Infection, Genetics and Evolution 31 (2015) 245–249

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Geographical segregation of *Cryptosporidium parvum* multilocus genotypes in Europe

Simone M. Cacciò [a,*], Valerie de Waele [b], Giovanni Widmer [c]

[a] Istituto Superiore di Sanità, Department of Infectious, Parasitic and Immunomediated Diseases, Viale Regina Elena 299, Rome 00161, Italy
[b] Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerpen, Belgium
[c] Cummings School of Veterinary Medicine at Tufts University, Department of Infectious Disease and Global Health, North Grafton, MA 01536, United States

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form 22 December 2014
Accepted 6 February 2015
Available online 14 February 2015

Keywords:
Cryptosporidium parvum
Multilocus genotypes
Population structure
Principal coordinate analysis
Linkage disequilibrium

ABSTRACT

Cryptosporidium parvum is a common enteric protozoan pathogen of humans and livestock. Multilocus genotyping based on simple sequence repeat polymorphisms has been used extensively to identify transmission cycles and to investigate the structure of *C. parvum* populations and of the related pathogen *Cryptosporidium hominis*. Using such methods, the zoonotic transmission of *C. parvum* has been shown to be epidemiologically important. Because different genetic markers have been used in different surveys, the comparison of *Cryptosporidium* genotypes across different laboratories is often not feasible. Therefore, few comparisons of *Cryptosporidium* populations across wide geographical areas have been published and our understanding of the epidemiology of cryptosporidiosis is fragmented. Here we report on the genotypic analysis of a large collection of 692 *C. parvum* isolates originating primarily from cattle and other ruminants from Italy, Ireland and Scotland. Because the same genotypic markers were used in these surveys, it was possible to merge the data. We found significant geographical segregation and a correlation between genetic and geographic distance, consistent with a model of isolation by distance. The presence of strong LD and positive I_A^S values in the combined MLG dataset suggest departure from panmixia, with different population structures of the parasite prevailing in each country.

1. Introduction

Species within the genus *Cryptosporidium* are obligate intracellular parasites that multiply in the epithelial cells of the gastrointestinal tract of vertebrates. These pathogens have a complex life cycle that comprises two generations of asexual multiplication followed by a sexual cycle (Tzipori, 1988). In humans, *Cryptosporidium hominis* and *Cryptosporidium parvum* account for the vast majority of cases worldwide (Cacciò and Putignani, 2014; Widmer and Sullivan, 2012; Xiao, 2010). *C. hominis* is considered a human parasite, albeit experimental and natural infection of animals with this species have occasionally been reported (Connelly et al., 2013; Giles et al., 2009; Ryan et al., 2005; Smith et al., 2005; Tanriverdi et al., 2003). *C. parvum* is a common parasite of ruminants, particularly of young animals, and has an established zoonotic potential (Learmonth et al., 2004; Sopwith et al., 2005).

Infection is initiated by the ingestion of oocysts. Direct human-to-human or animal-to-human contact, or ingestion of contaminated water and food are common routes of transmission. The oocysts are remarkably resistant to environmental stress and can withstand chlorination of drinking water. These properties, coupled with a low infectious dose (Chappell et al., 2006; Chappell et al., 1996), explain the large number of waterborne outbreaks, including the largest ever recorded outbreak in Milwaukee in the United States (Mac Kenzie et al., 1994).

Understanding how natural parasite populations are structured and how population structures can vary in relation to ecological and epidemiological conditions has attracted a considerable interest, due to the implication such studies can have on the development of control measures. Indeed, as highly polymorphic genetic markers (mini- and micro-satellites) were identified (Cacciò et al., 2000; Cacciò et al., 2001; Feng et al., 2000; Mallon et al., 2003a), population genetics studies were initiated in different countries to evaluate genetic diversity among parasite isolates, to estimate the occurrence of mixed infections (Tanriverdi et al., 2003; Widmer et al., 2014) and to identify host-associated subpopulations (Drumo et al., 2012) and other population structures (Feng et al., 2014). These studies have shown the existence of human-adapted *C. parvum* multi-locus genotypes (MLGs) (Leav et al., 2002; Mallon et al., 2003b), or have demonstrated the influence of different husbandry practices on the structure of bovine

* Corresponding author. Tel.: +39 06 4990 3016; fax: +39 06 4990 3561.
E-mail address: simone.caccio@iss.it (S.M. Cacciò).

http://dx.doi.org/10.1016/j.meegid.2015.02.008
1567-1348/

C. parvum populations (Tanriverdi et al., 2006). Most studies have focused on isolates collected from single countries (Gatei et al., 2007; Hunter et al., 2007) or individual host populations (Cama et al., 2008). Progress on an informative comparison across studies has been hampered by the lack of a standardized genotyping scheme. Consequently, meta-analyses encompassing multiple surveys have been rare (Wang et al., 2014).

To investigate the structure of *C. parvum* on a large geographical scale, we combined MLGs from three published surveys of *C. parvum* isolates from Scotland, Italy and Ireland (Morrison et al., 2008; Drumo et al., 2012; De Waele et al., 2013). This approach enabled us to assemble a large and geographically diverse MLG dataset comprising 692 *C. parvum* isolates. By applying different statistical tests, we found evidence of geographical segregation and linkage disequilibrium, and observed a significant correlation between genetic and geographical distance.

