between the genomes under comparison are labelled with gene IDs.



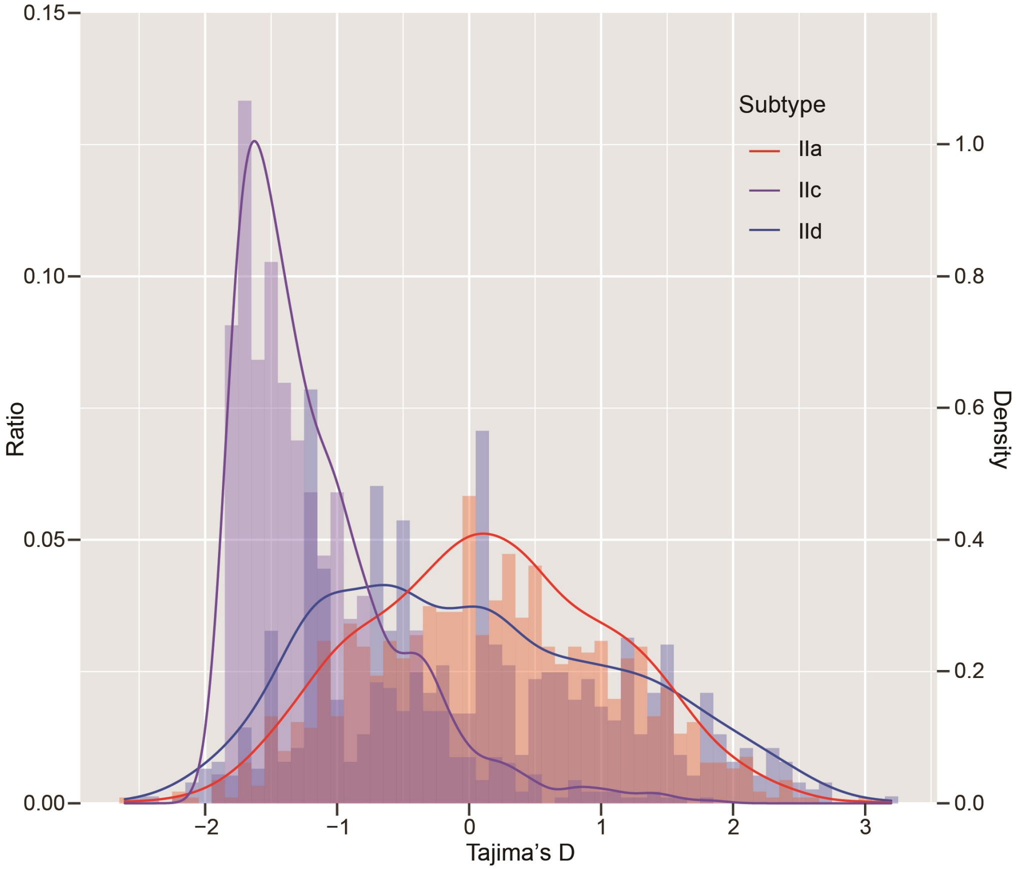
Supplementary fig. S10. Topology weighting along the genome showing sequence characteristics in the genome of US44513. A total of 15 possible topologies were grouped into five types according to the phylogenetic position of US44513 and were colored in different colors as shown in the upper panel. The bottom panel shows the proportion of each type of topology, which is shaded with colors corresponding to those in the upper panel and also in fig. 5*C*. A sliding window of 10 kb was used for calculation.



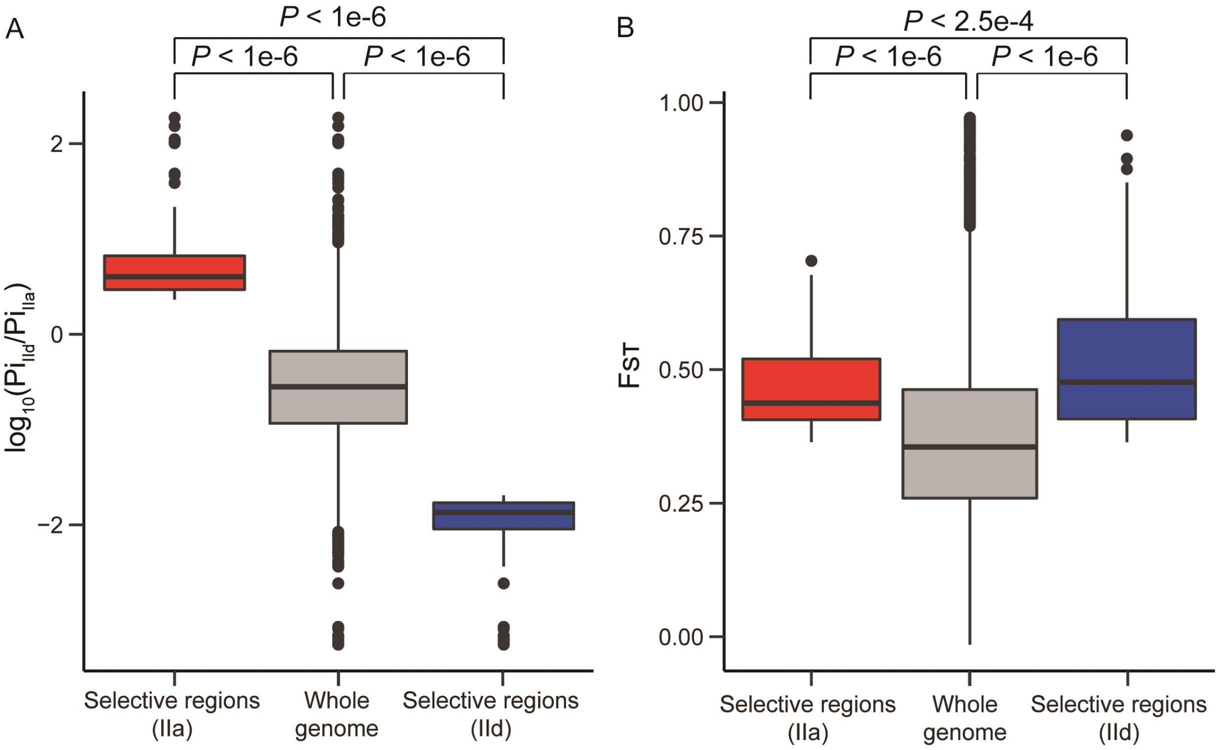
Supplementary fig. S11. Concatenated SNP alignment of US44513 and three *gp60* subtype families of *Cryptosporidium parvum*. (***A***) Concatenated SNP alignment in one area of chromosome 1 (99 k-120 k). Three genomes of each subtype family were selected from the comparison. The sequence of US44513 is placed at the top as the reference. Grey indicates sequence identity to US44513 while red indicates nucleotide substitution. A region marked with black box indicates high sequence identity between US4513 and IIa genomes, covering the gene encoded mucin glycoproteins protein (cgd1_470). (***B***) Concatenated SNP alignment in the first 550 kb of chromosome 2. In most of the alignment, US44513 shows high identity to IId. The black box at the end indicates high sequence identity between US44513 and IIc genomes. (***C***) Concatenated SNP alignment in one area of chromosome 6 (15 k-750 k). The black box indicates high sequence identity between US44513 and other IId genomes.



