Host Association of *Cryptosporidium parvum* Populations Infecting Domestic Ruminants in Spain

Joaquín Quílez,[a] Claudia Vergara-Castiblanco,[a] Luis Monteagudo,[b] Emilio del Cacho,[a] Caridad Sánchez-Acedo[a]

Department of Animal Pathology, Faculty of Veterinary Sciences, University of Zaragoza, Zaragoza, Spain[a]; Department of Anatomy, Embryology and Genetics, Faculty of Veterinary Sciences, University of Zaragoza, Zaragoza, Spain[b]

A stock of 148 *Cryptosporidium parvum* DNA extracts from lambs and goat kids selected from a previous study examining the occurrence of *Cryptosporidium* species and GP60 subtypes in diarrheic lambs and goat kids in northeastern Spain was further characterized by a multilocus fragment typing approach with six mini- and microsatellite loci. Various degrees of polymorphism were seen at all but the MS5 locus, although all markers exhibited two major alleles accounting for more than 75% of isolates. A total of 56 multilocus subtypes (MLTs) from lambs (48 MLTs) and goat kids (11 MLTs) were identified. Individual isolates with mixed MLTs were detected on more than 25% of the farms, but most MLTs (33) were distinctive for individual farms, revealing the endemicity of cryptosporidial infections on sheep and goat farms. Comparison with a previous study in calves in northern Spain using the same six-locus subtyping scheme showed the presence of host-associated alleles, differences in the identity of major alleles, and very little overlap in MLTs between *C. parvum* isolates from lambs and those from calves (1 MLT) or isolates from lambs and those from goat kids (3 MLTs). The Hunter-Gaston index of the multilocus technique was 0.976 (95% confidence interval [CI], 0.970 to 0.982), which supports its high discriminatory power for strain typing and epidemiological tracking. Population analyses revealed the presence of two host-associated subpopulations showing epidemic clonality among the *C. parvum* isolates infecting calves and lambs/goat kids, respectively, although evidence of genetic flow between the two subpopulations was also detected.

PCR-based methods have become essential tools for the proper identification of *Cryptosporidium* spp. in both human and animal hosts, given the limitations in the specific detection of this protozoan using microscopic techniques. Molecular methods were initially used for *Cryptosporidium* species differentiation by using genetic markers having relatively low intraspecific/high interspecific sequence variation, providing the basis for the current classification of members within the genus (1). More recently, a second generation of more discriminatory subtyping approaches using highly variable markers has enabled the study of intraspecies variation, leading to a better understanding of the population structure and transmission dynamics of *Cryptosporidium* spp. (2). Sequencing of the 60-kDa glycoprotein (GP60) gene has been used extensively in studies of the transmission of *Cryptosporidium hominis* and *Cryptosporidium parvum* and is currently the most widely subtyping method, but other techniques based on the characterization of short variable-number tandem-repeat (VNTR) loci, known as microsatellites and minisatellites, are being increasingly used in multilocus analyses. VNTR markers were identified in protozoan parasites as early as 1994, but their use as a typing tool for *Cryptosporidium* spp. is more recent and has mainly been restricted to isolates from humans and calves (1, 3).

Many contributions of molecular tools have focused on assessing the human-infective potential of *Cryptosporidium* species in animals, with *C. parvum* being the most common zoonotic species. Calves have been referred to as the only major reservoir for *C. parvum* infection in humans (4), but the potential role of other livestock species such as sheep and goats has received limited attention; in spite of this, the protozoan is one of the major enteropathogens associated with neonatal diarrheic outbreaks on small-ruminant farms, and these hosts represent an important sector of the global livestock population in many countries (5, 6). Two major *Cryptosporidium* species with zoonotic potential have been

identified in sheep and goats. *C. parvum* has been reported in diarrheic and asymptomatic lambs and goat kids (5, 7) as well as adult sheep and goats (8–10) from Europe, North America, Australia, and Africa. Other studies have reported that *Cryptosporidium ubiquitum* (formerly known as the *Cryptosporidium* cervine genotype) is much more prevalent that *C. parvum* in sheep (11–14).

The intraspecific variation of *C. parvum* from small ruminants is not well documented, since modest numbers of specimens have been genetically characterized by using either GP60 sequencing (7, 15) or VNTR markers (7, 16–22). A large-sample-size GP60 typing study previously conducted in Spain has shown that lambs and goat kids are mostly infected by *C. parvum* isolates belonging to the uncommon subtype family IId, compared to the major zoonotic family IIa reported as the most prevalent in calves, which suggests the uniqueness of cryptosporidial infection on small-ruminant farms (23). In the current study, the ovine and caprine *C. parvum* isolates analyzed in the latter contribution were further characterized using VNTR markers. For this purpose, we used a multilocus fragment typing approach with six mini- and microsatellites previously optimized for the analysis of genetic polymorphisms of *C. parvum* isolates from diarrheic preweaned calves in northern Spain (24). Multilocus subtype (MLT) data were compared with those reported in cattle and used to explore the population structure of *C. parvum* from domestic ruminants in Spain.

MATERIALS AND METHODS

Parasite isolates. A stock of 148 *C. parvum* DNA extracts selected from a previous study examining the occurrence of *Cryptosporidium* species and GP60 subtypes involved in the etiology of neonatal diarrhea in sheep and goat farms in Spain was used in this analysis (23). *Cryptosporidium* isolates were obtained from diarrheic preweaned lambs (*n* = 131) and goat kids (*n* = 17) randomly selected on 71 sheep and 7 goat farms (mean, 1.90 ± 1.08 animals/farm) from three provinces across Aragón (northeastern Spain). The size of the flocks ranged from 72 to 815 mature animals, and all farms were based on traditional extensive grazing systems, with shed lambing and breeding of their own female lambs or goat kids for replacement. Grazing areas were not shared with other livestock species. A single isolate from each of 33 farms and 2 to 6 isolates from each of the remaining 45 farms were included in the molecular analysis. Both the *Cryptosporidium* species and subtypes of all isolates were successfully identified in the previous study based on PCR-restriction fragment length polymorphism (PCR-RFLP) and sequence analyses of the small-subunit (SSU) rRNA and GP60 genes, respectively (25, 26).

PCR primers and conditions. Each isolate was genotyped at two minisatellite (MSB and MS5) and four microsatellite (ML1, ML2, TP14, and 5B12) loci, using primers and conditions described previously (24). Briefly, the satellite fragments were amplified by using simple (MSB, ML2, and 5B12), heminested (MS5 and ML1), and nested (TP14) PCRs. In order to allocate alleles with overlapping peaks to a specific locus, the reverse primers used in simple PCRs and the internal reverse primers used in heminested/nested PCRs were 5′ labeled with one of three spectrally separate fluorophores according to the predicted fragment size, including HEX (4,7,2′,4′,5′7′-hexachloro-6-carboxyfluorescein), 6-FAM (6-carboxyfluorescein), or TAMRA (6-carboxytetramethylrhodamine). Cycling conditions for simple PCRs included 40 cycles consisting of 94°C for 30 s, the marker-specific annealing temperature for 30 s, and 72°C for 1 min, with an initial denaturation step at 94°C for 5 min and a final extension at 72°C for 7 min. Heminested and nested PCRs were subjected to 35 cycles consisting of 95°C for 50 s, the marker-specific annealing temperature for 50 s, and 72°C for 1 min, with an initial denaturation step at 95°C for 3 min and a final extension at 72°C for 10 min.