2. Materials and methods

2.1. Source of data

MLG data were taken from three published studies (Morrison et al., 2008; Drumo et al., 2012; De Waele et al., 2013). The data from Scotland (Morrison et al., 2008) included isolates from cattle ($n = 212$), lambs ($n = 9$) and humans ($n = 64$). The data from Italy comprised isolates from cattle ($n = 122$), lambs ($n = 21$), goat kids ($n = 21$), and humans ($n = 9$). The data from Ireland comprised 234 isolates all from cattle (De Waele et al., 2013). The seven genetic markers included were MS1, GP15, MS9, TP14, MM5, MM18, and MM19. The MS1 marker contains a GGTGGTATGCCA repeat in the heat shock protein 70 gene (cgd2_20) located at positions 3136–5184 on chromosome 2. The GP15 marker contains a TCA repeated motif in a 975-bp gene (cgd6_1080) encoding a sporozoite surface protein located at positions 266,434–267,408 on chromosome 6. The MS9 marker contains a TGGACT repeat in a 2016-bp gene (cgd5_2850) encoding a hypothetical protein located at positions 640,137–642,152 on chromosome 5. The TP14 marker contains a CAA repeat in an 8421-bp gene (cgd8_1340) encoding a hypothetical protein located at positions 365,790–374,210 on chromosome 8. The MM5 marker contains a TCCTCCTCT repeat located in an 11,418-bp gene (cgd6_4290) located at positions 1002,285–1013,702 on chromosome 6. The MM18 marker contains a GGACCA repeat in the 5004-bp gene (cgd8_660) located at positions 165,295–170,298 on chromosome 8. The MM19 marker contains a GGAGCT repeat in the 7230-bp gene (cgd8_4840) located at position 1208,520–1215,749 on chromosome 8.

Since Morrison et al. (2008) used a different reverse PCR primer for the GP15 locus than that used by Drumo et al. (2012) and De Waele et al. (2013), GP15 alleles from Scotland were recoded by subtracting 66 nucleotides, based on the fact that the 5′ end of the two GP15 reverse primers are 66 nucleotides apart (see Supplementary Table S1 for the size of recoded GP15 alleles).

2.2. MS9 genotype of bovine isolates from Ireland

MS9 alleles were determined and added to the original dataset from Ireland. This marker was PCR amplified for this study from 234 of 245 bovine isolates from Ireland as described (Mallon et al., 2003a). The size of each PCR product was estimated by electrophoresis using a capillary apparatus (QiaXcel; Qiagen, Milan, Italy) by comparison to size standards. Samples representative of each allele were sequenced on both strands to confirm the estimated size. Each distinct allele was assigned a unique number indicating the estimated size in nucleotides.

2.3. MLG data analyses

The program LIAN 3.6 (Haubold and Hudson, 2000) (available at http://adenine.biz.fh-weihenstephan.de/cgi-bin/lian/lian.cgi.pl) was used to calculate the standardized index of association (I_A^S). This index, a derivation of the Maynard-Smith index of association (Maynard-Smith et al., 1993), measures the strength of the linkage disequilibrium (LD) and is independent of the number of loci analyzed. The null hypothesis of linkage equilibrium was tested by Monte Carlo simulation (10.000 iterations).

The MLG for each isolate was defined by combining the alleles from the seven genetic loci (Drumo et al., 2012). The eBURST software (http://eburst.mlst.net/default.asp) was used to visualize the structure of the combined *C. parvum* population from the three countries. Using this method, the clonal nature of related genotypes and putative "founder" genotypes were visualized (Feil et al., 2004). The most stringent setting was used, and only single-locus variants (SLVs) which differ at one locus only were assigned to the same cluster.

Principal Coordinate Analysis (PCoA) was applied to graphically display pairwise genetic distances between MLGs. A matrix of pairwise distances between MLGs was computed using the SSR distance metric. The SSR distance between two MLGs was calculated as the squared difference in estimated amplicon length summed over seven loci. This distance matrix was input into the PCoA calculator in GenAlex (Peakall and Smouse, 2012).

A Mantel test (Mantel, 1967) was used to assess whether genetic and geographical distances among isolates were correlated. Geographical distance was estimated using the latitude and longitude of the capital of the province or county from which the isolates originated. MLGs from 619 isolates obtained from all ruminants or isolated from calves only ($n = 568$) were included. The computations were performed with GenAlex.

3. Results

3.1. MLG analysis

To create a MLG dataset based on the same seven markers, the alleles at the MS9 locus were determined for 234 of the 245 originally studied Irish *C. parvum* isolates. The analysis of MS9 amplicons revealed that one allele of 450 bp in length predominated (231 of 234 isolates), whereas 3 other alleles (432 bp, 444 bp and 456 bp) were each found in a single isolate. Combining the MS9 locus data with six additional markers resulted in 75 different MLGs in Ireland, to which 102 unique MLGs from Italy and 89 unique MLGs from Scotland were added for the analyses. The complete list of alleles identified at each locus in each country is provided (Supplementary Table S1).

Most MLGs were found in one country only. Only 7 of 266 unique MLGs were found in *C. parvum* isolated from Italy and Ireland and were designated with code "EU" in Fig. 1. Specifically, MLG IT59 (found in 2 Italian isolates) was identical to IE02 (found in 10 Irish isolates, EU1), MLG IT60 (found in 1 Italian isolate) was identical to IE06 (found in 3 Irish isolates, EU2), MLG IT64 (found in 3 Italian isolates) was identical to IE07 (found in 6 Irish isolates, EU3), MLG IT72 (found in 5 Italian isolates) was identical to IE10 (found in 1 Irish isolate, EU4), MLG IT67 (found in 9 Italian isolates) was identical to IE11 (found in 7 Irish isolates, EU5), MLG IT88 (found in 1 Italian isolate) was identical to IE22 (found in 1 Irish isolate, EU6), and MLG IT97 (found in 1 Italian isolate) was identical to IE46 (found in 25 Irish isolates, EU7). Of note, the shared MLGs were all of bovine origin (the Italian panel also comprises 21 goats and 21 sheep samples). No MLGs were shared between Italy and Scotland, or between Ireland and Scotland (Fig. 1).