Received: 18 October 2021 | Revised: 19 May 2022 | Accepted: 20 May 2022

DOI: 10.1111/mec.16556

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Recent genetic exchanges and admixture shape the genome and population structure of the zoonotic pathogen *Cryptosporidium parvum*

Giulia I. Corsi[1,2] | Swapnil Tichkule[3,4] | Anna Rosa Sannella[5] | Paolo Vatta[5] | Francesco Asnicar[1] | Nicola Segata[1] | Aaron R. Jex[3] | Cock van Oosterhout[6] | Simone M. Cacciò[5]

[1]CIBO, University of Trento, Trento, Italy

[2]Department of Veterinary and Animal Sciences, Center for non-coding RNA in Technology and Health, University of Copenhagen, Frederiksberg, Denmark

[3]Population Health and Immunity, Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia

[4]Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Victoria, Australia

[5]Department of Infectious Diseases, European Union Reference Laboratory for Parasites, Istituto Superiore di Sanità, Rome, Italy

[6]School of Environmental Sciences, University of East Anglia, Norwich, UK

Correspondence

Simone M. Cacciò, Department of Infectious Diseases, European Union Reference Laboratory for Parasites, Istituto Superiore di Sanità, Viale Regina Elena, 299, Rome 00161, Italy.
Email: simone.caccio@iss.it

Cock van Oosterhout, School of Environmental Sciences, University of East Anglia, Norwich, UK.
Email: c.van-oosterhout@uea.ac.uk

Funding information

Horizon 2020 Framework Programme, Grant/Award Number: 643476

Handling Editor: Katherine Amato

Abstract

Cryptosporidium parvum is a globally distributed zoonotic pathogen and a major cause of diarrhoeal disease in humans and ruminants. The parasite's life cycle comprises an obligatory sexual phase, during which genetic exchanges can occur between previously isolated lineages. Here, we compare 32 whole genome sequences from human- and ruminant-derived parasite isolates collected across Europe, Egypt and China. We identify three strongly supported clusters that comprise a mix of isolates from different host species, geographic origins, and subtypes. We show that: (1) recombination occurs between ruminant isolates into human isolates; (2) these recombinant regions can be passed on to other human subtypes through gene flow and population admixture; (3) there have been multiple genetic exchanges, and most are probably recent; (4) putative virulence genes are significantly enriched within these genetic exchanges, and (5) this results in an increase in their nucleotide diversity. We carefully dissect the phylogenetic sequence of two genetic exchanges, illustrating the long-term evolutionary consequences of these events. Our results suggest that increased globalization and close human-animal contacts increase the opportunity for genetic exchanges between previously isolated parasite lineages, resulting in spillover and spillback events. We discuss how this can provide a novel substrate for natural selection at genes involved in host–parasite interactions, thereby potentially altering the dynamic coevolutionary equilibrium in the Red Queens arms race.

KEYWORDS

Cryptosporidium parvum, evolution, gene flow, population admixture, recombination, whole genome sequencing

Giulia I. Corsi and Swapnil Tichkule should be considered joint first author.

Cock van Oosterhout and Simone M. Cacciò should be considered joint senior author.

1 | INTRODUCTION

Cryptosporidium (Phylum Apicomplexa) comprises many species with a global distribution that are important causes of diarrhoeal disease in mammals, including humans (Innes et al., 2020). Human disease burden is particularly high among children living in low-income countries, where *Cryptosporidium* is a leading cause of moderate-to-severe diarrhoea (Kotloff et al., 2013) and exerts a negative long-term impact on childhood growth and wellbeing (Khalil et al., 2018). In the absence of broadly effective drugs and no vaccine, control of cryptosporidiosis is heavily dependent on the prevention of infection, and novel interventions are urgently needed (Bhalchandra et al., 2018; Chavez & White, 2018).

The epidemiology of human cryptosporidiosis is complex, with transmission occurring indirectly via contaminated food or water or directly via contact with animals or other infected people (McKerr et al., 2018). Although up to 17 species cause human infection, most cases are due to *Cryptosporidium hominis*, which is anthroponotic, or *C. parvum*, which is zoonotic (Feng et al., 2018). Animal reservoirs, such as calves and lambs, play an essential role in the spillover and spillback to humans. In Europe, three zoonotic subtypes are thought to be responsible for about 85% of human *C. parvum* infections, occurring as both sporadic cases or in outbreaks (Cacciò & Chalmers, 2016). The reasons for the high prevalence of these specific subtypes are presently unknown.

The life cycle of *Cryptosporidium* consists of an obligate sexual phase within a single host. The transmission stage is a meiotic spore (the oocyst), and this is generated by sexual reproduction. This lifecycle is completed in 3 days (Guérin & Striepen, 2020), and hence, each infection probably offers multiple opportunities for recombination. However, if the host is infected by only a single strain, sexual reproduction effectively amounts to self-fertilization (or selfing). In the absence of multiple-infections, the reproduction of *Cryptosporidium* thus resembles that of parasites such as *Giardia*, *Toxoplasma*, *Cryptococcus*, and *Trypanosoma* that have a predominant clonal evolution (PCE) (Tibayrenc et al., 1990). If two or more *Cryptosporidium* strains infect a single host (i.e., multiple infections), the sexual phase can result in genetic exchanges, and the population structure may resemble that of panmixia. In these incidences, recombination can rapidly generate genetic novelty that forms the new substrate for selection and adaptive evolution. Several recent whole genome sequencing studies (Gilchrist et al., 2018; Nader et al., 2019; Tichkule et al., 2021) have explored the evolutionary genetics of *Cryptosporidium*. For example, a study of *C. hominis* in children from a Bangladeshi community found high levels of genomic recombination, possibly due to the high transmission rate and likelihood of mixed infections (Gilchrist et al., 2018). Tichkule et al. (2021) found evidence of population admixture between distinct geographical lineages of *C. hominis* in Africa, showing that these lineages have been exchanging genetic information through recombination. Similarly, Nader et al. (2019) found evidence of recent genetic exchanges between zoonotic *C. p. parvum* and anthroponotic *C. p. anthroponosum* subspecies and documented genetic introgression between *C. hominis* and *C. parvum*. The authors concluded that despite genetic adaptation to specific host species, *Cryptosporidium* (sub)species continue to exchange genetic variation through hybridisation.

In this study, we examined the role of gene flow and recombination in the population structure and evolution of *C. parvum*. We generated whole genome sequences of 22 human- or ruminant-derived infections from Europe and compared these with genomic data for 10 publicly available *C. parvum* infections from other continents. We studied the population structure of this parasite at a geographic scale, and tested the host specificity of different subtypes. We also analysed the evidence of gene flow and genetic exchange between human and ruminant isolates. In particular, we studied the effects of genetic exchange on the pattern of nucleotide diversity and differentiation. We then tested whether genes involved in host–parasite coevolution are more significantly affected by these exchanges. The overall aim of this study is to improve our understanding of the evolutionary epidemiology of this important zoonotic pathogen.

2 | MATERIALS AND METHODS

2.1 | Parasite isolates

Table 1 lists the information available for the 33 *Cryptosporidium parvum* isolates from humans and ruminants included in this study. These samples comprise 22 isolates sequenced in the present study, 10 samples from previous studies (Feng et al., 2017; Hadfield et al., 2015), and the reference genome (Baptista et al., 2022). An aliquot of the faecal samples was used to extract genomic DNA and to identify the species and the *gp60* subtype, using previously published protocols (Alves et al., 2003; Ryan et al., 2003). We point out that the isolates from which genome information was derived were not specifically collected for this study and that they represented a convenience (and suboptimal) sampling. Furthermore, as the number of isolates from each country is relatively small, inferences on the population structure and diversity should be still interpreted as working hypotheses.

2.2 | Oocyst purification and genomic DNA extraction

Faecal samples were washed with distilled water and pelleted by centrifugation (1100 *g*) for 5 min in a refrigerated centrifuge. The procedure was repeated three times, and the sediment suspended in ~1–2 ml of distilled water. A commercial immunofluorescent assay (Merifluor *Cryptosporidium*/*Giardia* kit) was used to estimate oocyst number and integrity (i.e., presence of intact nuclei).