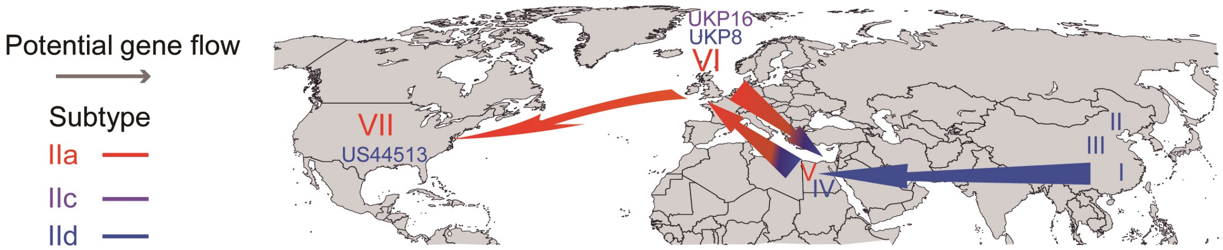
Supplementary fig. S12. Selection pressure and nucleotide diversity of genes in *Cryptosporidium parvum*. Each dot represents one gene. Genes extracted from 11 representative genomes were used in the analysis, with gene-based dN/dS (ω) and Pi values being calculated. Genes with dN/dS over 1 (in red) are thought under positive selection. Genes with dN/dS between 0 and 1 (in grey) are thought under purifying selection.



Supplementary fig. S13. Comparison of Tajima's D distribution among three major subtype families of *Cryptosporidium parvum*. The frequency distribution of Tajima's D values for IIc shows more negatively skewed values compared with the other two subtype families ($P <$ 1e-6 by Kruskal-Wallis test). Among them, the distribution of values in IIa is normally distributed ($P = 0.58$ by Kolmogorov-Smirnov test). The values were calculated using VCFtools with a non-overlapping 10-kb window. For the IId subtype family, UKP8 and US44513 were removed from the calculation to avoid biases due to the presence of considerable numbers of rare mutations in them.



Supplementary fig. S14. Box plot of Pi ratios (IId/IIa) (*A*) and FST values (*B*) in regions of IIa and IId genomes that are under selective sweep in comparison with the full genome. The statistical significance was assessed by using the Mann-Whitney U test.



Supplementary fig. S15. Gene flows in *Cryptosporidium parvum*. Potential gene flow events are marked with arrowed lines colored by *gp60* subtype family to indicate the ancestral sources.

Table S1. Information of 112 *Cryptosporidium* spp. isolates used in this study and summary statistics of whole genome sequencing (WGS) data