Automated fragment analysis. PCR products were first separated by electrophoresis in 1.5% agarose gels and visualized by staining with GelRed nucleic acid gel stain (Biotium, Hayward, CA) to confirm DNA amplification. According to the amplicon intensity, between 0.5 and 2 μl of the satellite-labeled PCR products for each *C. parvum* isolate was mixed with 0.3 μl of Et400-R size standard (GE Healthcare, USA) and 8.5 μl of deionized formamide. This mixture was then denatured (95°C for 2 min) and run on a linear polyacrylamide long read matrix (GE Healthcare, USA) in a 40-cm-detection-length/75-μm-internal-diameter capillary for genetic analysis. Signals were read with an automatic sequencer (Mega-BACE500; GE Healthcare, USA), and the data were stored and analyzed with the aid of Fragment Profiler software (version 1.2). PCR product sizes were then translated into allele numbers. At least two representative isolates of each allele from each host species were analyzed by bidirectional DNA sequencing for length confirmation. Size discrepancies between fragment analysis and sequencing were identical to those previously reported in calves, and the same correction factors for each marker were established (24). All alleles shared by *C. parvum* isolates from lambs and goat kids provided identical sequences. Allele nomenclature was based on the fragment size in base pairs identified using automated fragment analysis adjusted after comparison with this reference sequenced material. Alleles were compared and correlatively numbered according to those identified within *Cryptosporidium* isolates from calves (24).

MLT identification. The multilocus subtype (MLT) for each isolate was defined by the combination of alleles at the six loci. Only isolates that were amplified at all six loci were included in the analysis. The presence of two separate peaks for a specific locus differing by multiples of the repeat unit was designated a mixed infection, and the two potential MLTs were considered. Multilocus subtypes for isolates showing several alleles at

more than one locus could not be determined, and these isolates were excluded from the genetic analysis. The discriminatory power of each individual marker and the multilocus approach for subtyping *C. parvum* isolates was assessed by calculating the Hunter-Gaston discriminatory index (HGDI), which estimates the probability of randomly picking two unrelated isolates and finding them to be different (27). The VNTR diversity and confidence extractor software (V-DICE) available at the Health Protection Agency bioinformatics tools website was used for this purpose (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl).

Data analysis. The relationships among the MLTs identified in lambs/goat kids as well as calves from the previous Spanish study were estimated using the eBURST algorithm (28) (http://eburst.mlst.net/). The most stringent setting was used, and only MLTs that differ from one another at one locus (single-locus variants [SLVs]) were assigned to the same cluster. Clusters of linked SLVs are referred to as clonal complexes. Each cluster is formed around a founder member, which has the highest number of SLVs and is considered to be the ancestral type. Data from both satellite markers and the previously reported GP60 subtypes were used to assess the genetic structure of *C. parvum* isolates from lambs/goat kids and calves using several population genetic programs. The standardized index of association (I_A^S) was calculated using LIAN v. 3.5 as a measure to determine allelic linkage disequilibrium (LD) among different loci (http://pubmlst.org/perl/mlstanalyse/mlstanalyse.pl?site=pubmlst&page=lian&referer=pubmlst.org) (29). Isolates were assigned to different genetic groups using the Structure software, which uses a Bayesian model-based clustering approach, estimating for each isolate the fraction of its genotype that belongs to each cluster on the basis of allele frequencies (30). For each *K* value (number of genetic groups), between 1 and 10,000 burn-in iterations followed by a run of 100,000 Markov chain Monte Carlo repetitions were performed and 3 iterations of each calculation were done to evaluate variability. A factorial correspondence analysis (FCA) implemented in the Genetix software was performed according to the instructions of the authors (K. Belkhir, P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme, Genetix 4.05 [http://www.genetix.univ-montp2.fr/genetix/genetix.htm]). This test places the isolates in a three-dimensional space according to the similarity of their allelic state. Genetix software was also used to assess genetic differentiation and genetic flow by F statistics, calculating the fixation index values (Fst) for the different origins of the samples (31). An unweighted pair group method analysis (UPGMA) based on the method of Nei (32) was completed by means of the Tools for Population Genetics Analysis (TFPGA) software (M. P. Miller [http://www.marksgeneticsoftware.net/tfpga.htm]).

Nucleotide sequence accession numbers. Unique nucleotide sequences generated in this study were deposited in the GenBank database under accession numbers JQ954680 to JQ954688.

RESULTS

Allele frequencies. A total of 140 *C. parvum* isolates from lambs (*n* = 123) and goat kids (*n* = 17) were typeable at all six loci. These isolates originated from 66 sheep and 7 goat farms, respectively. Table 1 summarizes the numbers and sizes of the alleles identified as well as the Hunter-Gaston index for each locus. The TP14, MSB, and ML2 loci provided similar discriminatory powers, with HGDI values ranging from 0.629 to 0.655. These values were lower than that reported for GP60 sequencing, which differentiated a total of 12 subtypes in a previous study with the same stock of isolates (HGDI, 0.775; 95% confidence interval [CI], 0.740 to 0.810) (23). A total of six alleles of sizes ranging from 191 to 227 bp were identified within the ML2 locus, although 75% of isolates were assigned to only two alleles, namely, ML2-191 and ML2-193. These alleles were identical to the *C. parvum* sequences under GenBank accession numbers AJ308566 and AJ308565, respectively (17), and differed by contractions of the AG repeats to the previously undescribed ML2 alleles of 197, 201, and 221 bp. Re-

TABLE 1 Allele sizes and number allocation for each of the six minisatellite and microsatellite loci in *C. parvum* isolates from diarrheic preweaned lambs and goat kids in northeastern Spain