Oocysts were purified by immunomagnetic separation (IMS) using the Dynabeads anti-*Cryptosporidium* kit (Idexx), according to the manufacturer's instructions. To degrade residual contaminants, purified oocysts were treated with an equal volume of 0.6% sodium hypochlorite, washed three times with nuclease-free water, and

TABLE 1 List of the *Cryptosporidium parvum* isolates included in the study

Isolate	Host	Year of collection	Country of origin	*gp60* subtype	Reference
IT-C320	Goat kid	2013	Italy	IIaA15G2R1	This study
IT-C366	Goat kid	2014	Italy	IIdA17G2R1	This study
IT-C385	Goat kid	2015	Italy	IIaA15G2R1	This study
IT-C386	Goat kid	2015	Italy	IIaA15G2R1	This study
IT-C388	Lamb	2015	Italy	IIaA15G2R1	This study
IT-C389	Lamb	2015	Italy	IIaA15G2R1	This study
IT-C390	Lamb	2016	Italy	IIaA15G2R1	This study
IT-C391	Calf	2016	Italy	IIaA19G2R1	This study
IT-C392	Lamb	2016	Italy	IIaA17G1R1	This study
IT-C393	Calf	2016	Italy	IIaA16G3R1	This study
IT-C394	Lamb	2016	Italy	IIaA16G1R1	This study
IT-C395	Goat kid	2016	Italy	IIaA18R1	This study
DK-C6	Calf	1996	Denmark	IIaA18G1R1	This study
Slo1	Human	2016	Slovenia	IIaA15G2R1	This study
Slo2	Human	2016	Slovenia	IIaA15G1R1	This study
Slo4	Human	2016	Slovenia	IIaA20G1R1	This study
Slo5	Human	2016	Slovenia	IIaA15G2R1	This study
Slo7	Human	2016	Slovenia	IIaA19G1R1	This study
Slo9	Human	2016	Slovenia	IIaA15G2R1	This study
Spa1	Human	2015	Spain	IIaA15G2R1	This study
IT-To	Calf	1990	Italy	IIdA22G2R1	This study
IT-Ve	Calf	2015	Italy	IIaA15G2R1	This study
UKP2	Human	2012	United Kingdom	IIaA19G1R2	Hadfield et al. (2015)
UKP3	Human	2013	United Kingdom	IIaA18G2R1	Hadfield et al. (2015)
UKP4	Human	2012	United Kingdom	IIaA15G2R1	Hadfield et al. (2015)
UKP5	Human	2012	United Kingdom	IIaA15G2R1	Hadfield et al. (2015)
UKP6	Human	2013	United Kingdom	IIaA15G2R1	Hadfield et al. (2015)
UKP7	Human	2013	United Kingdom	IIaA17G1R1	Hadfield et al. (2015)
UKP8	Human	2013	United Kingdom	IIdA22G1	Hadfield et al. (2015)
31,727	Calf	2008	China	IIdA19G1	Feng et al. (2017)
34,902	Buffalo calf	2011	Egypt1	IIdA20G1	Feng et al. (2017)
35,090	Calf	2011	Egypt2	IIaA15G1R1	Feng et al. (2017)
IOWA-ATCC	Calf	2020	USA	IIaA17G2R1	Baptista et al. (2022)

pelleted by centrifugation (1100*g* for 5 min). The pellets were suspended in 100 μl of nuclease-free water, of which 5 μl were used to estimate the number of oocysts by microscopy, and 95 μl were used for DNA extraction. To maximize DNA yield, purified oocysts were submitted to five cycles of freezing in liquid nitrogen and thawing at 55°C. Genomic DNA was extracted with the DNA extraction IQ System kit (Promega), following the manufacturer's instructions, and eluted in 50 μl of elution buffer. DNA concentration was measured using Qubit dsDNA HS Assay Kit and the Qubit 1.0 fluorometer (Invitrogen), according to the manufacturer's instructions. To assess the presence of residual bacterial DNA in the genomic extracts, a previously described single round PCR targeting the 16S rRNA gene was used (Kommedal et al., 2011). PCR products were visualized by agarose gel electrophoresis.

2.3 | Whole genome amplification and next-generation sequencing

Whole genome amplification (WGA) was performed using the REPLI-g Midi-Kit (Qiagen), according to the manufacturer's instructions. Briefly, 5 μl of genomic DNA (corresponding to 1–10 ng of genomic DNA) were mixed with 5 μl of denaturing solution and incubated at room temperature for 3 min. Next, 10 μl of stop solution were added to stabilize denatured DNA fragments. The reaction mixture was completed with 29 µl of buffer and 1 µl of phi29 polymerase, allowed to proceed for 16 h at 30°C, and then stopped by heating for 5 min at 63°C. WGA products were visualized by electrophoresis on a 0.7% agarose gel, purified and quantified by Qubit as described above.

For next-generation sequencing, about 1 µg of purified WGA product per sample was used to generate each Illumina TruSeq paired-end library. Libraries were sequenced on an Illumina HiSeq 4000 platform to a read length of 100 bp (for four isolates: DK-C6, IT-C366, IT-To and IT-Ve) or 150 bp (for the remaining 18 isolates). Library preparation and NGS experiments were performed by a sequencing service (Biodiversa, Italy). An average of 25.72 million paired-end reads per isolate was generated.

2.4 | Retrieval of public *C. parvum* whole genome sequence data

Raw reads and assembled genomes of *C. parvum* were retrieved from the NCBI database from bioprojects PRJNA253836, PRJNA253840, PRJNA253843, PRJNA253845, PRJNA253846, PRJNA253847, PRJNA253848 (Hadfield et al., 2015), and PRJNA320419 (Feng et al., 2017). The assembled reference genome of *C. parvum* IOWA-ATCC (Baptista et al., 2022) was downloaded from CryptoDB v.52 (Warrenfeltz et al., 2020).

2.5 | Processing of *C. parvum* sequencing data

Raw sequencing reads of 32 isolates (Table 1) were preprocessed to filter low-quality bases and adapters using Trimmomatic v.0.36 (Bolger et al., 2014), with default parameters. The *C. parvum* IOWA-ATCC (Baptista et al., 2022) was set as the reference genome to map the filtered reads with BWA-MEM v.0.7 (Li & Durbin, 2010). Read depth was computed by using SAMtools (Li et al., 2009). Multiple sequence alignments at chromosomal level were generated with a module included in a pipeline previously developed (https://github.com/EBI-COMMUNITY/ebi-parasite,module"build_chr_multiAlign_for_recombi"), provided with SAMtools v.1.9 (Li et al., 2009), Pilon v.1.24 (Walker et al., 2014), and progressive Mauve v.2.4.0 (Aaron et al., 2010). Variant calling was performed with the HaplotypeCaller module of GATK v.3.7, applying hard filters as suggested in the GATK's Best Practices (van der Auwera et al., 2013) Furthermore, after joint genotyping of 32 individual variant call format (VCF) files, we excluded SNPs with alleles <5X coverage, missingness >50%, MAF < 0.05 and alternate allele depth < 50%. The moimix R package (https://github.com/bahlolab/moimix) was used to estimate multiplicity of infection. The FWS statistic was calculated, which is a type of fixation index to assess the within-host genetic differentiation (Manske et al., 2012). In pure isolates with haploid genomes, FWS is expected to approach unity. Isolates with FWS < 0.95 were excluded, as they are likely to represent multiple infections (Manske et al., 2012). For isolates with FWS > 0.95, each multiallelic SNP position was reduced to a biallelic position by retaining the dominant allele.

2.6 | Phylogenetic and cluster analyses

Phylogenetic trees were inferred on the concatenated set of genomic SNPs using the maximum likelihood method and general time reversible model, as implemented in the MEGA software, version 7 (Kumar et al., 2018). The reliability of the clusters was evaluated by the bootstrap method with 1000 replicates. The set of genomic SNPs was used to visualize clustering among isolates using a principal component analysis (PCA), as implemented in Clustvis (https://biit.cs.ut.ee/clustvis/) (Metsalu & Vilo, 2015). Population structure analysis of the isolates was performed with the STRUCTURE v.2.3 program (Pritchard et al., 2000). Phylogenetic networks were generated by using the neighbour-net algorithm implemented in SplitsTree v.5 (Huson & Bryant, 2006). DensiTree 2 was used to construct a consensus tree (Bouckaert & Heled, 2014).

2.7 | Recombination analyses

The multiple sequence alignments of each chromosome were analysed by the Recombination Detection Program software, version 4 (RDP4) (Murrell et al., 2015) using RDP, Geneconv, Bootscan, MaxChi, and Chimæra. A *p*-value cutoff <10E−5 was used to filter significant events. The *p*-value represents the probability that the identified recombination block results from the accumulation of mutations rather than by recombination. The "major parent" is related to the greater part of the recombinant's sequence (the recipient), whereas the "minor parent" is related to the sequences in the proposed recombinant region (the donor).

To visualize the recombinant blocks identified by RDP4, we used the HybridCheck software (Ward & van Oosterhout, 2016), with a step size of 1 and a window size of 500 SNPs. The same settings were used to date the recombination events, assuming a mutation rate of 10^{-8} and a generation time of 48 h.

Furthermore, to visualize the relationships between the isolates and infer the population history, we used the POPART package (Leigh & Bryant, 2015) and generated haplotype networks of the sequences where the recombinant blocks were identified.