Bioproject	Sample name	Subtype	Depth (×)	Mapping Coverage	Number of contigs	Assembly length(bp)	N50(bp)	Region	Host	SRA run	Species	Group
This study	CN42632[*,§,¶]	IIdA19G1	184.09	0.9961	45	9128551	524034	China-Shanghai	Cattle	SRR15694545	*C. parvum*	I
This study	CN42658	IIdA19G1	332.66	0.9962	60	9124876	490577	China-Shanghai	Cattle	SRR15694544	*C. parvum*	I
This study	CN42669	IIdA19G1	344.59	0.9961	59	9124522	524979	China-Shanghai	Cattle	SRR15694543	*C. parvum*	I
This study	CN42672	IIdA19G1	310.71	0.9958	70	9127957	468340	China-Shanghai	Cattle	SRR15694542	*C. parvum*	I
This study	CN42962	IIdA19G1	110.45	0.9958	84	9182750	281029	China-Shanghai	Cattle	SRR15694541	*C. parvum*	I
This study	CN42963	IIdA19G1	313.01	0.9961	61	9128705	404983	China-Shanghai	Cattle	SRR15694540	*C. parvum*	I
This study	CN42964	IIdA19G1	300.38	0.9962	40	9126269	777450	China-Shanghai	Cattle	SRR15694538	*C. parvum*	I
This study	CN42965	IIdA19G1	226.89	0.9961	53	9129166	598907	China-Shanghai	Cattle	SRR15694537	*C. parvum*	I
This study	CN42966	IIdA19G1	243.97	0.9959	48	9120749	488810	China-Shanghai	Cattle	SRR15694536	*C. parvum*	I
This study	CN42967	IIdA19G1	314.57	0.9962	56	9124555	503195	China-Shanghai	Cattle	SRR15694535	*C. parvum*	I
This study	CN42970	IIdA19G1	265.78	0.9962	54	9124585	503194	China-Shanghai	Cattle	SRR15694534	*C. parvum*	I
This study	CN42972	IIdA19G1	285.83	0.9961	66	9125347	416655	China-Shanghai	Cattle	SRR15694533	*C. parvum*	I
This study	CN42973[*]	IIdA19G1	275.52	0.9961	48	9126731	579008	China-Shanghai	Cattle	SRR15694532	*C. parvum*	I
This study	CN42978	IIdA19G1	264.99	0.9959	62	9147499	410818	China-Shanghai	Cattle	SRR15694531	*C. parvum*	I
This study	CN42980	IIdA19G1	268.09	0.9962	56	9128775	502860	China-Shanghai	Cattle	SRR15694530	*C. parvum*	I
This study	CN42981	IIdA19G1	254.63	0.9963	44	9126983	578875	China-Shanghai	Cattle	SRR15694529	*C. parvum*	I
This study	CN42982	IIdA19G1	263.15	0.9962	58	9121770	450740	China-Shanghai	Cattle	SRR15694527	*C. parvum*	I
This study	CN42983	IIdA19G1	310.76	0.9963	46	9126704	502859	China-Shanghai	Cattle	SRR15694526	*C. parvum*	I
This study	CN42984	IIdA19G1	213.56	0.9960	61	9146565	509977	China-Shanghai	Cattle	SRR15694525	*C. parvum*	I
This study	CN42985	IIdA19G1	301.12	0.9960	46	9126927	510468	China-Shanghai	Cattle	SRR15694524	*C. parvum*	I
This study	CN42986	IIdA19G1	241.38	0.9966	50	9128227	501852	China-Shanghai	Cattle	SRR15694523	*C. parvum*	I
This study	CN42987	IIdA19G1	263.53	0.9961	58	9128378	345407	China-Shanghai	Cattle	SRR15694522	*C. parvum*	I
This study	CN45011[*,†,§,¶]	IIdA19G1	212.08	0.9958	52	9123057	526044	China-Guangdong	Cattle	SRR15694573	*C. parvum*	I
This study	CN45012	IIdA19G1	207.84	0.9959	59	9125623	489222	China-Guangdong	Cattle	SRR15694572	*C. parvum*	I
This study	CN45016	IIdA19G1	179.90	0.9957	55	9113046	389601	China-Guangdong	Cattle	SRR15694561	*C. parvum*	I
This study	CN45017	IIdA19G1	189.92	0.9957	81	9114623	377563	China-Guangdong	Cattle	SRR15694550	*C. parvum*	I
This study	CN45022	IIdA19G1	196.94	0.9958	77	9119104	255912	China-Guangdong	Cattle	SRR15694539	*C. parvum*	I
This study	CN6294	IIdA19G1	194.55	0.9939	2519	10283072	151175	China-Guangdong	Cattle	SRR15694528	*C. parvum*	I
This study	CN22958	IIdA19G1	88.22	0.9935	3174	10134354	155871	China-Guangdong	Cattle	SRR15694517	*C. parvum*	I
This study	CN22971	IIdA19G1	91.87	0.9936	642	9297791	165889	China-Guangdong	Cattle	SRR15694506	*C. parvum*	I
This study	CN22972	IIdA19G1	97.88	0.9937	766	9371599	154288	China-Guangdong	Cattle	SRR15694495	*C. parvum*	I
This study	CN22973	IIdA19G1	117.79	0.9944	713	9448049	229861	China-Guangdong	Cattle	SRR15694484	*C. parvum*	I
PRJNA320419	CN31727[*]	IIdA19G1	131.15	0.9941	195	9102487	115541	China-Henan	Cattle	SRR3473975	*C. parvum*	I

This study	CN35419[*,§,¶]	IIdA19G1	572.25	0.9953	85	9121358	277190	China-Henan	GKO mice model	SRR15694552	*C. parvum*	I
This study	CNXJ-CP	IIdA19G1	104.76	0.9937	3023	10568153	166037	China-Henan	Cattle	SRR15694551	*C. parvum*	I
This study	CN44290[*,§,¶]	IIdA19G1	305.63	0.9963	53	9125710	735705	China-Jiangsu	Cattle	SRR15694548	*C. parvum*	I
This study	CN44291	IIdA19G1	298.87	0.9962	55	9134005	410973	China-Jiangsu	Cattle	SRR15694547	*C. parvum*	I
This study	CN44298	IIdA19G1	287.28	0.9948	248	9115440	101649	China-Jiangsu	Cattle	SRR15694546	*C. parvum*	I
This study	CN11707[*]	IIdA20G1	182.82	0.9941	543	9250306	147454	China-Heilongjiang	Cattle	SRR15694564	*C. parvum*	II
This study	CN11722	IIdA19G1	118.68	0.9942	1512	9890051	159500	China-Heilongjiang	Cattle	SRR15694562	*C. parvum*	II
This study	CN11730	IIdA20G1	88.08	0.9942	1343	9647162	164303	China-Heilongjiang	Cattle	SRR15694560	*C. parvum*	II
This study	CN11738	IIdA20G1	86.87	0.9941	1278	9634356	173609	China-Heilongjiang	Cattle	SRR15694559	*C. parvum*	II
This study	CN11744	IIdA20G1	100.67	0.9944	5064	11796809	157673	China-Heilongjiang	Cattle	SRR15694558	*C. parvum*	II
This study	CN11760[*,†,§,¶]	IIdA20G1	158.06	0.9943	393	9190902	154551	China-Heilongjiang	Cattle	SRR15694563	*C. parvum*	II
This study	CN11783	IIdA20G1	82.01	0.9942	1375	9673954	158386	China-Heilongjiang	Cattle	SRR15694557	*C. parvum*	II
This study	CN11785	IIdA20G1	85.58	0.9943	668	9371699	157408	China-Heilongjiang	Cattle	SRR15694556	*C. parvum*	II
This study	CN9592[*,§,¶]	IIdA20G1	170.33	0.9943	484	9238425	138937	China-Inner Mongolia	Bactrian Camel	SRR15694549	*C. parvum*	II
This study	CN11789	IIdA20G1	98.14	0.9943	648	9333149	157437	China-Heilongjiang	Cattle	SRR15694555	*C. parvum*	II
This study	CN12405[*]	IIdA20G1	80.09	0.9928	782	9331055	148802	China-Hebei	Cattle	SRR15694570	*C. parvum*	III
This study	CN12535	IIdA20G1	80.63	0.9931	823	9345137	134110	China-Hebei	Cattle	SRR15694569	*C. parvum*	III
This study	CN12536	IIdA20G1	79.51	0.9937	653	9279294	143163	China-Hebei	Cattle	SRR15694568	*C. parvum*	III
This study	CN24696	IIdA20G1	110.75	0.9943	8583	13864095	134116	China-Hebei	Cattle	SRR15694567	*C. parvum*	III
This study	CN24716[*,†,§,¶]	IIdA20G1	88.67	0.9940	731	9390991	154382	China-Hebei	Cattle	SRR15694566	*C. parvum*	III
This study	CN24721	IIdA20G1	94.57	0.9944	1077	9534436	171959	China-Hebei	Cattle	SRR15694565	*C. parvum*	III
This study	CN39904[*,†,§,¶]	IIdA15G1	12.00	0.9787	1077	8807820	19195	China-XinJiang	Cattle	SRR15694520	*C. parvum*	-
This study	EG34902[*,§]	IIdA20G1	104.45	0.9956	104	9202795	212062	Egypt	Cattle	SRR15694516	*C. parvum*	IV
This study	EG43713[*,†,§]	IIdA20G1	132.80	0.9956	61	9125628	336482	Egypt	Human	SRR15694514	*C. parvum*	IV
This study	EG35090[§]	IIaA15G1R1	17.42	0.9907	656	9056186	39405	Egypt	Cattle	SRR15694515	*C. parvum*	V
This study	EG44409[*,§]	IIaA15G1R1	152.80	0.9963	126	9131737	263713	Egypt	Cattle	SRR15694512	*C. parvum*	V
This study	EG44412[*,§]	IIaA15G1R1	200.20	0.9959	129	9130189	245366	Egypt	Cattle	SRR15694511	*C. parvum*	V
This study	EG44425[*,†,§]	IIaA15G1R1	206.81	0.9962	109	9129248	265428	Egypt	Cattle	SRR15694510	*C. parvum*	V
This study	EG44458[*,§]	IIaA15G1R1	211.31	0.9934	492	9124046	50680	Egypt	Cattle	SRR15694509	*C. parvum*	V
This study	EG44492[*,§]	IIaA15G1R1	230.54	0.9958	137	9160198	216140	Egypt	Cattle	SRR15694508	*C. parvum*	V
This study	EG44493[§]	IIaA15G1R1	140.67	0.9943	2131	9490644	11037	Egypt	Cattle	SRR15694507	*C. parvum*	V