Locus and allele size (bp) (allele no.)[a]	Lambs		Goat kids	
	No. (%) of isolates ($n = 123$)	No. of farms ($n = 66$)	No. (%) of isolates ($n = 17$)	No. of farms ($n = 7$)
ML1 (HGDI = 0.519 [0.445–0.593])[b]				
226 (1)	80 (65)	42	14 (82.3)	6
238 (2)	24 (19.5)	17		
223 (3)	13 (10.6)	7	2 (11.8)	1
241 (4)	2 (1.6)	1		
238 + 223	2 (1.6)	1		
238 + 241	2 (1.6)	1		
223 + 226			1 (5.9)	1
ML2 (HGDI = 0.655 [0.596–0.714])[b]				
191 (2)	65 (52.8)	34	7 (41.2)	6
227 (3)	3 (2.4)	2		
193 (9)	29 (23.6)	18	4 (23.5)	2
197 (10)	2 (1.6)	1		
201 (11)	7 (5.7)	2	5 (29.4)	1
221 (12)	17 (13.8)	10	1 (5.9)	1
TP14 (HGDI = 0.629 [0.593–0.665])[b]				
324 (1)	59 (47.9)	33	8 (47)	3
333 (2)	32 (26)	17	7 (41.2)	4
342 (3)	24 (19.5)	14	1 (5.9)	1
324 + 333	5 (4.1)	5		
324 + 342	3 (2.4)	3		
333 + 342			1 (5.9)	1
MS5 (HGDI = 0.025 [0.000–0.058])[b]				
215 (1)	1 (0.8)	1		
239 (2)	120 (97.6)	64	17 (100)	7
239 + 191 (3)	1 (0.8)	1		
239 + 215	1 (0.8)	1		
5B12 (HGDI = 0.499 [0.438–0.559])[b]				
167 (2)	37 (30.1)	25	3 (17.6)	3
169 (3)	77 (62.6)	39	13 (76.5)	6
171 (4)	7 (5.7)	4	1 (5.9)	1
167 + 169	1 (0.8)	1		
169 + 171	1 (0.8)	1		
MSB (HGDI = 0.641 [0.602–0.679])[b]				
304 (1)	52 (42.3)	31	2 (11.8)	2
322 (3)	14 (11.4)	7	1 (5.9)	1
328 (4)	3 (2.4)	2	3 (17.6)	2
310 (5)	50 (40.6)	26	10 (58.8)	5
310 + 304	1 (0.8)	1		
310 + 316 (2)	2 (1.6)	2		
310 + 322	1 (0.8)	1		
310 + 328			1 (5.9)	1

[a] Alleles were compared and correlatively numbered according to those identified within *Cryptosporidium* isolates from calves by Quílez et al. (24).
[b] Discriminatory power (95% confidence interval).

peated attempts to sequence allele ML2-227 were unsuccessful due to the presence of underlying signals in the electropherogram that prevented the accurate readout of sequences. A similar distribution was seen at the MSB minisatellite, with a total of five alleles identified within this marker but two predominant alleles exhibited by more than 80% of isolates, including MSB-304 and the novel allele MSB-310. Sequence analysis of selected isolates revealed that MSB-322 was identical to the *C. parvum* isolate from

calves under GenBank accession number JF342570, but alleles MSB-304 and MSB-328 differed by three and one nucleotide polymorphism within the repeat region to the homologous alleles from calves in Spain with accession numbers JF342568 and JF342571, respectively (24).

Only three alleles were identified at the TP14 locus, although they were evenly distributed within the *C. parvum* isolates, which explains the relatively high HDGI value for this marker. The alleles

TP14-324 and TP14-333 were identical to those previously reported in calves under accession numbers JF342561 and JF342562, but the allele TP14-342 differed by multiple nucleotide polymorphisms from the *C. parvum* reference sequence JF342563 from GenBank (24). Nine samples showed two alleles at this locus, which displayed the highest rate of mixed infections. Two major alleles with sizes of 226 bp and 169 bp accounted for more than 64% of isolates at the ML1 and 5B12 loci, respectively. Isolates selected for sequencing at the ML1 marker matched the C1 (ML1-238), C2 (ML1-226), and C3 (ML1-223) alleles deposited in GenBank (accession numbers AJ249582 to AJ249584) (16) and differed by contractions of the microsatellite repeats from the novel *C. parvum* allele ML1-241. At the 5B12 locus, sequence analyses revealed that fragments of 167, 169, and 171 bp were identical to the sequences JF342565 to JF342567 from calves in Spain (24). The MS5 locus displayed the poorest genetic diversity, since all but one isolate showed the MS5-239 allele, which demonstrated 100% identity to the bovine *C. parvum* reference sequence under GenBank accession number JF342572 (24). Repeated attempts to sequence alleles of 215 bp and 191 bp at the MS5 locus were unsuccessful. Comparison of these results with the distribution of satellite alleles previously reported in calves in a larger area of northern Spain, which included the geographic area analyzed in the current study, revealed the presence of host-associated alleles for *C. parvum* isolates from both lambs and goat kids (ML1-223, ML1-241, ML2-193, ML2-197, ML2-201, and ML2-221) and calves (ML2-185, ML2-229 to ML2-237, and 5B12-165). Differences were also seen in the identity of major alleles at the ML1 (226 bp versus 238 bp), MSB (310 bp and 304 bp versus 322 bp), ML2 (191 and 193 bp versus 231 and 233 bp), and TP14 (324 bp versus 333 bp) loci for isolates from ovine/caprine and cattle specimens, respectively (24).

Multilocus subtypes. The composition and frequency of MLTs identified based on the combination of alleles at all the six mini- and microsatellite loci are shown in Table 2. A total of 48 MLTs were identified within the *C. parvum* isolates from lambs, and 11 MLTs were identified within those from goat kids. Individual isolates with mixed MLTs were seen at 19/73 farms, including 18 isolates from lambs and three isolates from goat kids which showed two alleles at one locus and were allocated to the corresponding MLTs. One lamb isolate had a biallelic profile at two different loci and was excluded from the genetic analyses. A single MLT was identified for 27 of 45 farms where two or more isolates were collected. Most MLTs identified in both lambs (25/48) and goat kids (8/11) were farm specific, and comparison with MLTs from calves revealed that the majority of them were host associated. Namely, only three MLTs (L11, L14, and L21) were shared by lambs and goat kids, and only one MLT from lambs (L39) was also found in calves (C49). All these shared MLTs except for L21 also shared the GP60 subtype. The HGDI value of the six-satellite typing method was 0.976 (95% CI, 0.970 to 0.982), and the discriminatory power hardly increased when results of GP60 sequencing were added to the multilocus typing (0.980; 95% CI, 0.976 to 0.986).

Population analysis. When all the MLTs from ruminants were analyzed using the eBURST algorithm, a picture of two groups of related clonal complexes with a star-like topology was identified, with the MLT L6 connecting the two groups and no outlier in the network (Fig. 1). The first group included all but one of the MLTs of bovine origin, as well as some MLTs of ovine and caprine origin

exhibiting the minor allele MSB-322 (L6, L18, L19, L20, L37, L38, L48, and K2). The other group included all but the above-mentioned MLTs from lambs and goat kids as well as the MLT C2, which matched the only isolate from calves showing the allele MSB-304. A founder MLT exhibiting the most prevalent alleles at each marker for either calves or lambs/goat kids was identified in each group, i.e., C42 and L11, with bootstrap values of 79% and 94%, respectively, as well as related MLTs which could be considered subfounders of additional clonal lineages, such as C27 or L3 (bootstrap values of 83% and 82%, respectively). Values of the standardized index of association were all above zero, and the pairwise variance (V_D) was greater than the 95% critical value (L) when all isolates from ruminants were included in the analyses, revealing the presence of linkage disequilibrium (LD) among and within *C. parvum* subpopulations from the three host species. This finding indicates a predominant clonal genetic structure, although the low values of I_A^S suggest that reproduction is not exclusively clonal within the population. In fact, the I_A^S value did not differ from zero (*P* values > 0.05) when isolates from either calves or goat kids exhibiting the same MLT were scored as a single individual, which suggests an epidemic population structure. Interestingly, the I_A^S value was near zero but "statistically significant" for isolates from lambs, although the V_D and L values were similar (1.496 and 1.477, respectively), and LD was broken when isolates from both lambs and goat kids were analyzed as a whole (Table 3).