2.8 | Population diversity and genetic variation

The level of pairwise divergence (Dxy) and intrapopulation nucleotide diversity (π) of the three clusters inferred from the phylogenetic analyses were computed along each chromosome using nonoverlapping genomic windows of 5 kb. Peritelomeric and subtelomeric regions were defined as the first and second 5% of the chromosome sequence at each end, respectively (Nader et al., 2019). The significance of the changes in nucleotide diversity associated with each recombination event was evaluated with a binomial test given the chromosomal SNP probability in the recombinant cluster. For events involving recombinants from multiple clusters, the recombinant cluster was set as the modal cluster among the recombinant samples. The RDP4 results were filtered by removing regions >100 kb and entries with <20 SNPs according to our variant calling analysis to avoid calls triggered by singletons in the multiple sequence alignments. The analysis of genes overlapping highly diverse and divergent regions was conducted using the annotations available in CryptoDB (n = 3894 genes with annotated coding sequences, CDSs). Gene descriptions were obtained from the supplementary material of the *C. parvum* IOWA-ATCC genome (Baptista et al., 2022) and by matching mRNA sequences against genes annotated in the previous reference (Abrahamsen et al., 2004) with BLASTn (min identity = 95%, min coverage = 75%). Genetic variability was estimated as the average number of SNPs between all isolates for each gene (considering only the regions annotated as CDS), normalized by the reference coding sequence length. Genes were labelled as highly polymorphic if their variability was greater than 10 times that computed over the whole genome (>6 SNPs/kb, n = 26 highly polymorphic genes). The enrichment for highly polymorphic genes in recombinant regions was computed via the Chi-squared test of independence on a contingency table containing gene counts split by variability level and location (highly polymorphic, in recombinant region $n = 16$ or outside $n = 10$; not highly polymorphic in recombinant region $n = 318$ or outside $n = 3550$). The odds ratio was used to determine the direction of the enrichment. Functional annotations such as EC numbers, GO functions, predicted signal peptides and transmembrane domains were retrieved from CryptoDB. The amino acid sequences annotated for each protein-coding gene, downloaded from CryptoDB, were screened for the presence of single amino acid Repeats (SAARs) by moving over the amino acids string while looking for at least five consecutive instances of the same letter. For genes with more than one annotated protein variant (CPATCC_000022, CPATCC_000589, CPATCC_000639) all the SAARs found in any of the variants were listed.

2.9 | De novo assembly of *C. parvum* genomes

Raw reads were assembled as contigs using SPAdes v3.10.1 (flag "--careful" on) (Bankevich et al., 2021). Contigs shorter than 1 kb were removed, and contaminants were identified for each raw genome using MegaBLAST (Zhang et al., 2000) by assigning each contig to the genus matching with the highest percentage of coverage (at least 5%) and similarity (at least 75%). Given the high similarity between *C. hominis* and *C. parvum*, contigs matching either species with coverage >90% were assigned to *C. parvum* and retained for further analyses. Contigs matching with coverage <90% to the *Cryptosporidium* genus were discarded.

3 | RESULTS

3.1 | Data set of new and available *C. parvum* whole genomes

We generated whole genome sequences from 7 human- and 15 ruminant-derived *C. parvum* isolates collected in Denmark, Italy, Slovenia or Spain (Table 1). In addition, we retrieved raw sequence data from seven human-derived *C. parvum* isolates from the United Kingdom (UK) and three ruminant-derived isolates from China and Egypt (Table 1). Each isolate was compared to a recently published and essentially complete assembly of the *C. parvum* IOWA-ATCC isolate (Table 1).

3.2 | Identification of mixed populations in raw sequence data

We used moimix and the FWS statistics to estimate the presence of mixed infections in the raw sequence data of this study. We identified three ruminant-derived isolates from Italy, that is, It-C394, It-To and It-Ve (Figure S1) as mixed infections; these were removed from subsequent analyses. We also identified and removed contigs and associated reads that matched bacterial, fungal or host-derived contaminants using BLASTn, following the de novo assembly of each of the 22 newly sequenced isolates in this study (Table S1). The number of *Cryptosporidium*-derived contigs for the 22 newly sequenced isolates ranged from 30 (IT-C389) to 65 (IT-C366), whereas it ranged from 91 (UKP6) to 2435 (Egypt2) for those retrieved from GenBank (Table S2). The larger fragmentation of the retrieved assemblies was further shown by the substantially lower N50 (range, 22 to 222 kb, average 113 Kb) compared to that of the newly sequenced assemblies (range, 258 to 874 kb, average 558 kb) (Table S2). These data confirm that high-quality *Cryptosporidium* whole genome sequences can be obtained directly from clinical stool samples (Feng et al., 2017; Hadfield et al., 2015; Nader et al., 2019).

3.3 | Genetic variability among isolates

We identified 9806 biallelic SNPs across all genomes relative to the IOWA ATCC reference genome (CryptoDB release 52, 2021-05-20). The number of SNPs ranged from 709 for the human isolate UKP7 to 5650 for the calf isolate Egypt1 (Table S3). The four isolates of the IId lineage had a larger number of SNPs (4711 ± 778) compared to the 13 isolates of the IIaA15G2R1 subtype in the IIa lineage (1087 ± 277). The SNPs were not randomly distributed across the eight chromosomes, with chromosomes 1 and 6 showing statistically higher density than the other chromosomes (binomial tests, $p < 10^{-15}$ for both chromosomes). Chromosomes 2 and 8 showed the lowest SNP densities (binomial tests, $p = 1.85 \times 10^{-24}$ and $p = 1.07 \times 10^{-16}$, respectively).

3.4 | Phylogenetic and clustering analyses

Phylogenetic relationships among the *C. parvum* isolates (Figure 1a), inferred using all biallelic SNPs ($n = 9806$), identified three clusters. However, the isolates did not cluster by host species, geographic origin or gp60 subtype. These conclusions were further supported by a consensus tree generated using the DensiTree 2 software (Figure S2A), by a Principal Component Analysis (PCA) (Figure S2B) and an analysis of the population structure executed with STRUCTURE (Figure S2C). Interestingly, the latter analysis revealed a moderate level of admixture among the three clusters. We examined admixture in further detail and generated a phylogenetic network with SplitsTree (Figure 1b). This confirmed the presence of three distinct clusters that were connected by several loops, shown in the network as parallel interconnected edges forming reticulates which we interpreted as evidence for recombination events within and between clusters. Loops in a phylogenetic network indicate phylogenetic inconsistencies, where part of the focal sequence resembles that of a second sequence, and part is more similar to a third sequence. If a fourth sequence is added to such a network, the first three sequences pull a loop. If those loops are large, this suggests that the sequence similarity between the focal, second and third sequence is high at the points where they match. This indicated that the genetic exchange between those three sequences is likely to have occurred recently.

3.5 | Genome-wide recombination analysis

We identified 73 genomic recombination events among all isolates (Table S4), and found that the recombinant regions were not homogeneously distributed across chromosomes (Table S5). Recombinant regions were detected significantly more frequently on chromosomes 1 and 3 (binomial tests: $p = 3.002\text{E}{-5}$ and $p = .00195$, respectively), and chromosomes 7 and 8 showed significantly fewer recombinant regions than expected (binomial tests: $p = .00241$, and $p = 2.336\text{E}{-5}$, respectively). Recombination events were not homogeneously distributed within chromosomes (Table S6), with significantly more events detected at peri- and subtelomeric regions, particularly for chromosomes 4 and 6 (binomial tests: $p = .004672$, and $p = 1.011\text{E}{-6}$, respectively).

We then asked if the level of polymorphism differed depending on whether genes were found inside or outside of a recombinant region. Genomic regions involved in recombination were significantly enriched for highly polymorphic genes compared to other regions ($\chi^2 = 86.94$; $p = 1.11\text{E}{-20}$, df = 1). Moreover, genes in recombinant regions displayed a higher level of intrapopulation diversity compared to the rest of the genome, independent of the population clusters (two-tailed *t*-test, $p = 2.78\text{E-}18$, $p = 5.72\text{E-}17$ and $p = 2.09\text{E}{-21}$, for clusters 1, 2 and 3, respectively). Among the highly variable protein-coding genes located in recombinant regions,