This study	CZ44619[*]	IIaA16G1R1	106.24	0.9962	57	9115842	327909	Czech Republic	Cattle	SRR15694519	*C. parvum*	VI
This study	CZ44621[*,†]	IIaA16G1R1	88.46	0.9962	54	9116233	365301	Czech Republic	Human	SRR15694518	*C. parvum*	VI
PRJNA308172	Uppsala1499[*]	IIaA16G1R1	1105.23	0.9959	194	9124225	134156	Sweden	Cattle	SRR3091751-61	*C. parvum*	VI
PRJNA253848	UKP8[*]	IIdA22G1	240.58	0.9943	411	9074754	71069	UK-England	Human	SRR6148259	*C. parvum*	VI
This study	US44513[*,†,¶]	IIdA15G1	191.51	0.9964	45	9123389	653246	USA-Alabama	Human	SRR15694483	*C. parvum*	-
This study	US42561[*,†,¶]	IIaA15G2R1	539.98	0.9976	46	9118496	487688	USA-NewYork	Cattle	SRR15694505	*C. parvum*	VII
This study	US42429	IIaA15G2R1	365.84	0.9973	133	9116940	308087	USA-Washington	Cattle	SRR15694504	*C. parvum*	VII
This study	US42554	IIaA15G2R1	447.22	0.9973	53	9117393	365249	USA-Washington	Cattle	SRR15694503	*C. parvum*	VII
This study	US41886	IIaA16G3R1	64.62	0.9962	127	9088468	220027	USA-Wisconsin	Cattle	SRR15694502	*C. parvum*	VII
This study	US42562[¶]	IIaA14G3R2	388.75	0.9974	36	9118212	583742	USA-Colorado	Cattle	SRR15694500	*C. parvum*	VII
This study	US38783	IIaA15G2R1	256.72	0.9962	81	9098756	360396	USA-Iowa	Cattle	SRR15694499	*C. parvum*	VII
This study	US41883[*,¶]	IIaA15G2R2	76.85	0.9966	68	9103353	340033	USA-Iowa	Cattle	SRR15694498	*C. parvum*	VII
This study	US41560[*,¶]	IIaA15G2R1	66.51	0.9966	74	9096277	412143	USA-Minnesota	Cattle	SRR15694497	*C. parvum*	VII
This study	US42435	IIaA14G1R2	504.27	0.9977	48	9120099	463248	USA-Minnesota	Cattle	SRR15694496	*C. parvum*	VII
This study	US41566	IIaA15G2R1	101.52	0.9970	55	9106202	485811	USA-NewYork	Cattle	SRR15694493	*C. parvum*	VII
This study	US41889	IIaA15G2R1	87.37	0.9968	62	9100857	461240	USA-NewYork	Cattle	SRR15694492	*C. parvum*	VII
This study	US42564[¶]	IIaA16G3R1	474.92	0.9975	41	9120664	591984	USA-Ohio	Cattle	SRR15694491	*C. parvum*	VII
This study	US41565[*,¶]	IIaA15G2R1	146.20	0.9970	53	9102599	563104	USA-Pennsylvania	Cattle	SRR15694489	*C. parvum*	VII
This study	US44519[¶]	IIaA15G2R1	314.54	0.9973	67	9118026	333344	USA-Tennessee	Cattle	SRR15694488	*C. parvum*	VII
This study	US42556[¶]	IIaA17G2R1	478.05	0.9975	50	9114216	503360	USA-Washington	Cattle	SRR15694487	*C. parvum*	VII
This study	US42557	IIaA17G2R1	465.50	0.9973	58	9114383	424976	USA-Washington	Cattle	SRR15694486	*C. parvum*	VII
This study	US41562[¶]	IIaA17G2R1	95.38	0.9968	70	9095786	333020	USA-Wisconsin	Cattle	SRR15694485	*C. parvum*	VII
PRJNA439211	UKP1[*]	IIaA17G1R1	1004.69	0.9972	55	9211997	429156	UK	Human	SRR6871415	*C. parvum*	VII
PRJNA253836	UKP2[*]	IIaA19G1R2	43.52	0.9958	147	9098259	174184	UK-England	Human	SRR6117460	*C. parvum*	VII
PRJNA253840	UKP3[*]	IIaA18G2R1	113.06	0.9952	301	9081779	107032	UK-Wales	Human	SRR6147472	*C. parvum*	VII
PRJNA253843	UKP4	IIaA15G2R1	224.63	0.9954	285	9080095	95982	UK-England	Human	SRR6147581	*C. parvum*	VII
PRJNA253845	UKP5	IIaA15G2R1	19.10	0.9945	560	9020719	48689	UK	Human	SRR6147587	*C. parvum*	VII
PRJNA253846	UKP6[*]	IIaA15G2R1	80.52	0.9963	112	9109773	200532	UK	Human	SRR6147945	*C. parvum*	VII
PRJNA253847	UKP7[*]	IIaA17G1R1	98.39	0.9922	593	8990035	47383	UK-England	Human	SRR6147964	*C. parvum*	VII
This study	US45179[*,¶]	IIcA5G3d	259.38	0.9954	45	9102376	593612	USA-Nebraska	Human	SRR15694494	*C. parvum*	IIc
This study	US43801[*,¶]	IIcA5G3a	320.80	0.9953	80	9127151	341506	USA-Oregon	Human	SRR15694490	*C. parvum*	IIc
This study	US45175[*,†,¶]	IIcA5G3a	255.96	0.9956	47	9104505	410630	USA-Alabama	Human	SRR15694501	*C. parvum*	IIc
This study	EG43849[*]	IIcA5G3a	109.73	0.9943	61	9109671	387299	Egypt	Human	SRR15694513	*C. parvum*	IIc
PRJNA437480	TU114[*]	IIcA5G3b	195.30	0.9946	47	9109832	406808	Uganda	Human	SRR10996883	*C. parvum*	IIc