The Bayesian statistical approach identified $K = 2$ to be the best estimation of ancestral clusters for the populations studied, represented by different colors in Fig. 2A. An isolate was considered to belong to one of the clusters only if the probability of it belonging was higher than 0.8. Otherwise, the isolate was considered to have a mixed ancestry. Most isolates from calves (95.8%) and one isolate from a lamb belonged to one cluster. In contrast, all isolates from goat kids and most from lambs (98.3%) were assigned to the other cluster, which also included one isolate from a calf. A few isolates from both calves ($n = 4$) and lambs ($n = 1$) could not be allocated to only one population, which mirrors their mixed origin. The three-dimensional distribution of correspondence analysis (FCA) of the MLT data also showed the split between *C. parvum* isolates from calves and lambs/goat kids, with similar genotypic diversities within the two clusters, some isolates with no clear membership in any population, and a few contacts between cattle and sheep isolates (Fig. 2B). The same two major subpopulations were detected by the UPGMA (data not shown) and were supported by F statistics, with Fst values indicating strong genetic differentiation between *C. parvum* isolates from calves and lambs (0.32409) or calves and goat kids (0.36516). These values were highly significant since random permutation tests based on up to 10,000 reiterations provided 0.00% values superior to the ones obtained for calves and lambs/goat kids, while a notably lower Fst coefficient was obtained when lambs and goat kids were compared (0.03702). The effective number of migrants between populations per generation (Nm) was estimated from Fst values (31). Comparison between calves and lambs or calves and goat kids provided Nm values of 0.52 and 0.43, respectively. In contrast, a high Nm value (6.50) was obtained when lambs and goat kids were compared, as expected for groups more genetically related.

TABLE 2 MLTs of *C. parvum* isolates from lambs and goat kids based on the combination of alleles at six microsatellite and minisatellite loci and comparison with previously reported GP60 subtypes

Livestock type	MLT	Allele at locus[a]						No. of isolates[b]	No. of farms	GP60 subtype(s)
		ML1	MS5	TP14	MSB	5B12	ML2			
Lamb								122	65[c]	
	L1	1	1	2	1	3	10	2	1	IIdA22G1
	L2	1	2	1	1	2	2	4	3	IIdA17G1a
	L3	1	2	1	1	3	2	4	2	IIdA19G1
	L4	1	2	1	1	3	9	5	3	IIdA17G1a, IIdA14G1
	L5	1	2	1	1	3	12	1	1	IIdA18G1
	L6	1	2	1	3	3	2	1	1	IIdA17G1a
	L7	1	2	1	4	3	2	2	1	IIdA24G1
	L8	1	2	1	5	2	2	3	3	IIdA17G1b, IIdA17G1a
	L9	1	2	1	5	2	9	5	2	IIdA17G1b, IIdA17G1a
	L10	1	2	1	5	2	11	1	1	IIdA17G1a
	L11	1	2	1	5	3	2	9	3	IIdA17G1a
	L12	1	2	1	5	3	3	2	1	IIdA17G1b
	L13	1	2	1	5	3	9	1	1	IIdA19G1
	L14	1	2	1	5	3	11	6	2	IIdA17G1a
	L15	1	2	2	1	3	2	8	5	IIdA19G1
	L16	1	2	2	1	3	9	2	1	IIdA17G1b, IIdA18G1
	L17	1	2	2	1	3	10	1	1	IIdA22G1
	L18	1	2	2	3	2	2	1	1	IIdA19G1
	L19	1	2	2	3	2	12	1	1	IIdA17G1b
	L20	1	2	2	3	3	12	8	4	IIdA17G1b, IIdA17G1a
	L21	1	2	2	5	2	2	1	1	IIdA17G1b
	L22	1	2	2	5	3	2	4	2	IIdA17G1a
	L23	1	2	2	5	3	11	2	2	IIdA17G1a
	L24	1	2	3	1	2	2	2	2	IIdA17G1b, IIdA17G1a
	L25	1	2	3	1	3	2	1	1	IIdA17G1a
	L26	1	2	3	1	3	9	2	1	IIdA17G1a
	L27	1	2	3	1	3	12	3	3	IIdA18G1
	L28	1	2	3	1	4	12	7	4	IIdA18G1
	L29	2	2	1	1	2	9	5	3	IIdA17G1b
	L30	2	2	1	1	3	2	2	2	IIdA19G1
	L31	2	2	1	1	3	9	1	1	IIdA17G1a
	L32	2	2	1	2	2	9	1	1	IIdA17G1a
	L33	2	2	1	5	2	2	1	1	IIdA17G1b
	L34	2	2	1	5	2	9	4	3	IIdA17G1a
	L35	2	2	1	5	3	2	3	2	IIdA19G1
	L36	2	2	1	5	3	9	1	1	IIdA19G1
	L37	2	2	2	3	2	3	1	1	IIaA15G2R1
	L38	2	2	2	3	3	2	2	1	IIdA18G1
	L39	2	2	2	4	4	2	1	1	IIaA18G3R1
	L40	2	2	2	5	2	2	2	2	IIdA17G1b, IIdA18G1
	L41	2	2	3	1	2	2	5	3	IIdA19G1, IIdA17G1b
	L42	2	2	3	5	3	2	2	1	IIdA17G1a
	L43	3	2	1	5	3	2	6	3	IIdA19G1, IIdA15G1
	L44	3	2	1	5	3	9	2	2	IIdA17G1a, IIdA17G1b
	L45	3	2	3	1	2	2	4	2	IIdA19G1
	L46	3	2	3	5	2	2	3	1	IIdA19G1
	L47	3	2	3	5	3	9	1	1	IIdA17G1a
	L48	4	2	2	3	3	2	4	1	IIdA18G1
Goat kids								17	7	
	K1	1	2	1	1	2	11	1	1	IIdA17G1a
	K2	1	2	1	3	3	12	1	1	IIdA19G1
	L11	1	2	1	5	3	2	2	2	IIdA17G1a
	L14	1	2	1	5	3	11	4	1	IIdA17G1a
	K3	1	2	2	1	4	2	1	1	IIdA17G1a
	K4	1	2	2	4	3	9	4	2	IIdA26G1, IIdA19G1
	L21	1	2	2	5	2	2	2	2	IIdA19G1
	K5	1	2	2	5	3	9	1	1	IIdA26G1
	K6	3	2	2	1	4	2	1	1	IIdA17G1a
	K7	3	2	2	5	3	2	1	1	IIdA25G1
	K8	3	2	3	5	3	2	2	1	IIdA25G1