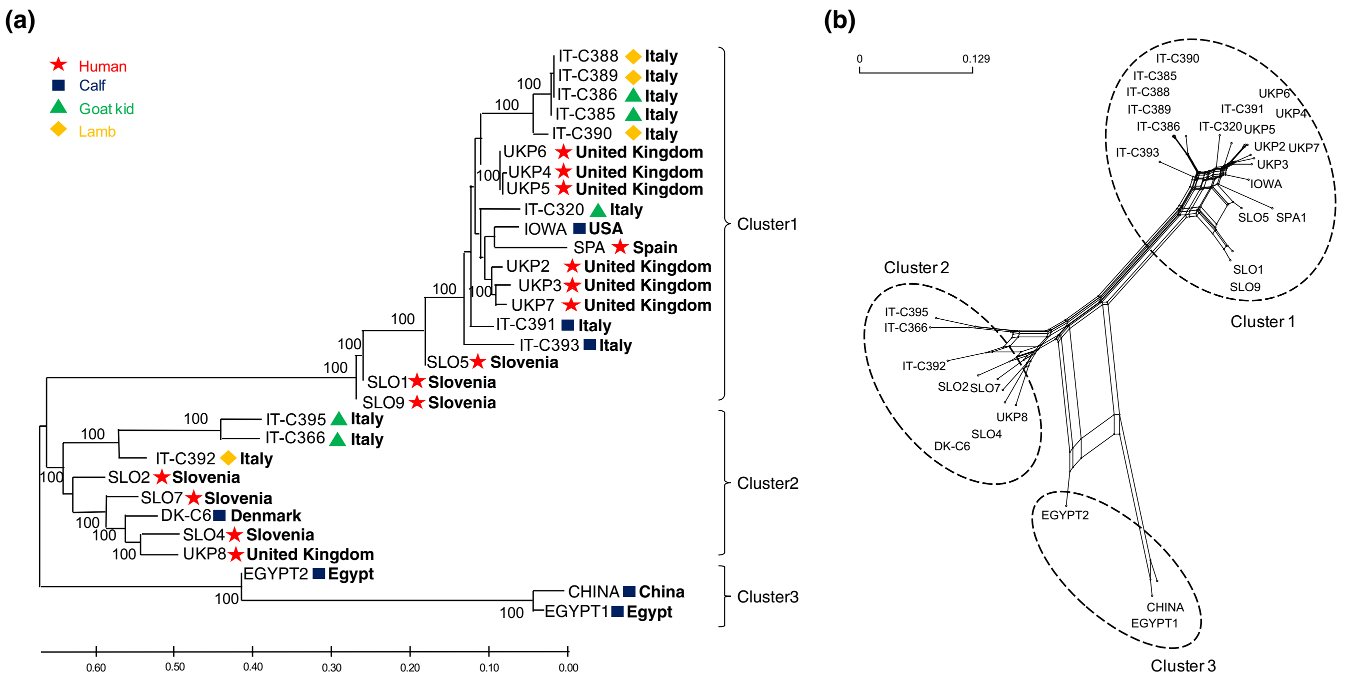

FIGURE 1 (a) Phylogenetic relationships based on concatenated genomic SNPs. Trees were inferred using maximum likelihood with bootstrap values. Branches are labelled with isolate ID, host of origin and geographical origin. The symbols indicate the host species the sample was collected from. Branch lengths represent the number of substitutions per site. (b) a phylogenetic network generated by SplitsTree shows the presence of three distinct clusters connected by several loops, indicative of recombination and outcrossing [Colour figure can be viewed at wileyonlinelibrary.com] nine have a signal peptide, and seven have single amino acid repeats (SAARs) (Table S7). In general, genes with a signal peptide or a SAAR located in recombinant regions ($n = 117$) have higher nucleotide variation compared to the other genes ($n = 1069$) with the same properties but located elsewhere in the genome (two-tailed *t*-test, $p = 4.11E{-}18$). This suggests that recombination might have enabled the adaptive evolution of genes potentially involved in virulence by increasing their nucleotide diversity.

3.6 | Recombination events occurred recently

We estimated the mean age of the recombination events between the three *C. parvum* linages after excluding those for which the parental sequence was missing and/or the breakpoints were ambiguous. For the events where multiple isolates were identified by RDP4 as recombinant or minor parents, one representative of each was chosen, but multiple combinations were tested to ensure that the age estimates were consistent. Using these criteria, we examined 24 recombination events and showed that 12 (50%) probably occurred very recently, that is, in the past ~200 years (Figure 2a). We did detect much more ancient events, which shows that we were able to detect genetic exchanges that occurred in the distant past.

To examine whether the rate of genetic exchanges might have increased only recently, we calculated the cumulative percentage of recombinant nucleotides against the age of the recombination event (in years). This showed that about 22% nucleotides have been exchanged between the three lineages in the past ~200 years (Figure 2b). We argue that, if the historic rate of recombination would have been as high as that of the past ~200 years, then the three lineages would have been homogenized (i.e., panmictic), and they would not appear genetically distinct. This suggests that the rate of genetic exchanges has increased in the past two centuries.

Next, we focus on recombination events in chromosomes 1 and 4 to illustrate the effect of genetic exchanges on nucleotide variation between isolates within a cluster and genetic divergence between clusters, and to better understand the evolutionary epidemiology of such genetic exchanges.

3.7 | Chromosome 1

Chromosome 1 shows two interesting cases of recombination involving human- and ruminant-derived isolates from different clusters. Network analysis shows that a human isolate from Spain (Spa1) and a calf isolate from Egypt (Egypt2) have markedly different branching patterns (Figures 3a–b). Indeed, both the isolates are clustered outside their original clusters. In the phylogenetic network (Figure 3b), large loops connecting isolate Spa1 and Egypt 2 with cluster 3 isolates are observed, suggesting that these two isolates have parts of their chromosome 1 sequences highly similar to cluster 3 as a result of admixture. Examining the recombination signal in this chromosome using HybridCheck analysis pinpoints the exact chromosome location of the genetic exchanges (Figure 3c).

In the case of Spa1, the exchange involved a region of 22 kb that, according to HybridCheck, is between positions 773,084 and 796,671. This region is characterized by a significantly higher level of nucleotide diversity (π) in cluster 1 than expected (Figure S3). This is reflected by the 13 genes found in this region (CPATCC_000384 to CPATCC_000397), two of which (CPATCC_000384 and CPATCC_000385) are among the top 1%–2% most polymorphic genes (Table S7).

In the case of Egypt2, the exchange involved a region of 26 kb located between position 745,529 and 771,337, and this region is also characterized by a significantly higher than expected level of nucleotide diversity (π) (Figure S3). As for Spa1, this is reflected by the 11 genes found in this region (CPATCC_000372 to CPATCC_000383), three of which (CPATCC_000376, CPATCC_000378, CPATCC_000383) are among the top 1%–2% most polymorphic genes (Table S7). A schematic representation of the evolutionary events involving genetic exchanges between the isolates of cluster 1 and cluster 3 is illustrated in Figure 3d. This diagram shows that the ancestors of isolates Egypt2 and Spa1 both received genetic variation from a cluster 3, around 49 (21–96; 95% CI) and 289 (204–395; 95% CI) years ago, respectively.

3.8 | Chromosome 4

Chromosome 4 shows an interesting case of genetic exchange, wherein exactly the same ~8 kb region of cluster 3 is found in isolates from both cluster 1 and cluster 2. Two isolates from cluster 2, namely Slo2 (human) and C392 (lamb), and two human-derived isolates from cluster 1 (Slo1 and Slo9) are found outside their original clusters in the phylogenetic tree (Figure S4A of chromosome 4) and show signs of admixture with cluster 3 in the population structure analysis performed with STRUCTURE (Figure S4B of chromosome 4). Furthermore, large loops are seen in the phylogenetic network (Figure 4a) suggesting recombination may have occurred.

A visual inspection of the multiple sequence alignment shows that, at a region close to the 5′ telomere (from position 751 to position 8057) of chromosome 4, these isolates share 242 SNPs with cluster 3 isolates. The recombination detection software RDP4 identifies recombination at this region (Table S4). Further inspection of this chromosome with HybridCheck supports that the same ~8 kb region at the 5′ telomere of chromosome 4 from cluster 3 is now found in isolates of both cluster 1 and cluster 2 (Figure 4c). We estimated that both clusters received genetic variation from cluster 3 circa 472 (286–723; 95% CI) years ago through recombination. It is unlikely these were two independent events, given that the start and end position of the blocks are identical in both clusters. Rather, we propose cluster 3 first recombined with a genome of one of the other clusters (e.g., cluster 1), and that the descendants of this recombinant subsequently recombined with the