PRJNA315506	UKP14	IIcA5G3a	60.83	0.9625	2276	8316505	6565	UK-England	Human	SRR6813718	*C. parvum*	IIc
PRJNA315507	UKP15	IIcA5G3a	101.79	0.9896	994	8978913	23227	UK-Wales	Human	SRR6813719	*C. parvum*	IIc
PRJNA315508	UKP16[*,†]	IIcA5G3j	117.30	0.9910	914	8986137	24053	UK-Wales	Human	SRR7898459	*C. parvum*	-
This study	CN11747[‡]	IIdA19G1	1.17	0.4438	-	-	-	China-Heilongjiang	Cattle	SRR15694553	*C. parvum*	-
This study	CN11768[‡]	IIdA20G1	15.61	0.6505	-	-	-	China-Heilongjiang	Cattle	SRR15694554	*C. parvum*	-
This study	CN45023[‡]	IIdA19G1	38.61	0.9521	-	-	-	China-Guangdong	Addax	SRR15694571	*C. parvum*	-
This study	CN42649[‡]	IIdA19G1	52.11	0.7297	-	-	-	China-Shanghai	Cattle	SRR15694521	*C. parvum*	-
PRJNA315504	UKP12[‡]	IIcA5G3p	29.37	0.9289	-	-	-	UK-England	Human	SRR6813717	*C. parvum*	-
PRJNA315505	UKP13[‡]	IIcA5G3a	51.67	0.9023	-	-	-	UK-England	Human	SRR6813716	*C. parvum*	-
PRJNA222836	TU502[§,¶]	IbA10G2	91.29	0.9895	27	9088522	879981	Uganda	Human	SRR1015747	*C. hominis*	OG
PRJNA222837	UKH1[§,¶]	IbA10G2	113.06	0.9898	35	9094195	1076561	UK-Wales	Human	SRR1015721	*C. hominis*	OG
PRJNA252787	US30976[§,¶]	IaA28R4	446.06	0.9898	89	9058648	172076	USA-Missouri	Human	SRR1557959	*C. hominis*	OG
PRJNA253838	UKH4[§,¶]	IaA14R3	257.25	0.9874	393	9017179	72253	UK	Human	SRR6143718	*C. hominis*	OG
PRJNA634705	Afr1[§,¶]	IeA11G3T3	196.36	0.9907	275	9031482	95430	Ghana	Human	SRR11843813	*C. hominis*	OG

Note:

[*] Genomes selected for ANI analysis.

[†] Genomes selected for calculating selection pressure.

[‡] Not used for following analysis except mapping and assembly.

[§] Genomes used for topology weighting of Egypt-IId.

[¶] Genomes used for topology weighting of US44513.

Table S2. Summary of related pairs identified with Identity By Descent (IBD) within the same subtype family

Subtype	No. samples	No. all pairs	Percentage of pairs inferred IBD (%)
IIa	34	561	26.6
IIc	8	28	17.9
IId	59	1711	44.7

Table S3. Summary of related pairs identified with Identity By Descent (IBD) within the same country (samples over 3)

Country	No. samples	No. all pairs	Percentage of pairs inferred IBD (%)
China	55	1485	51.4
Egypt	10	45	26.7
UK	11	55	38.2
USA	21	210	37.6

Table S4. Gene gains and losses in genomes of *Cryptosporidium parvum*

Chr	Gene family	*C. parvum* IIa-IOWA	*C. parvum* IIa-CZ44621	*C. parvum* IIa-EG44425	*C. parvum* IId-EG43713	*C. parvum* IId-CN35419	*C. parvum* IId-UKP8	*C. parvum* IId-US44513	*C. parvum* IIc-US45179	*C. parvum* IIc-UKP16
1	dUTP diphosphatase	cgd1_3860	cgd1_3860	cgd1_3860	cgd1_3860	cgd1_3860	cgd1_3860	cgd1_3860	-	cgd1_3860
3	SKSR*	-	cgd3_10 paralog	cgd3_10 paralog	cgd3_10 paralog	cgd3_10 paralog	cgd3_10 paralog	cgd3_10 paralog	-	cgd3_10 paralog
5	Insulinase-like proteases	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	-	-	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog
		-	-	cgd3_4260 paralog	cgd3_4260 paralog	cgd3_4260 paralog	-	-	-	-
	MEDLE secreted proteins	cgd6_5480	cgd6_5480	-	-	-	-	cgd6_5480	cgd6_5480	-
		cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490	cgd6_5490
	Hypothetical protein	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470	cgd6_5470
		cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500
6	Insulinase-like proteases	cgd6_5510_5520	cgd6_5510_5520	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	cgd6_5510_5520 paralog	-	-
	Hypothetical protein	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	cgd6_5500	-	cgd6_5500
7	MEDLE secreted proteins	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580	cgd5_4580
		cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590	cgd5_4590
		cgd5_4600	cgd5_4600	cgd5_4600 paralog	cgd5_4600	cgd5_4600	cgd5_4600	cgd5_4600	cgd5_4600	-
		-	cgd5_4600 paralog	-	-	-	-	cgd5_4610 paralog	-	-
		cgd5_4610	cgd5_4610	cgd5_4610 paralog	cgd5_4610	cgd5_4610	cgd5_4610	cgd5_4610	cgd5_4610	cgd5_4610
Subtotal	MEDLE secreted proteins	6	7	5	5	5	5	7	6	4
	Insulinase-like proteases	2	2	3	3	3	1	1	1	1