[a] The number allocation for alleles is indicated in Table 1.
[b] Samples with mixed infections at a single locus were allocated to the corresponding multilocus subtype.
[c] The single isolate from a sheep farm showed mixed infections at more than one locus and was excluded from the genetic analyses.
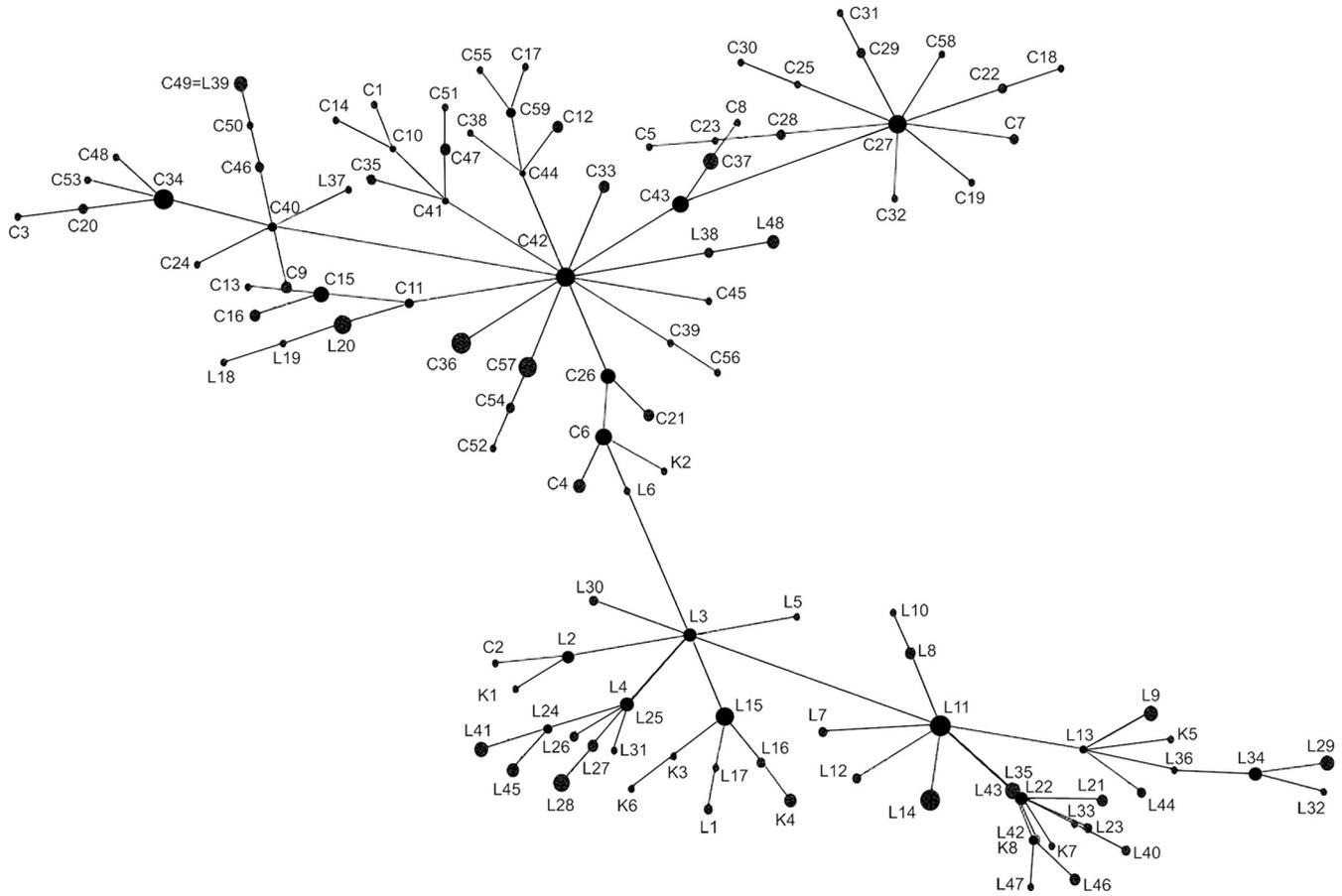
FIG 1 Single-locus variant eBURST network for multilocus subtypes (MLTs) of *Cryptosporidium parvum* isolates from calves (C; 59 MLTs), lambs (L; 48 MLTs), and goat kids (K; 11 MLTs) in Spain. Dots represent MLTs, with diameters proportional to numbers of isolates. Single-locus variants are joined by lines.

TABLE 3 Analysis of linkage disequilibrium in *C. parvum* population from ruminants in Spain

Hosts	Analysis	I_A^S	P value[b]	$V_D > L$
Calves	All	0.0289	<0.001	Yes
	Distinct MLT[a]	0.0064	0.253	No
Lambs	All	0.0521	<0.001	Yes
	Distinct MLT	0.0156	0.03	Yes
Goat kids	All	0.0832	<0.001	Yes
	Distinct MLT	0.0490	0.082	No
Calves + lambs	All	0.0868	<0.001	Yes
	Distinct MLT	0.0485	<0.001	Yes
Calves + goat kids	All	0.0752	<0.001	Yes
	Distinct MLT	0.0480	<0.001	Yes
Lambs + goat kids	All	0.0468	<0.001	Yes
	Distinct MLT	0.0044	0.266	No
All	All	0.0714	<0.001	Yes
	Distinct MLT	0.0452	<0.001	Yes

[a] Each group of isolates with identical MLTs was scored as a single individual.
[b] *P* values are based on a Monte Carlo test with 1,000 permutations.

DISCUSSION

During the last few years, contributions to molecular epidemiology of cryptosporidial infections in sheep and goats have mainly focused on the *Cryptosporidium* species differentiation, revealing the role of domestic small ruminants as a reservoir of *C. parvum* and other potentially zoonotic species (5). However, the intraspecies diversity of this protozoan in these hosts is not well documented, and limited studies have been performed with VNTR markers, restricting most comparisons to results obtained with *C. parvum* of human or bovine origin. Analysis of the genetic diversity through the panel of loci selected for the current study indicates that the MS5 locus was far too homogenous to be a useful marker for *C. parvum* typing in our geographical area, as previously reported in isolates from humans and livestock in different countries (19, 22, 33). In contrast, various degrees of polymorphism and discriminatory power were seen at the remaining loci, although all of them exhibited two major alleles accounting for more than 75% of isolates.