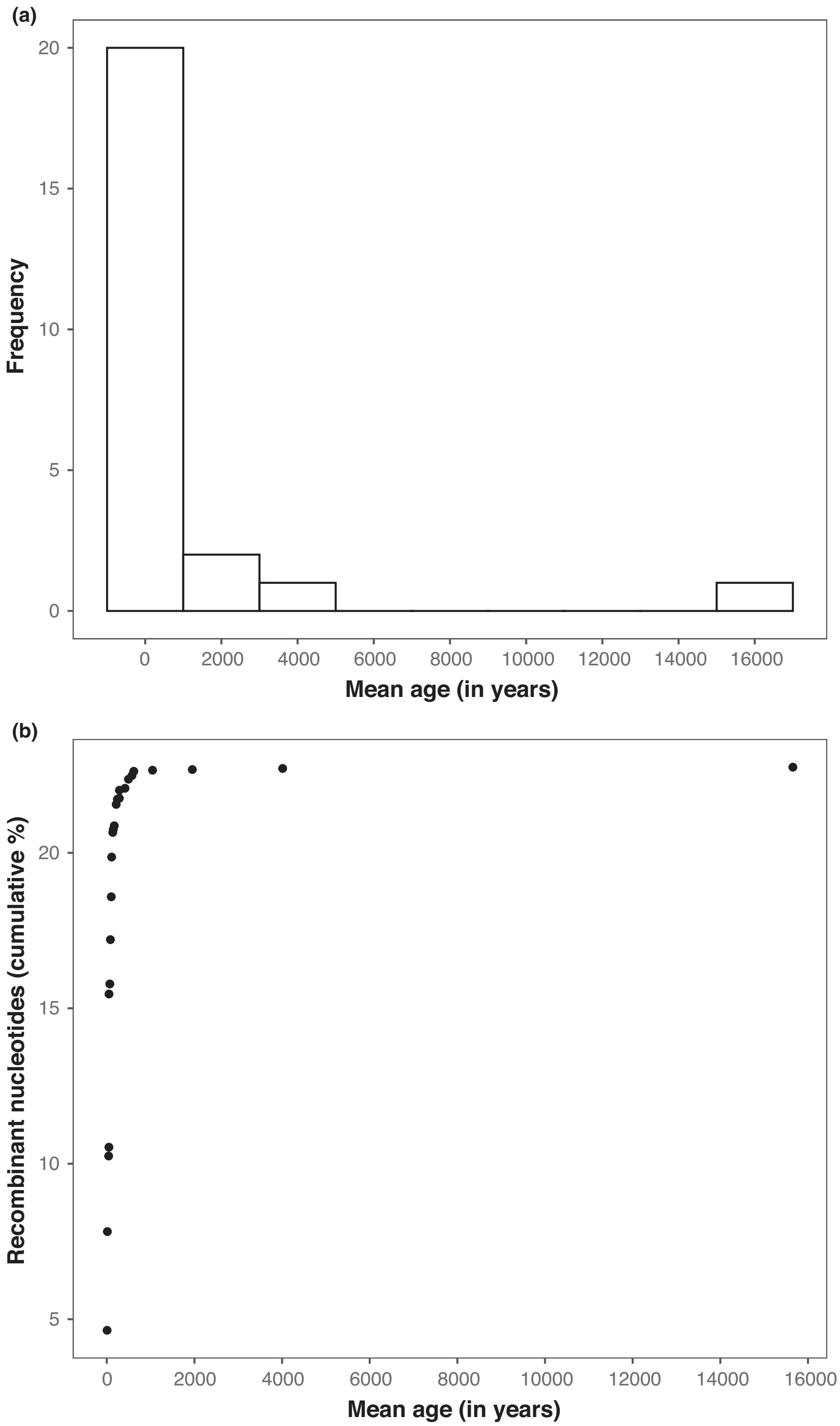

FIGURE 2 (a) Distribution of the mean age (in years) of 24 recombination events between the three *C. parvum* linages. Most events are recent, with 12 out of 24 (50%) having occurred in the past 200 years. Older events can still be detected, as evidenced by some events dating back ~16,000 years ago. (b) Cumulative percentage of recombinant nucleotides against the age of the recombination event (in years) ranked from recent to most ancient. Recent recombination events exchanged ~22% nucleotides between the three lineages in the past ~200 years. If the historic rate of recombination would have been as high as the recent rate in the past ~200 years, the three lineages would have been homogenized (i.e., panmictic), and they would not be genetically distinct. This suggests that the recombination rate has increased in recent times

third cluster (e.g., cluster 2). In other words, this is consistent with a potential spillover and spillback event, from calf into human, and back into lamb. The haplotype network plot generated with Popart (Figure 4b), further supports the reconstruction of the sequence of events, which is schematically illustrated in Figure 4d. The recombinant region, now shared between extant isolates from cluster 1 (Slo1 and Slo9) and cluster 2 (C392 and Slo2), comprises four genes (CPATCC_001472-CPATCC 001475), which are among the top 1%–2% most polymorphic in the genome, encoding for proteins with still unknown function (Table S7).

4 | DISCUSSION

We performed an evolutionary genetic analysis, studying whole genome sequence data of 32 isolates of *C. parvum*, which include 22

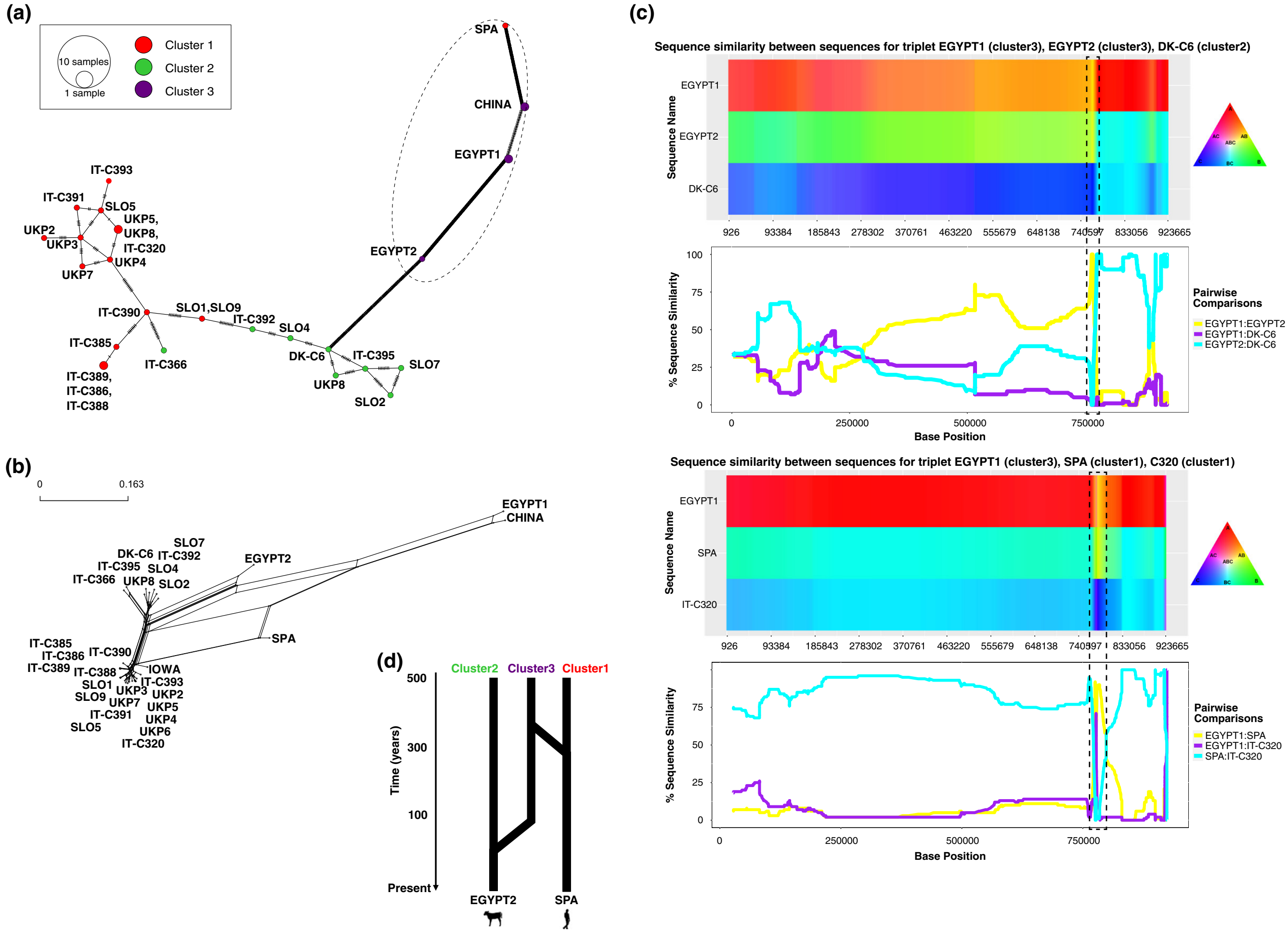

FIGURE 3 Recombination analysis of chromosome 1 identifies two hybrid sequences, Spa1 (a human isolate from cluster 1) and Egypt 2 (a calf isolate from cluster 3). (a) Network showing that haplotypes of hybrid isolates Spa1 and Egypt2 are diverged from the rest of the isolates and are closely associated with their minor parental sequences (cluster 3 isolates). (b) Phylogenetic network showing loops between hybrid isolates, representing potential recombination. (c) Sequence similarity plots obtained with HybridCheck for different triplets of isolates involved in a recombination event (hybrid, major parent, minor parent). The sequence similarity is shown along chromosomal positions (x-axis) by a colour map in which regions with the same polymorphism share colours (top) and by line graphs reporting the percentage similarity on the y-axis (bottom). Graphs recombinant regions that present showing high similarity between hybrid isolates and their minor parental sequences at the recombinant regions are enclosed in dashed boxes. (d) Schematic representation of recombination events in hybrid isolates, which shows that Egypt2 and Spa1 both received genetic variation from a cluster 3, around 49 (21–96; 95% CI) and 289 (204–395; 95% CI) years ago, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