Table S5. Information of genes under positive selection among 11 selected *Cryptosporidium parvum* genomes

Chr	Gene name	dN/dS	# SNPs	Pi value	Subcellular location prediction	Signal peptide prediction	Product
1	cgd1_120	2.11	49	0.01501	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd1_130	1.37	47	0.01024	Extracellular	Yes	Hypothetical protein
	cgd1_620	1.52	19	0.00326	Golgi apparatus	Yes	Hypothetical protein
	cgd1_1260	1.31	6	0.00097	Extracellular	Yes	Hypothetical protein
	cgd1_2810	1.04	6	0.0008	Nucleus	No	Hypothetical protein
	cgd1_3090	1.81	8	0.00196	Nucleus	No	Hypothetical protein
	cgd1_3120	1.11	6	0.00055	Cytoplasm	No	Peptide synthase-like condensation domain
	cgd1_3180	1.06	5	0.00098	Cytoplasm	No	Hypothetical protein
2	cgd2_450	3.16	39	0.00232	Extracellular	Yes	Mucin-like protein
	cgd2_410	1.89	18	0.00393	Extracellular	Yes	Mucin-like protein
	cgd2_400	2.29	4	0.0012	Extracellular	Yes	Mucin-like protein
	cgd2_940	2.45	13	0.00068	Cell membrane	Yes	Signal peptide plus transmembrane domain or GPI anchor
	cgd2_1380	1.03	5	0.00152	Mitochondrion	No	Hypothetical protein
	cgd2_1710	1.08	6	0.00027	Lysosome/Vacuole	No	Putative signal peptide protein and transmembrane domain containing protein
	cgd2_1720	1.19	6	0.00091	Cytoplasm	No	Hypothetical protein
	cgd2_3330	1.48	8	0.00044	Endoplasmic reticulum	Yes	APG-1-like HSP70 domain-containing protein
	cgd2_3460	1.07	5	0.00042	Nucleus	No	Zinc finger-containing and myb DNA binding domain-containing protein
	cgd2_4370	3.32	43	0.00872	Endoplasmic reticulum	No	Signal peptide-containing protein
3	cgd3_1580	1.89	9	0.00044	Cytoplasm	No	Hypothetical protein
	cgd3_1690	2.08	28	0.00483	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd3_1700	1.04	13	0.00143	Extracellular	No	Hypothetical protein
	cgd3_1710	4.21	15	0.0035	Extracellular	Yes	Hypothetical protein
	cgd3_1770	1.44	37	0.0054	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd3_1780	1.76	28	0.00332	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd3_3420	1.38	6	0.00125	Lysosome/Vacuole	No	Transmembrane domain and EF hand containing protein
	cgd3_3650	1.33	6	0.00087	Cytoplasm	No	Centrin-like EF hand protein
	cgd3_3700	1.23	6	0.00043	Endoplasmic reticulum	No	Hypothetical protein
4	cgd4_1110	1.07	13	0.00084	Cytoplasm	No	Hypothetical protein
	cgd4_1140	1.6	9	0.00115	Cytoplasm	No	Hypothetical protein
	cgd4_3880	1.62	9	0.00174	Cytoplasm	No	Emap RNA binding domain-containing protein
	cgd4_3720	1.08	5	0.00033	Cytoplasm	No	Putative TPR domain-containing protein
	cgd4_3670	1.66	15	0.00322	Nucleus	No	Hypothetical protein
	cgd4_3420	1.92	9	0.00081	Cytoplasm	No	Putative Vps53-N protein
	CPATCC_001824	1.41	6	0.00069	Cytoplasm	No	Hypothetical protein
	cgd4_1850	1.34	7	0.00067	Cytoplasm	No	RNAse Ccr4p
	cgd4_4300	1.72	7	0.00061	Cytoplasm	No	Beta adaptin