The major alleles at both the ML1 (226- and 238-bp) and ML2 (191- and 193-bp) loci have been described in humans and ruminants in different European countries, the United States, Japan, or Australia, although previously undescribed alleles were also identified at the last markers and the MSB locus (7, 16–19, 21, 22, 25, 34–36). In fact, the novel alleles at both the ML2 (197-, 201-, and 221-bp) and MSB (310-bp) loci were exhibited by more than 22%
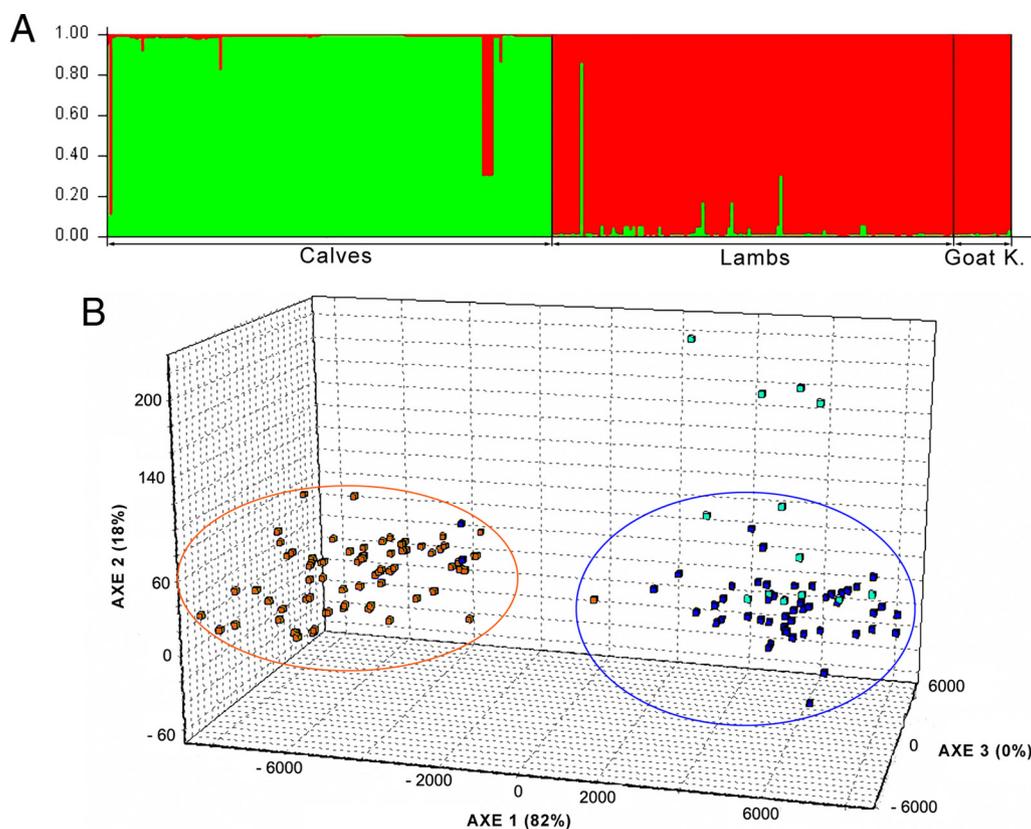
FIG 2 Clustering analyses of *C. parvum* isolates based on total multilocus information, including information on 6 mini- and microsatellite markers and GP60 subtypes. Altogether, 139 isolates from diarrheic preweaned lambs ($n = 122$) and goat kids ($n = 17$) as well as 131 isolates from diarrheic preweaned calves from a previous Spanish study were included. (A) Bayesian genetic structure analysis as inferred by Structure software. The bar plot shows the most probable number of clusters ($K = 2$). Each isolate is represented by a single vertical line. The green and red backgrounds represent one cluster, and the length of each line shows the isolate's estimated proportion of membership in that population. The host origin of the *C. parvum* isolates for each cluster is indicated at the bottom. (B) Factorial correspondence analysis (FCA) including isolates from calves (orange squares), lambs (blue squares), and goat kids (green squares).

and 42% of the samples, respectively, revealing the genetic distinctiveness of numerous isolates. Other studies have also reported the presence of geographically restricted alleles such as ML1-220, so far identified only in humans and ruminants in Italy and occasionally identified in the United Kingdom (7, 16, 17). It is also worth noting that none of the above-mentioned major alleles apart from ML1-238 were seen in lambs in a recent study in northwestern Spain, confirming that differences in the geographical distribution of *C. parvum* isolates could apply even to areas in close proximity (37). Comparison of the identity of alleles identified at the remaining loci and *C. parvum* alleles reported by other authors could not be done because of either the different methods used for sizing of PCR fragments or the use of different primer sets.

None of the six markers alone improved the discriminatory power previously reported by GP60 sequencing (23), but the multilocus analyses substantially increased the HGDI value and identified up to 56 MLTs within 140 isolates. This finding indicates an extensive genetic diversity, which has been related to regions where *Cryptosporidium* transmission is assumed to be intense (38) and was proportionally higher in goat kids (11 MLTs/17 isolates) than in lambs (48 MLTs/122 isolates). Other studies with different panels of markers have also revealed extensive polymorphism within *C. parvum* populations in France and Haiti (26 MLTs/61 isolates), the United Kingdom (31 MLTs/141 isolates), Scotland

(95 MLTs/297 samples), Italy (102 MLTs/173 samples), or the United States (94 MLTs/212 isolates), although all these studies included isolates from humans in addition to different livestock species (19, 20, 22, 39, 40).

The high level of individual isolates with mixed MLTs, which were seen on more than 25% of the farms, reflects the fact that odds of sexual recombination between genetically distinct isolates are common in this geographical area. Nevertheless, the finding of a single MLT in most farms (27/45) where several isolates were analyzed and the fact that most MLTs (33/56) were distinctive for individual farms reveal the endemicity of cryptosporidial infections on small-ruminant farms and the usefulness of this VNTR-based multilocus approach for strain typing and epidemiological tracking. Tanriverdi et al. (33) also reported that a majority of MLTs were limited to single cattle farms in Israel and Turkey, with a high proportion of mixed infections in individual isolates, especially in the latter country. Mixed infections were also commonly found within *C. parvum* from humans and ruminants in Scotland (10 to 37% of the isolates) and Italy (11.6%), with 7/12 farms in the latter country harboring a single and unique multilocus type (20, 39).

Probably one of the most interesting findings of this study is that a majority of isolates of *C. parvum* infecting domestic ruminants in our geographical area are host associated, which supports

the data previously reported by GP60 sequencing concluding that calves and lambs/goat kids are mostly infected by different allelic subtype families (IIa and IId, respectively) (23, 41). Comparison with the previous multilocus study in calves using the same six-locus subtyping scheme revealed the presence of host-associated alleles and even nucleotide polymorphisms in shared alleles between calves and lambs/goat kids (24). Namely, some alleles that were rare or absent in cattle at both the MSB (304- and 310-bp) and ML2 (from 191- to 221-bp) loci were responsible for most infections in small ruminants (81.4% and 97.8%, respectively). The allele ML1-226 accounted for 67.1% of infections in lambs and goat kids, while calves were mainly infected (71.5%) by isolates exhibiting the allele ML1-238. Similarly, a fragment of 165 bp reported as the second most common allele at the 5B12 marker in calves was not seen in either lambs or goat kids. Evidence of host association was further supported by findings of the multilocus analysis, since very little overlap in MLTs was seen between lambs and calves (1 MLT) or lambs and goat kids (3 MLTs).