newly sequenced isolates from humans and young ruminants across European, and 10 published genomes from previous studies (Feng et al., 2017; Hadfield et al., 2015). We identified three strongly supported clusters that grouped independently of host species and geographic location. Previous studies, mostly based on length variation of tandem DNA repeats, showed that the *C. parvum* population structure can be either panmictic, epidemic or clonal, probably as a result of differences in ecological factors, such as transmission intensity (Feng et al., 2018). In general, studies have reported a correlation between genetic and geographic distance, consistent with a model of isolation by distance, while support for host-adapted genotypes was less consistent. As the data set investigated here was small and based on a convenience sampling, a direct comparison with previous data is difficult. However, our data suggests a recent increase in genetic exchanges between the three lineages affected a substantial proportion of the genome (~22%). Furthermore, these lineages were independently of host species and geographic location. The elevated rate of sequence exchange implies that the population structure is dynamic, possibly as the result of new opportunities caused by global environmental changes. We anticipated that the current efforts to sequence additional genomes will allow comparing the population structure in further detail and that this will need to be monitored closely to track the population dynamics and continued (recombination-fuelled) evolution of this parasite. Despite any apparent reproductive barriers, reproductive isolation in *C. parvum* appears to be sufficiently strong for "near-clades" (NCs) to evolve, which are reasonably stable in time and space. Such population structuring is consistent with the predominant clonal evolution

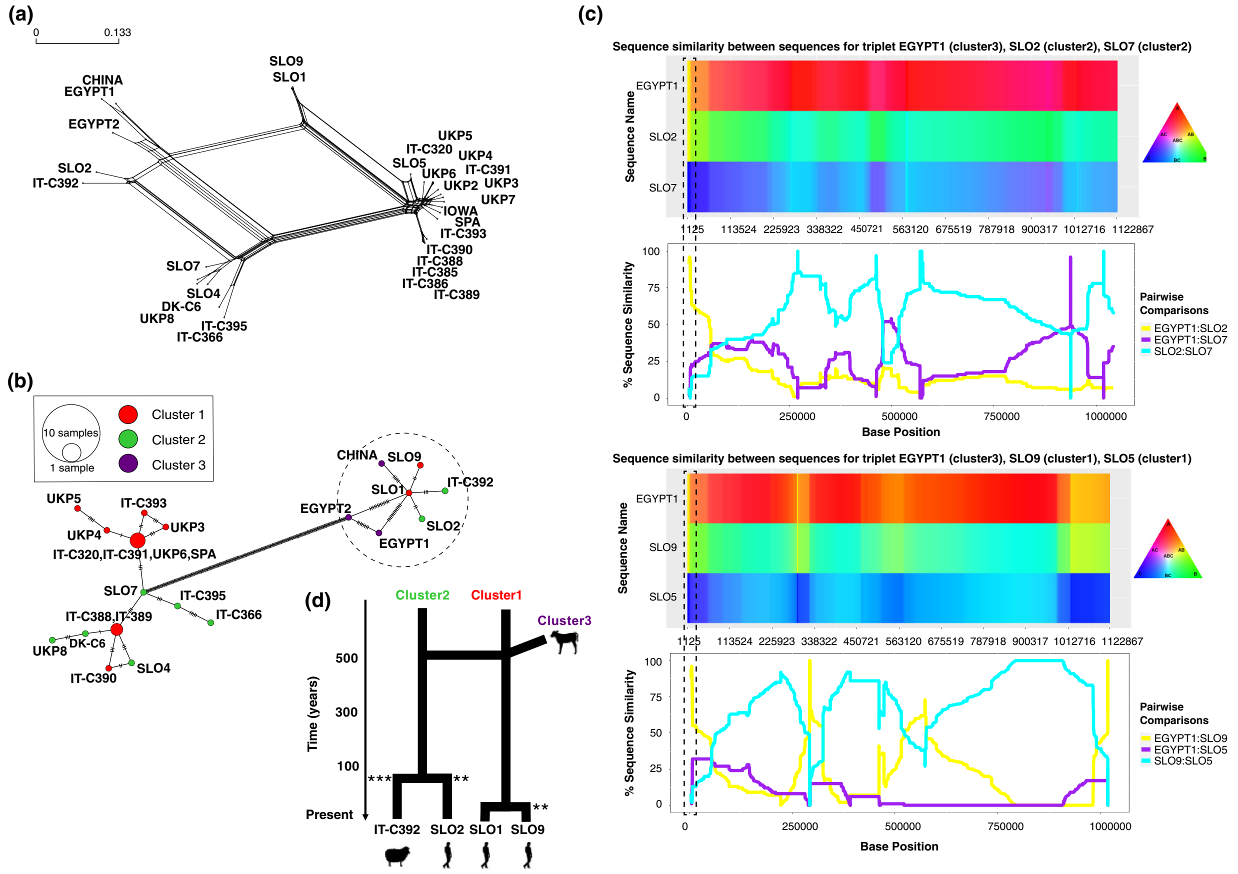

FIGURE 4 Recombination analysis of chromosome 4 identifies an exchange of the same 8 kb genomic region from cluster 3 into isolates from cluster 1 and 2. (a) the phylogenetic network shows the signature of recombination and admixture in four hybrid isolates, that is, Slo1 and Slo9 (cluster 1), and Slo2 and IT-C392 (cluster 2) placed outside their designated clusters. (b) Haplotype network shows that the recombination event first involved cluster 1 and was then passed on to cluster 2. (c) the sequence similarity graphs obtained with HybridCheck graphs show high similarity between hybrid isolates and their minor parental sequences at the recombinant regions (dashed boxes). For each triplet of isolates examined, the sequence similarity is shown along chromosomal positions (x-axis) by a colour map in which regions with the same polymorphism share colours (top) and by line graphs reporting the percentage similarity on the y-axis (bottom). (d) Schematic representation of introgression events in hybrid isolates. Subsequent recombination within cluster 1 and cluster 2 occurred circa 472 (286–723; 95% CI) years ago and is consistent with a potential spillover and spillback event between different zoonotic hosts and human host. Stars indicate the number of SNPs among the hybrid isolates [Colour figure can be viewed at wileyonlinelibrary.com]

(PCE), resulting in robust phylogenetic clustering, broken down only by occasional recombination (Tibayrenc et al., 1990). PCE is typical for several parasites and pathogens, including *Giardia*, *Toxoplasma*, *Cryptococcus*, *Trypanosoma* and others. The clonal lineages form due to a lack of outcrossing caused by their low transmission rates, resulting in only few multiclonal infections (compared with the "starving sex hypothesis") (Tibayrenc & Ayala, 2004). However, *C. parvum* may be on the "clonality threshold", right on the tipping point where recombination starts to break down the structuring typical for clonal evolution of many pathogens (Tibayrenc & Ayala, 2021) (Figure S5). Our analysis shows that the rate of genetic exchanges between lineages has significantly increased in the past 200 years (Figure 2a), which suggests that the distinct population structuring typical for PCE is moving towards increased panmixia. We first discuss our observations, and then consider the evolutionary consequences, hypothesizing how increased globalization and human-induced changes in the environment may influence the evolution of this and other parasites.

We investigated the role of recombination in shaping the population structure and the genome of the *C. parvum* isolates. We noticed that the three lineages show signs of admixture, and we detected 73 strongly supported recombination events. These were not distributed randomly across the genome but occurred more frequently than expected on chromosomes 1 and 3, and at peri- and subtelomeric regions of chromosomes 4 and 6. The exact recombinant and the major and minor parents were not identified in many of these events, due to the limited data set and the genomic similarity among isolates within each lineage. However, some events for which we did have this information were analysed in detail, allowing us to document potential spillover and spillback events. Our analyses suggest that recombination occurred from ruminant isolates into human isolates, and back into another zoonotic host species (i.e., lamb). We thus show that the recombinant regions are transmitted through the population by gene flow and admixture. Although this suggests that recombination is directional, we should clarify that these are the evolutionary events we witness after the filter of natural selection, and in a data set comprising only 32 isolates. In other words, genetic exchanges may have occurred into the opposite direction (e.g., from human strains into strains of ruminants), but they were not detected in our data set.

Importantly, we can, however, conclude that these genetic exchanges are probably recent; they are estimated to have occurred between 49 to 472 years ago, that is, within the Anthropocene. Indeed, the rate of genetic exchanges appears to have significantly increased over the past ~200 years. This is not due to our inability to detect ancient recombination events, as a small number of much older events were detected by our analyses (Figure 2a). Rather, the recently elevated rate of recombination between lineages is consistent with increased rates of contact caused by international travel and globalization. The huge biomass of our livestock (Bar-On et al., 2018) and close human-animal contacts have also increased the opportunity for genetic exchanges between previously isolated parasite lineages (van Oosterhout, 2021). This is resulting in spillover and spillback events, which may alter the dynamic coevolutionary equilibrium in the Red Queens arms race (van Oosterhout, 2021).