5	cgd5_5470	1.42	53	0.02608	Extracellular	No	Putative signal peptide-containing protein
	cgd5_20	2.44	11	0.00079	Extracellular	Yes	Signal peptide containing protein
	cgd5_50	1.67	15	0.00111	Nucleus	No	Signal peptide containing protein
	cgd5_1660	1.09	5	0.00049	Golgi apparatus	Yes	Putative signal peptide-containing protein and acidic stretch
	cgd5_2100	1.09	7	0.00079	Endoplasmic reticulum	No	DNA methyltransferase PMT1-like protein
	cgd5_2560	1.02	9	0.00059	Cytoplasm	No	MutL-like ATPase
	cgd5_4040	1.24	9	0.00058	Nucleus	No	Hypothetical protein
	cgd5_3820	1.29	6	0.00109	Endoplasmic reticulum	No	Vacuole membrane protein%2C VMP1 like integral membrane protein
	CPATCC_002492	1.55	7	0.0015	Extracellular	No	Hypothetical protein
6	cgd6_160	1.05	4	0.00049	Cytoplasm	No	Hypothetical protein
	cgd6_1040	1	8	0.00216	Nucleus	No	RNAse P Rpr2/Rpp21 subunit domain
	cgd6_1130	1.09	5	0.0005	Endoplasmic reticulum	Yes	Putative signal peptide-containing protein
	cgd6_2860	1.03	5	0.00079	Cytoplasm	No	Putative coiled coil protein
	cgd6_3160	2.98	14	0.00086	Cytoplasm	No	Hypothetical protein
	cgd6_3200	1.27	9	0.00105	Cytoplasm	No	Hypothetical protein
	cgd6_3600	1.03	5	0.00121	Nucleus	No	Hypothetical protein
	cgd6_4090	1.29	7	0.00134	Cytoplasm	No	Hypothetical protein
	cgd6_4640	1.58	10	0.00088	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd6_5320	2	8	0.00062	Cytoplasm	No	Hypothetical protein
7	cgd7_930	1.37	6	0.00077	Endoplasmic reticulum	No	Membrane domain-containing protein
	cgd7_1350	1.13	5	0.00087	Cell membrane	No	Signal peptide-containing protein
	cgd7_1370	1.11	4	0.00125	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd7_2230	1.34	6	0.0006	Cytoplasm	No	Coiled coil-containing protein
	cgd7_2600	1.69	15	0.00063	Cytoplasm	No	Hypothetical protein
	cgd7_2830	1.08	5	0.0005	Cytoplasm	No	Hypothetical protein
	cgd7_2880	1.11	5	0.00092	Cytoplasm	No	SLBP-like protein
	cgd7_4710	1.2	7	0.00142	Cytoplasm	No	Hypothetical protein
	cgd7_5470	1.45	4	0.00045	Cytoplasm	No	Hypothetical protein
8	cgd8_60	10.77	23	0.00864	Extracellular	Yes	Signal peptide-containing protein
	cgd8_380	1.67	9	0.00136	Mitochondrion	No	Putative oxidase or dehydrogenase
	cgd8_770	1.63	7	0.0005	Cytoplasm	No	Hypothetical protein
	cgd8_1620	1.28	6	0.00074	Nucleus	No	DNA polymerase delta2
	cgd8_2450	1.22	12	0.0007	Cytoplasm	No	Formin homology 2 domain (FH2) protein
	cgd8_4540	1.06	6	0.00039	Cytoplasm	No	Putative coiled coil-containing protein
	cgd8_4570	1.01	4	0.0007	Lysosome/Vacuole	No	Transmembrane domain-containing protein
	cgd8_4640	1.56	6	0.00141	Nucleus	No	WD40 repeat
	cgd8_5210	1.34	13	0.00039	Cytoplasm	No	Hypothetical protein
	cgd8_5300	1.43	13	0.00065	Endoplasmic reticulum	Yes	Signal peptide containing protein
	cgd8_5380	1.01	37	0.00122	Extracellular	Yes	Putative signal peptide-containing secreted protein

Table S6. Information of polymorphic genes under purifying selection among 11 selected *Cryptosporidium parvum* genomes

Chr	Gene name	# SNPs	dN/dS	Pi value	Subcellular location prediction	Signal peptide prediction	Product
1	cgd1_120	49	2.113	0.015	Endoplasmic reticulum	Yes	Hypothetical protein
	cgd1_130	44	1.366	0.010	Extracellular	Yes	Hypothetical protein
	cgd1_460	47	0.278	0.012	Cytoplasm	No	Hypothetical protein
	cgd1_470	204	0.963	0.033	Extracellular	Yes	Mucin-like protein
	cgd1_480	18	0.244	0.018	Mitochondrion	No	Hypothetical protein
	cgd1_530	7	0.000	0.011	Mitochondrion	No	Hypothetical protein
2	cgd2_420	32	-	0.010	Extracellular	Yes	Mucin-like protein
4	cgd4_10	20	-	0.025	Nucleus	No	Hypothetical protein
	cgd4_20	47	0.252	0.021	Endoplasmic reticulum	Yes	Putative signal peptide-containing protein
	cgd4_4500	242	-	0.022	Extracellular	Yes	Signal peptide containing cysteine rich protein
	CPATCC_001985	70	-	0.017	Mitochondrion	No	Hypothetical protein
5	cgd6_5500	46	-	0.027	Cytoplasm	No	Hypothetical protein
	cgd5_5490	16	-	0.010	Extracellular	Yes	Hypothetical protein
	cgd5_5470	53	1.418	0.026	Extracellular	No	Putative signal peptide-containing protein
6	chro.60010 homolog	52	-	0.015	Cytoplasm	No	Hypothetical protein
	cgd6_40	37	-	0.025	Extracellular	Yes	Hypothetical protein
	cgd6_60	251	0.423	0.011	Extracellular	No	Putative signal peptide-containing protein
	cgd6_1060	51	0.316	0.010	Nucleus	No	Spectrin repeat-containing protein
	cgd6_1070	22	0.189	0.011	Endoplasmic reticulum	No	Hypothetical protein
	cgd6_1080	94	0.556	0.039	Cell membrane	Yes	Hypothetical protein
	cgd6_5270	50	0.855	0.019	Extracellular	Yes	Putative signal peptide-containing protein
7	cgd7_4531	31	-	0.013	Endoplasmic reticulum	Yes	Signal peptide-containing protein
	cgd7_4330	137	-	0.014	Cell membrane	No	Putative glycoprotein
	cgd7_4500	104	0.864	0.014	Extracellular	Yes	Signal peptide-containing secreted protein
8	cgd8_10	107	-	0.024	Cytoplasm	No	Hypothetical protein
	cgd8_20	241	-	0.017	Golgi apparatus	Yes	Putative secreted signal peptide-containing protein
	cgd8_30	219	-	0.015	Endoplasmic reticulum	Yes	Putative secreted signal peptide-containing protein
	cgd8_40	134	-	0.012	Endoplasmic reticulum	Yes	Putative secreted signal peptide-containing protein

Table S7. List of genes with over 3 SNPs in zoonotic *Cryptosporidium parvum* (IIa and IId genomes) within regions under selective sweep in IIa