The presence of host-associated *C. parvum* populations has been documented by GP60 typing, with some widespread allelic subtype families such as IIc being so far found only in humans and related to anthroponotic transmission (2). Multilocus studies of isolates from humans and livestock with VNTR markers have also identified *C. parvum* groups apparently restricted to humans in the United Kingdom, France, or Haiti, which supports the anthroponotic characteristics of these isolates or the occurrence of cycles that do not involve livestock (20, 22, 34). In contrast, no apparent host association was seen in the upper Midwest of the United States, with *C. parvum* being transmitted freely between cattle and humans (40). Data on the existence of specific *C. parvum* subpopulations within livestock species are limited. Mallon et al. (21) found no evidence of host specificity with regard to the parasites infecting sheep and cattle in Scotland, but a low number of ovine isolates were analyzed. However, a clear host separation was seen in Italy, with MLTs from goats found to differ from those of bovine and ovine origin (39). In the current study, only three MLTs were shared by lambs and goat kids, but additional isolates of caprine origin should be analyzed to confirm the distinctiveness between isolates infecting the two hosts. It is also worth mentioning that all the above-mentioned multilocus studies included the gp15/45/60 locus, reported as the single most polymorphic marker identified so far in the *Cryptosporidium* genome (2), which supports the power of VNTR markers other than the GP60 glycoprotein to discriminate *C. parvum* subpopulations.

Ecological factors have been reported to influence the population structure of *C. parvum*, with propagation pathways ranging from panmictic to clonal depending on local and host-related determinants (42). A complex picture was seen in Scotland, with panmictic, epidemic, and clonal subpopulations among *C. parvum* isolates from humans and livestock (20). In France, the *C. parvum* structure ranged from basic clonality (humans) to epidemic clonality (livestock) (22), while a predominantly clonal structure was seen in Italy (39). In contrast, the *C. parvum* population in humans and cattle was predominantly panmictic in the upper Midwest of the United States, with limited geographic or host substructuring (40). Population analysis in the current study confirmed the evolutionary divergence of *C. parvum* isolates infecting domestic ruminants in our geographical area. Linkage analyses showed linkage disequilibrium among and within *C. parvum* subpopulations from the three host species, revealing an overall clonal genetic structure. Nevertheless, the I_A^S values did not differ from zero when each of the MLTs from either calves or lambs/goat kids was treated as a unit, revealing that overrepresentation of some MLTs could have masked genetic exchange, and LD observed within both groups of hosts could arise from the clonal expansion of one or more MLTs to produce epidemic clones, i.e., an epidemic population structure (20, 43).

The eBURST network topology also showed a predominantly clonal population structure indicative of two populations, since two main groups of connected clonal clusters, including MLTs from either calves or lambs/goat kids, were identified, with some founder MLTs which could be responsible for the epidemic structure (28). Nevertheless, the finding of several MLTs from lambs/goat kids or calves in the main cluster of bovine or ovine/caprine origin, respectively, as well as the presence of an MLT exhibiting an SLV which connected the two clusters, indicates that the two subpopulations are not completely separated and that cross-infections and subsequent recombinations are possible, with the minor alleles at the MSB locus for each host being a consistent indicator of this cross-infectivity. Both the Bayesian and vectorial analyses also detected a considerable genetic structure, and Fst values indicated strong genetic differentiation between the two *C. parvum* subpopulations, although traces of genetic flow were seen using the Bayesian algorithm, a finding which is consistent with the above-mentioned observation that genetic exchange between the two subpopulations occurs rarely but should not be excluded.

To the best of our knowledge, this is the first large study on the molecular characterization of *C. parvum* from lambs and goat kids by using microsatellite and minisatellite loci. The results reveal the high discriminatory power of the multilocus approach for the epidemiological investigation of outbreaks and confirm the uniqueness of cryptosporidial infections on small-ruminant farms. Our findings also demonstrate the host substructuring of *C. parvum* in livestock in northern Spain, which might be a result of husbandry practices limiting transmission between large- and small-ruminant farms rather than a strict host specificity, as revealed by the genetic flow between the two subpopulations.

ACKNOWLEDGMENTS

REFERENCES

1. **Jex AR, Smith HV, Monis PT, Campbell BE, Gasser RB.** 2008. *Cryptosporidium*—biotechnological advances in the detection, diagnosis and analysis of genetic variation. Biotechnol. Adv. **26:**304–317.
2. **Xiao L.** 2010. Molecular epidemiology of cryptosporidiosis: an update. Exp. Parasitol. **124:**80–89.
3. **Traub RJ, Monis PT, Robertson LD.** 2005. Molecular epidemiology: a multidisciplinary approach to understanding parasitic zoonoses. Int. J. Parasitol. **35:**1295–1307.
4. **Xiao L, Feng Y.** 2008. Zoonotic cryptosporidiosis. FEMS Immunol. Med. Microbiol. **52:**309–323.
5. **Robertson LJ.** 2009. *Giardia* and *Cryptosporidium* infections in sheep and goats: a review of the potential for transmission to humans via environmental contamination. Epidemiol. Infect. **137:**913–921.
6. **Santín M, Trout JM.** 2008. Livestock, p 451–483. *In* Fayer R, Xiao L (ed), *Cryptosporidium* and cryptosporidiosis. CRC Press, Boca Raton, FL.
7. **Chalmers RM, Ferguson C, Cacciò S, Gasser RB, Abs El-Osta YG, Heijnen L, Xiao L, Elwin K, Hadfield S, Sinclair M, Stevens M.** 2005. Direct comparison of selected methods for genetic categorisation of *Cryp-*

tosporidium parvum* and *Cryptosporidium hominis* species. Int. J. Parasitol. **35**:397–410.

8. **Hajdušek O, Ditrich O, Šlapeta J.** 2004. Molecular identification of *Cryptosporidium* spp. in animal and human hosts from the Czech Republic. Vet. Parasitol. **122**:183–192.

9. **Morgan UM, Sargent KD, Deplazes P, Forbes DA, Spano F, Hertzberg H, Elliot A, Thompson RCA.** 1998. Molecular characterization of *Cryptosporidium* from various hosts. Parasitology **117**:31–37.

10. **O'Brien E, McInnes L, Ryan U.** 2008. *Cryptosporidium* GP60 genotypes from humans and domesticated animals in Australia, North America and Europe. Exp. Parasitol. **118**:118–121.

11. **Elwin K, Chalmers RM.** 2008. Contemporary identification of previously reported novel *Cryptosporidium* isolates reveals *Cryptosporidium bovis* and the cervine genotype in sheep (*Ovis aries*). Parasit. Res. **102**:1103–1105.