Lastly, we show that genes potentially implicated in virulence (here characterized by the presence of signal peptide and SAAR in the encoded proteins) are significantly more affected by genetic exchanges, and that the nucleotide diversity of these genes is elevated by these genetic exchanges.

A similar pattern, with an uneven distribution of recombination events and frequent involvement of peri- and subtelomeric regions, has been reported in a study of zoonotic and anthroponotic isolates of *C. parvum* and by comparison of *C. parvum* with *C. hominis* (Nader et al., 2019). Previous studies also showed that genes involved in host cell invasion (e.g., those encoding for mucin-like glycoproteins, thrombospondin-related adhesive proteins, secreted MEDLE family proteins, insulinase-like proteases and rhomboid-like proteases) are often located in the proximity of telomeres and are characterized by changes in copy number and high divergence among different *Cryptosporidium* species (Liu et al., 2016; Tichkule et al., 2022). More recently, it has been demonstrated that MEDLE-containing proteins are translocated into the cytosol of infected cells, after the parasite has established its intracellular niche, that is, after invasion (Dumaine et al., 2021). These proteins can modulate signalling pathways for intracellular development of the parasite. Further, antisera raised against recombinant MEDLE1 or MEDLE2 have been show to diminish the efficiency of sporozoite infection by 40% (Fei et al., 2018; Li et al., 2017), stressing the important role played by these proteins.

Therefore, genetic exchanges involving these fast-evolving genes are likely to have an important impact on the evolution of host specificity and virulence in the *Cryptosporidium* genus. We argue that these processes are even more important for a parasite like *C. parvum* that has many host species. Beneficial mutations that arise in the parasite lineage of one host species may thus be transmitted to lineages or clusters infecting other host species. Given that the opportunities of genetic exchanges have increased recently (Cutler et al., 2010; Davies Calvan & Šlapetam, 2021; Easton et al., 2020; Rohr et al., 2019), previously isolated clonal lineages are now in contact. The novel variation that is introduced by recombination between genetically diverged lineages rapidly generates new substrate for adaptive evolution of parasites, and this could give parasites an important evolutionary advantage in the antagonistic host–parasite coevolution. The shift in the balance of the evolutionary forces towards more recombination and less genetic drift offers more opportunities for natural selection (Figure S5).

Our study shows that whole genome sequence data can improve our understanding of the evolutionary epidemiology of zoonotic pathogens by helping to identify host reservoirs, gene flow, recombination, spillover and spillback events. Such analyses are crucially important to underpin the One Health approach and mitigate the threat of emerging infectious diseases (King et al., 2015; Ogden et al., 2019; van Oosterhout, 2021; Webster et al., 2016).

AUTHOR CONTRIBUTIONS

Simone M. Cacciò conceived the study. Giulia I. Corsi, Swapnil Tichkule, Aaron R. Jex, Cock van Oosterhout, Paolo Vatta, Francesco Asnicar, and Nicola Segata performed the bioinformatic analyses. Anna Rosa Sannella and Simone M. Cacciò performed the bench work. Simone M. Cacciò, Aaron R. Jex and Cock van Oosterhout wrote the manuscript with input from all authors. The authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Dr Nicoletta D'Avino (Istituto Zooprofilattico Sperimentale, Perugia, Italy) for the parasite isolates of animal origin, Dr Barbara Soba (University of Ljubljana, Slovenia) and by Dr David Carmena (Spanish National Centre for Microbiology, Madrid) for the parasite isolates of human origin. This work was supported by the project "Collaborative management platform for detection and analyses of (re-) emerging and foodborne outbreaks in Europe" (COMPARE, www.compare-europe.eu), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 643476 to S.M.C. A.R.J was funded through an Australian National Health and Medical Research Council Leadership grant (APP1194330). The Walter and Eliza Hall Institute of Medical Research receives support through the Victorian State Government Operational Infrastructure Support and Australian Government National Health and Medical Research Council Independent Research Institute Infrastructure Support Scheme. S.T. was supported by Walter and Eliza Hall International PhD Scholarship and Melbourne Research Scholarship (MRS).

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw sequence data have been submitted to the Sequence Read Archive (SRA) under the project accession nos PRJNA634014, biosamples SAMN14979425 to SAMN14979431 (human isolates) and PRJNA633764, biosamples SAMN14969799 to SAMN14969813 (animal isolates).

ORCID

Giulia I. Corsi https://orcid.org/0000-0001-5932-0664
Swapnil Tichkule https://orcid.org/0000-0003-2940-9961
Paolo Vatta https://orcid.org/0000-0002-9333-8021
Francesco Asnicar https://orcid.org/0000-0003-3732-1468
Nicola Segata https://orcid.org/0000-0002-1583-5794
Cock van Oosterhout https://orcid.org/0000-0002-5653-738X
Simone M. Cacciò https://orcid.org/0000-0002-8561-1323

REFERENCES

Aaron, E., Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G. A., Xu, P., Bankier, A. T., Dear, P. H., Konfortov, B. A., Spriggs, H. F., Iyer, L., Anantharaman, V., Aravind, L., & Kapur, V. (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*, *304*, 441–445.

Alves, M., Xiao, L., Sulaiman, I., Lal, A. A., Matos, O., & Antunes, F. (2003). Subgenotype analysis of *cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. *Journal of Clinical Microbiology*, *41*, 2744–2747.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2021). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477.

Baptista, R. P., Li, Y., Sateriale, A., Sanders, M. J., Brooks, K. L., Tracey, A., Ansell, B. R. E., Jex, A. R., Cooper, G. W., Smith, E. D., Xiao, R., Dumaine, J. E., Georgeson, P., Pope, B. J., Berriman, M., Striepen, B., Cotton, J. A., & Kissinger, J. C. (2022). Long-read assembly and comparative evidence-based reanalysis of *cryptosporidium* genome sequences reveal new biological insights. *Genome Research*, *32*(1), 203–213. https://doi.org/10.1101/gr.275325.121

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on earth. *Proceedings of the National Academy of Sciences USA*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Bhalchandra, S., Cardenas, D., & Ward, H. D. (2018). Recent breakthroughs and ongoing limitations in *cryptosporidium* research. *F1000Res*, *7*, pii: F1000.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: seeing trees through the forest. *bioRxiv*, 012401. https://doi.org/10.1101/012401

Cacciò, S. M., & Chalmers, R. M. (2016). Human cryptosporidiosis in Europe. *Clinical Microbiology and Infection*, *22*, 471–480.

Chavez, M. A., & White, A. C. (2018). Novel treatment strategies and drugs in development for cryptosporidiosis. *Expert Review Anti Infection Therapy*, *16*, 655–661.

Cutler, S. J., Fooks, A. R., & van der Poelm, W. H. M. (2010). Public health threat of new, reemerging, and neglected zoonoses in the industrialised world. *Emerging Infectious Diseases*, *16*(1), 1–7.

Davies Calvan, N. E., & Šlapetam, J. (2021). *Fasciola* species introgression: Just a fluke or something more? *Trends in Parasitology*, *37*(1), 25–34.

Dumaine, J. E., Sateriale, A., Gibson, A. R., Reddy, A. G., Gullicksrud, J. A., Hunter, E. N., Clark, J. T., & Striepen, B. (2021). The enteric pathogen *Cryptosporidium parvum* exports proteins into the cytosol of the infected host cell. *eLife*, *10*, e70451. https://doi.org/10.7554/eLife.70451

Easton, A., Gao, S., Lawton, S. P., Bennuru, S., Khan, A., Dahlstrom, E., Oliveira, R. J., Kepha, S., Porcella, S. F., Webster, J., Anderson, R., Grigg, M. E., Davis, R. E., Wang, J., & Nutman, T. B. (2020). Molecular evidence of hybridisation between pig and human *ascaris* indicates an interbred species complex infecting humans. *eLife*, *9*, e61562.

Fei, J., Wu, H., Su, J., Jin, C., Li, N., Guo, Y., Feng, Y., & Xiao, L. (2018). Characterization of MEDLE-1, a protein in early development of *Cryptosporidium parvum*. *Parasites & Vectors*, *11*, 312. https://doi.org/10.1186/s13071-018-2889-2 PMID: 29792229.

Feng, Y., Li, N., Roellig, D. M., Kelley, A., Liu, G., Amer, S., Tang, K., Zhang, L., & Xiao, L. (2017). Comparative genomic analysis of the IId subtype family of *Cryptosporidium parvum*. *International Journal for Parasitology*, *47*, 281–290.