Chr	Gene ID	Number of SNPs			Product
		All genomes	IIa	IId	
2	cgd2_690	8	1	6	Lycine-rich repeat-containing protein
	cgd2_700	6	4	4	PMS1'MutL-like ATPase'
	cgd2_1970	4	2	2	Hypothetical protein
3	cgd3_2080	8	4	5	Transmembrane domain-containing protein
	cgd3_2120	4	2	2	Myosin fused to Q motif and Rcc1 domain
	cgd3_2180	16	5	13	Type I fatty acid synthase
	cgd3_2730	4	2	2	Transmembrane domain-containing protein
	cgd3_2960	5	3	2	Hypothetical protein
	cgd3_3060	4	1	3	Putative exportin 1
	cgd3_3560	4	0	4	Magnesium transporter CorA-like protein
	cgd3_3570	11	4	7	Similar to 80 kDa MCM3-associated protein
	cgd3_3600	4	0	4	Hypothetical protein
	cgd3_3620	12	3	10	Transmembrane domain-containing protein
4	cgd4_440	6	1	5	Flap endonuclease
6	cgd6_4700	9	3	6	ATP-binding cassette transporter
	cgd6_4730	4	3	1	Transmembrane domain-containing protein
	cgd6_4750	22	2	19	Splicing factor 3B subunit1-like HEAT repeat-containing protein
	cgd6_4760	5	0	1	Tubulin beta chain
	cgd6_4770	4	0	4	Similar to cell cycle control protein cwf2
	cgd6_4780	6	0	6	Putative transmembrane domain-containing protein
	cgd6_4790	11	1	10	Putative Zn finger-containing protein
	cgd6_4800	6	0	5	Hypothetical protein
	cgd6_4830	7	2	4	Drs1p
	cgd6_4840	11	6	7	Serine protease
	cgd6_4850	7	1	6	PRP19 non-snRNP sliceosome component required for DNA repair
	cgd6_4860	5	1	4	Nucleolar protein GU2
	cgd6_4910	6	1	5	CCCH RNA binding domains-containing protein
	cgd6_4920	4	0	4	Transmembrane domain-containing protein
	cgd6_4930	7	1	6	Beta-adaptin AP complex subunit-related protein
	cgd6_5190	4	1	3	Hypothetical protein
8	cgd8_3040	5	2	3	Putative transmembrane domain-containing protein

Table S8. List of genes with over 3 SNPs in zoonotic *Cryptosporidium parvum* (IIa and IId genomes) within regions under selective sweep in IId

Chr	Gene ID	Number of SNPs			Product
		All genomes	IIa	IId	
1	cgd1_470	90	1	0	Mucin-like protein
	cgd1_480	18	0	0	Hypothetical protein
	cgd1_490	7	0	0	Hypothetical protein
	cgd1_500	16	15	0	Hypothetical protein
	cgd1_510	24	24	0	Hypothetical protein
	cgd1_530	6	6	0	Hypothetical protein
	cgd1_540	5	5	0	Putative signal peptide-containing vacuolar ATP synthase subunit
	CMU_005380 homolog	10	10	0	Hypothetical protein
	cgd1_550	41	41	1	Putative signal peptide-containing secreted protein
	cgd1_1040	4	4	0	Hypothetical protein
	cgd1_1250	15	15	1	Putative signal peptide-containing secreted protein
	cgd1_1300	4	4	0	Hypothetical protein
	cgd1_1640	9	8	1	AT hook and a PHD finger domain-containing chromatin protein
	cgd1_3170	6	6	0	Hypothetical protein
2	cgd2_80	6	6	0	Putative ABC transporter-like protein
	cgd2_60	8	8	0	Hypothetical protein
	cgd2_2320	4	3	1	Ferlin-like protein
	cgd2_2360	4	3	0	Dop1p-like protein
	cgd2_2680	4	4	1	Putative magnesium-transporting ATPase
	cgd2_3090	5	5	0	Putative signal peptide protein
	cgd2_4250	4	4	0	Putative tuftelin-interacting protein
3	cgd3_950	4	4	0	Hypothetical protein
	cgd3_1780	5	5	0	Hypothetical protein
	cgd3_1800	4	4	0	Transmembrane domain-containing protein
	cgd3_4190	8	8	1	Secreted insulinase-like peptidase
	cgd3_4200	4	4	0	Signal peptide-containing secreted insulinase-like peptidase
	cgd4_510	8	8	1	Hypothetical protein
4	cgd4_550	4	3	1	Hypothetical protein
	cgd4_1350	4	4	0	Regulator of chromosome condensation
	cgd4_3630	15	15	0	Hypothetical protein
	cgd4_3610	4	4	0	Putative signal peptide-containing protein
	cgd4_3530	6	5	1	Putative signal peptide-containing protein
	cgd4_2900	17	13	2	Polyketide synthase
	cgd4_2850	4	4	0	Hypothetical protein
	cgd4_2690	9	9	0	Sec7 domain-containing protein
	cgd4_2670	17	16	0	FRP1 like and phosphatidylinositol kinase domain-containing protein
	cgd4_2470	9	8	0	Putative signal peptide-containing protein
	cgd4_2150	5	4	0	Hypothetical protein
5	cgd5_5460	16	15	1	Hypothetical protein
	cgd5_810	4	3	1	Leucine-rich repeat protein
	cgd5_2540	7	7	0	WD repeat protein
	cgd5_2580	6	5	1	Putative signal peptide-containing protein
6	cgd6_40	34	34	0	Hypothetical protein
	cgd6_50	22	22	0	Hypothetical protein
	cgd6_60	4	3	1	Putative signal peptide-containing protein
	cgd6_730	4	4	0	14-3-3 domain-containing protein
	cgd6_880	8	7	1	Glycogen debranching enzyme Gdb1p
	cgd6_2790	5	5	0	Putative kinesin heavy chain
	cgd6_3040	6	6	0	Hypothetical protein
	cgd6_3050	11	11	0	Putative signal peptide-containing protein
	cgd6_3690	4	4	0	Putative PUG and UBX domain-containing protein
	cgd6_3730	8	7	1	Signal peptide-containing protein

	cgd6_3760	4	4	0	Hypothetical protein
7	cgd7_1340	4	4	0	Hypothetical protein
	cgd7_1550	4	4	0	Hypothetical protein
	cgd7_1630	7	7	0	Putative RNA methylase
	cgd7_2340	6	6	0	Signal peptide-containing protein
	cgd7_4310	8	5	1	Signal peptide-containing protein
	cgd7_4600	4	3	0	Abstrakt protein SF II helicase
	cgd7_4990	13	13	1	Ubiquitin ligase
8	cgd8_550	4	4	0	Hypothetical protein
	cgd8_810	10	9	1	Hypothetical protein
	cgd8_850	6	6	0	Putative membrane associated protein
	cgd8_870	5	5	0	DNA dependent DNA polymerase alpha subunit
	cgd8_900	5	4	0	Utp6p-like protein
	cgd8_1340	11	9	2	Hypothetical protein
	cgd8_4620	8	7	1	ISWI related chromatinic protein