12. **Ryan UM, Bath C, Robertson I, Read C, Elliot A, McInnes L, Traub R, Besier B.** 2005. Sheep may not be an important zoonotic reservoir for *Cryptosporidium* and *Giardia* parasites. Appl. Environ. Microbiol. **71**:4992–4997.

13. **Santín M, Trout JM, Fayer R.** 2007. Prevalence and molecular characterization of *Cryptosporidium* and *Giardia* species and genotypes in sheep in Maryland. Vet. Parasitol. **146**:17–24.

14. **Wang Y, Feng Y, Cui B, Jian F, Ning C, Wang R, Zhang L, Xiao L.** 2010. Cervine genotype is the major *Cryptosporidium* genotype in sheep in China. Parasitol. Res. **106**:341–347.

15. **Alves M, Xiao L, Antunes F, Matos O.** 2006. Distribution of *Cryptosporidium* subtypes in humans and domestic and wild ruminants in Portugal. Parasitol. Res. **99**:287–292.

16. **Cacciò S, Homan W, Camilli R, Traldi G, Kortbeek T, Pozio F.** 2000. A microsatellite marker reveals population heterogeneity within humans and animal genotypes of *Cryptosporidium parvum*. Parasitology **120**:237–244.

17. **Cacciò S, Spano F, Pozio F.** 2001. Large sequence variation at two microsatellite loci among zoonotic (genotype C) isolates of *Cryptosporidium parvum*. Int. J. Parasitol. **31**:1082–1086.

18. **Enemark HL, Ahrens P, Juel CD, Petersen E, Petersen RF, Andersen JS, Lind P, Thamsborg SM.** 2002. Molecular characterization of Danish *Cryptosporidium parvum* isolates. Parasitology **125**:331–341.

19. **Leoni F, Mallon ME, Smith HV, Tait A, McLauchlin J.** 2007. Multilocus analysis of *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates from sporadic and outbreak-related human cases and *C. parvum* isolates from sporadic livestock cases in the United Kingdom. J. Clin. Microbiol. **45**:3286–3294.

20. **Morrison LJ, Mallon ME, Smith HV, MacLeod A, Xiao L, Tait A.** 2008. The population structure of the *Cryptosporidium parvum* population in Scotland: a complex picture. Infect. Genet. Evol. **8**:121–129.

21. **Mallon M, MacLeod A, Wastling JM, Smith H, Tait A.** 2003. Multilocus genotyping of *Cryptosporidium parvum* type 2: population genetics and sub-structuring. Infect. Genet. Evol. **3**:207–218.

22. **Ngouanesavanh T, Guyot K, Certad G, Le Fichoux I, Chartier C, Verdier RI, Cailliez JC, Camus D, Dei-Cas E, Bañuls AL.** 2006. *Cryptosporidium* population genetics: evidence of clonality in isolates from France and Haiti. J. Eukaryot. Microbiol. **53**(Suppl 1):S33–S36.

23. **Quílez J, Torres E, Chalmers RM, Hadfield SJ, Del Cacho E, Sánchez-Acedo C.** 2008. *Cryptosporidium* genotypes and subtypes in lambs and goat kids in Spain. Appl. Environ. Microbiol. **74**:6026–6031.

24. **Quílez J, Vergara-Castiblanco C, Monteagudo L, Del Cacho E, Sánchez-Acedo C.** 2011. Multilocus fragment typing and genetic structure of *Cryptosporidium parvum* isolates from diarrheic preweaned calves in Spain. Appl. Environ. Microbiol. **77**:7779–7786.

25. **Alves M, Xiao L, Sulaiman I, Lal AA, Matos O, Antunes F.** 2003. Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. J. Clin. Microbiol. **41**:2744–2747.

26. **Xiao L, Singh A, Limor J, Graczyk TK, Gradus S, Lal A.** 2001. Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater. Appl. Environ. Microbiol. **67**:1097–1101.

27. **Hunter PR, Gaston MA.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26**:2465–2466.

28. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**:1518–1530.

29. **Haubold B, Hudson RR.** 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics **16**:847–848.

30. **Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

31. **Cockerham CC, Weir BS.** 1993. Estimation of gene flow from F-statistics. Evolution **47**:855–863.

32. **Nei M.** 1972. Genetic distance between populations. Am. Nat. **106**:283–292.

33. **Tanriverdi S, Markovics A, Özkan A, Itik A, Shkap V, Widmer G.** 2006. Emergence of distinct genotypes of *Cryptosporidium parvum* in structured host populations. Appl. Environ. Microbiol. **72**:2507–2513.

34. **Mallon M, MacLeod A, Wastling J, Smith H, Reilly B, Tait A.** 2003. Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*. J. Mol. Evol. **56**:407–417.

35. **Wielinga PR, de Vries A, van der Goot TH, Mank T, Mars MH, Kortbeek LM, van der Giessen JWB.** 2008. Molecular epidemiology of *Cryptosporidium* in humans and cattle in The Netherlands. Int. J. Parasitol. **38**:809–817.

36. **Zintl A, Ezzaty-Mirashemi M, Chalmers RM, Elwin K, Mulcahy G, Lucy FE, de Waal T.** 2011. Longitudinal and spatial distribution of GP60 subtypes in human cryptosporidiosis cases in Ireland. Epidemiol. Infect. **139**:1945–1955.

37. **Díaz P, Hadfield SJ, Quílez J, Soilán M, López C, Panadero R, Díez-Baños P, Morrondo P, Chalmers RM.** 2012. Assessment of three methods for multilocus fragment typing of *Cryptosporidium parvum* from domestic ruminants in north west Spain. Vet. Parasitol. **186**:188–195.

38. **Widmer G, Sullivan S.** 2012. Genomics and population biology of *Cryptosporidium* species. Parasite Immunol. **34**:61–71.

39. **Drumo R, Widmer G, Morrison LJ, Tait A, Grelloni V, D'Avino N, Pozio E, Cacciò SM.** 2012. Evidence of host-associated populations of *Cryptosporidium parvum* in Italy. Appl. Environ. Microbiol. **78**:3523–3529.

40. **Herges GR, Widmer G, Clark ME, Khan E, Giddings CW, Brewer M, McEvoy JM.** 2012. Evidence that *Cryptosporidium parvum* populations are panmictic and unstructured in the upper Midwest of the United States. Appl. Environ. Microbiol. **78**:8096–8101.

41. **Quílez J, Torres E, Chalmers RM, Robinson G, Del Cacho E, Sánchez-Acedo C.** 2008. *Cryptosporidium* species and subtype analysis from dairy calves in Spain. Parasitology **135**:1613–1620.

42. **Tanriverdi S, Grinberg A, Chalmers RM, Hunter PR, Petrovic Z, Akiyoshi DE, London E, Zhang L, Tzipori S, Tumwine JK, Widmer G.** 2008. Inferences about the global population structure of *Cryptosporidium parvum* and *Cryptosporidium hominis*. Appl. Environ. Microbiol. **74**:7227–7234.

43. **Smith JM, Smith NH, O'Rourke M, Spratt BG.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. U. S. A. **90**:4384–4388.