Feng, Y., Ryan, U. M., & Xiao, L. (2018). Genetic diversity and population structure of *cryptosporidium*. *Trends in Parasitology*, *34*, 997–1011.

Gilchrist, C. A., Cotton, J. A., Burkey, C., Arju, T., Gilmartin, A., Lin, Y., Ahmed, E., Steiner, K., Alam, M., Ahmed, S., Robinson, G., Uz Zaman, S., Kabir, M., Sanders, M., Chalmers, R. M., Ahmed, T., Ma, J. Z., Haque, R., Faruque, A. S. G., … Petri, W. A. (2018). Genetic diversity of *Cryptosporidium hominis* in a Bangladeshi community as revealed by whole-genome sequencing. *Journal of Infectious Diseases*, *218*, 259–264.

Guérin, A., & Striepen, B. (2020). The biology of the intestinal intracellular parasite *cryptosporidium*. *Cell Host & Microbe*, *28*(4), 509–515.

Hadfield, S. J., Pachebat, J. A., Swain, M. T., Robinson, G., Cameron, S. J., Alexander, J., Hegarty, M. J., Elwin, K., & Chalmers, R. M. (2015). Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples. *BMC Genomics*, *16*, 650.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267.

Innes, E. A., Chalmers, R. M., Wells, B., & Pawlowic, M. C. (2020). A one health approach to tackle cryptosporidiosis. *Trends in Parasitology*, *36*, 290–303.

Khalil, I. A., Troeger, C., Rao, P. C., Blacker, B. F., Brown, A., Brewer, T. G., Colombara, D. V., De Hostos, L., Engman, C., Guerrant, R. L., Haque, R., Houpt, R. H., Kang, G., Korpe, P. S., Kotloff, K. L., Lima, A. L. M., Petri, W. A., Jr., Platts-Mills, J. A., Shoultz, D. A., … Mokdad, A. H. (2018). Morbidity, mortality, and long-term consequences associated with diarrhoea from *cryptosporidium* infection in children younger than 5 years: A meta-analyses study. *Lancet Global Health*, *6*, e758–e768.

King, K. C., Stelkens, R. B., Webster, J. P., Smith, D. F., & Brockhurst, M. A. (2015). Hybridisation in parasites: Consequences for adaptive evolution, pathogenesis, and public health in a changing world. *PLoS Pathogens*, *11*(9), e1005098.

Kommedal, Ø., Lekang, K., Langeland, N., & Wiker, H. G. (2011). Characterisation of polybacterial clinical samples using a set of group-specific broad-range primers targeting the 16S rRNA gene followed by DNA sequencing and RipSeq analysis. *Journal of Medical Microbiology*, *60*, 927–936.

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Wu, Y., Sow, S. O., Sur, D., Breiman, R. F., Faruque, A. S., Zaidi, A. K., Saha, D., Alonso, P. L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., … Levine, M. M. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, GEMS): A prospective, case-control study. *Lancet*, *382*, 209–222.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549.

Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, B., Wu, H., Li, N., Su, J., Jia, R., Jiang, J., Feng, Y., & Xiao, L. (2017). Preliminary characterization of MEDLE-2, a protein potentially involved in the invasion of *Cryptosporidium parvum*. *Frontiers in Microbiology*, *8*, 1647.

Liu, S., Roellig, D. M., Guo, Y., Li, N., Frace, M. A., Tang, K., Zhang, L., Feng, Y., & Xiao, L. (2016). Evolution of mitosome metabolism and invasion-related proteins in *cryptosporidium*. *BMC Genomics*, *17*(1), 1006. https://doi.org/10.1186/s12864-016-3343-5

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J. B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., Su, X. Z., Amaratunga, C., … Kwiatkowski, D. P. (2012). Analysis of *plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, *487*(7407), 375–379.

McKerr, C., O'Brien, S. J., Chalmers, R. M., Vivancos, R., & Christley, R. M. (2018). Exposures associated with infection with *cryptosporidium* in industrialised countries: A systematic review protocol. *Systematic Reviews*, *7*, 70.

Metsalu, T., & Vilo, J. (2015). Clustvis: A web tool for visualising clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Research*, *43*(W1), W566–W570.

Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*, vev003.

Nader, J. L., Mathers, T. C., Ward, B. J., Pachebat, J. A., Swain, M. T., Robinson, G., Chalmers, R. M., Hunter, P. R., van Oosterhout, C., & Tyler, K. M. (2019). Evolutionary genomics of anthroponosis in *cryptosporidium*. *Nature Microbiology*, *4*, 826–836.

Ogden, N. H., Wilson, J. R. U., Richardson, D. M., Hui, C., Davies, S. J., Kumschick, S., Le Roux, J. J., Measey, J., Saul, W., & Pulliam, J. R. C. (2019). Emerging infectious diseases and biological invasions: A call for a one health collaboration in science and management. *Royal Society for Open Science*, *6*(3), 181577.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Rohr, J. R., Barrett, C. B., Civitello, D. J., Craft, M. E., Delius, B., DeLeo, G. A., Hudson, P. J., Jouanard, N., Nguyen, K. H., Ostfeld, R. S., Remais, J. V., Riveau, G., Sokolow, S. H., & Tilman, D. (2019). Emerging human infectious diseases and the links to global food production. *Nature Sustainability*, *2*, 445–456.

Ryan, U., Xiao, L., Read, C., Zhou, L., Lal, A. A., & Pavlasek, I. (2003). Identification of novel *cryptosporidium* genotypes from The Czech Republic. *Applied and Environmental Microbiology*, *69*, 4302–4307.

Tibayrenc, M., & Ayala, F. J. (2004). New insights into clonality and panmixia in *plasmodium* and *toxoplasma*. *Advances in Parasitology*, *84*, 253–268.

Tibayrenc, M., & Ayala, F. J. (2021). Models in parasite and pathogen evolution: Genomic analysis reveals predominant clonality and progressive evolution at all evolutionary scales in parasitic protozoa, yeasts and bacteria. *Advances in Parasitology*, *111*, 75–117.

Tibayrenc, M., Kjellberg, F., & Ayala, F. J. (1990). A clonal theory of parasitic protozoa: The population structure of *Entamoeba*, *giardia*, *leishmania*, *naegleria*, *plasmodium*, *trichomonas* and *Trypanosoma*, and its medical and taxonomical consequences. *Proceedings of the National Academy of Sciences USA*, *87*, 2414–2418.

Tichkule, S., Cacciò, S. M., Robinson, G., Chalmers, R. M., Mueller, I., Emery-Corbin, S. J., Eibach, D., Tyler, K. M., van Oosterhout, C., & Jex, A. R. (2022). Population genomics of *Cryptosporidium hominis* across five continents identifies two subspecies that have diverged and recombined during 500 years of evolution. *Molecular Biology and Evolution*, *39*(4), msac056. https://doi.org/10.1093/molbev/msac056

Tichkule, S., Jex, A. R., van Oosterhout, C., Sannella, A. R., Krumkamp, R., Adrich, C., Maiga-Ascofare, O., Dekker, D., Lamshöft, M., Mbwana, J., Rakotozandrindrainy, N., Borrmann, S., Thye, T., Schuldt, K., Winter, D., Kremsner, P. G., Oppong, K., Manouana, P., Mbong, M., … Cacciò, S. M. (2021). Comparative genomics revealed adaptive admixture in *Cryptosporidium hominis* in Africa. *Microbial Genomics*, *7*(1). https://doi.org/10.1099/mgen.0.000493

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1110), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Van Oosterhout, C. (2021). Mitigating the threat of emerging infectious diseases; a coevolutionary perspective. *Virulence*, *12*(1), 1288–1295.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouellier, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Ward, B. J., & van Oosterhout, C. (2016). HYBRIDCHECK: Software for the rapid detection, visualisation and dating of recombinant regions in genome sequence data. *Molecular and Ecological Resources*, *16*(2), 534–539.

Warrenfeltz, S., Kissinger, J. C., & EuPathDB Team. (2020). Accessing *cryptosporidium* omic and isolate data via CryptoDB.org. *Methods in Molecular Biology*, *2052*, 139–192.

Webster, J. P., Gower, C. M., Knowles, S. C. L., Molineux, D. H., & Fenton, A. (2016). One health - an ecological and evolutionary framework for tackling neglected zoonotic diseases. *Evolutionary Applications*, *9*(2), 313–333.

Zhang, Z., Schwartz, S., Wagner, L., Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*: 203–214.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Corsi, G. I., Tichkule, S., Sannella, A. R., Vatta, P., Asnicar, F., Segata, N., Jex, A. R., van Oosterhout, C., & Cacciò, S. M. (2023). Recent genetic exchanges and admixture shape the genome and population structure of the zoonotic pathogen *Cryptosporidium parvum*. *Molecular Ecology*, *32*, 2633–2645. https://doi.org/10.1111/mec.16556