Meta-analysis of a polymorphic surface glycoprotein of the parasitic protozoa *Cryptosporidium parvum* and *Cryptosporidium hominis*

G. WIDMER*

Tufts Cummings School of Veterinary Medicine, Division of Infectious Diseases, North Grafton, MA, USA

SUMMARY

Due to its extensive polymorphism, a partial sequence of the *Cryptosporidium* surface glycoprotein gene gp60 has been frequently used as a genetic marker. I explored the global diversity of this protein, and compared its sequence diversity in *Cryptosporidium parvum* and *Cryptosporidium hominis*. In marked contrast to the geographical partition of *C. parvum* and *C. hominis* multi-locus genotypes, gp60 allelic groups showed no evidence of segregating in space, or of differing with respect to geographical diversity. Globally, genetic diversity of *C. hominis* gp60 exceeded that of *C. parvum*. Within *C. parvum*, gp60 alleles originating from human isolates were more diverse than those infecting ruminants. Phylogenetic analysis grouped gp60 sequences into a small number of relatively homogenous allelic groups, with only a small number of alleles having evolved independently. With the notable exception of a group of alleles restricted to humans, *C. parvum* alleles are found in ruminants and humans.

Key words: Cryptosporidiosis, *Cryptosporidium parvum*, *Cryptosporidium hominis*, gp60, gp40/15.

INTRODUCTION

Cryptosporidium parvum and *Cryptosporidium hominis* are related species of apicomplexan protozoa causing cryptosporidiosis, an enteric infection of humans and animals. *C. parvum* is considered a zoonotic pathogen, as it is often acquired from ruminants by faecal–oral transmission of environmentally resistant oocysts. In contrast, the host range of *C. hominis* is thought to be restricted to humans.

The completion of the *C. parvum* and *C. hominis* genome [1, 2] has facilitated the discovery of numerous genetic polymorphisms, which have been used as genetic markers for characterizing routes of transmission and parasite populations [3–6]. Since its first description in 2000, the merozoite/sporozoite surface protein gp60 [7, 8] has been uniquely popular as a tool for genotyping *C. parvum* and *C. hominis* isolates. The focus on a fragment of this gene, which also includes a polymeric tract of serine residues, has generated a large collection of partial gp60 sequences. To facilitate the interpretation of this sequence information, a gp60 coding system based on the originally proposed allele codes [7] has been widely adopted. As more gp60 sequences were discovered, the original codes were extended to indicate the number of serine residues in the repeat and silent single nucleotide polymorphisms commonly present in this repeat [9]. The extensively used gp60 genotyping system has led some investigators to rely primarily on this sequence to genotype *C. parvum* and *C. hominis* isolates, and to define what is frequently referred to as 'subtypes'. This approach is intuitively appealing, because it reduces the genetic complexity of these species to an

* Author for correspondence: Dr G. Widmer, Tufts Cummings School of Veterinary Medicine, Building 20, 200 Westboro Road, North Grafton, MA, USA.
(Email: giovanni.widmer@tufts.edu).

unambiguous typing method enabling easy comparison of genotypes from different surveys. What is often overlooked is that this approach is incompatible with the obligatory sexual phase in the *Cryptosporidium* life-cycle and the possibility that meiosis results in genetic recombination between dissimilar genotypes. Genetic recombination was confirmed experimentally [10, 11], and inferred from the analysis of multi-locus genotypes (MLGs) of natural parasite populations [3, 5, 12]. Isolates which are grouped into a 'subtype' based on them sharing the same gp60 genotype, may thus differ at other loci, and could in theory be genetically more distinct than two isolates with different gp60 alleles. Genotypes from multiple loci, including gp60, may be advantageous in defining sub-specific populations and predicting transmission cycles.

The publication of numerous gp60 sequences has been driven by individual surveys where specific locations were intensely sampled [5, 9, 13–27]. Although genotype information is by no means random in space, the relative large number of countries from which gp60 sequences are available supports a relatively unbiased global analysis of gp60 polymorphism in *C. parvum* and *C. hominis*, and a comparison of gp60 diversity in these species. An analysis of gp60 richness in *C. parvum* and *C. hominis* is reported, and evidence for the lack of geographical structuring of gp60 polymorphisms within these species is presented. These results are discussed in the context of recent evidence of geographical structuring of *C. parvum* and *C. hominis* populations [12].

METHODS

Amino-acid sequences

C. parvum and *C. hominis* gp60 amino-acid sequences were downloaded (November 2008) from the Entrez Protein Database at the National Center for Biotechnology Information (NCBI). The search terms included '*Cryptosporidium parvum*' or '*Cryptosporidium hominis*' together with gp60 or gp40/15. Records were individually inspected to ensure that the species designation was consistent with the current taxonomy of the genus *Cryptosporidium*. This was particularly necessary for older records deposited up to 2002, as such entries predate the naming of *C. hominis* [28]. If not amended, such sequences may still be identified as *C. parvum* type 1, instead of *C. hominis* [22]. The gp60 amino-acid sequence was downloaded together with the geographical origin and the host species from which the isolate originated. The geographical origin was mostly defined at the level of country. For some isolates originating from large countries a more specific designation was sometimes used, if available. For instance, this was the case for isolates from the city of Kolkata, India (Entrez Protein Database #ABG77411-8), Milwaukee, USA (AAQ01491-7), or the province of Ontario, Canada (ABB04251-8).

Consistent with the species' host specificity, all *C. hominis* isolates ($n = 118$) originated from humans. The 155 *C. parvum* isolates originated from humans ($n = 76$), cattle ($n = 73$) and sheep ($n = 6$). gp60 sequences which did not originate from natural isolates were excluded. This applied to 78 cloned sequences from a laboratory-propagated *C. hominis* isolate (AAT76052-129). From this collection only the last entry was retained.

Sequence analysis

Sequences were downloaded in FASTA format, imported into BioEdit [29], and aligned with Clustal W [30], accessed through the BioEdit Accessory Applications menu. A 98 amino-acid sequence starting at position 35 and ending at position 132 (defined according to GenBank protein sequence AAF78281 [7]) was retained based on the availability of this fragment in most gp60 entries. Amino-acid residues upstream and downstream from these positions were removed. The 98 amino-acid sequence begins 16 amino-acid residues downstream of the predicted signal peptide carboxy terminus, comprises the entire serine repeat and the upstream portion of what was originally identified as the *C. hominis* hypervariable region [7].

gp60 amino-acid polymorphism was analysed at two levels: (1) the number of serine residues in the above-mentioned repeat; (2) the complete 98 amino-acid sequence. The diversity of both variables was analysed using individual-based rarefaction [31, 32]. Coleman rarefaction curves were drawn using the program EstimateS (http://viceroy.eeb.uconn.edu/estimates). Alleles were numbered incrementally, such that serine repeats of equal length were assigned the same allele number. Silent substitutions were not considered. For the amino-acid sequence analysis, identical amino-acid sequences were assigned the same allele number, such that each allele was assigned a unique number. This coding system does not take into consideration the degree of sequence similarity, as each unique allele is assigned a code irrespective of the extent of sequence divergence. As above, the analysis
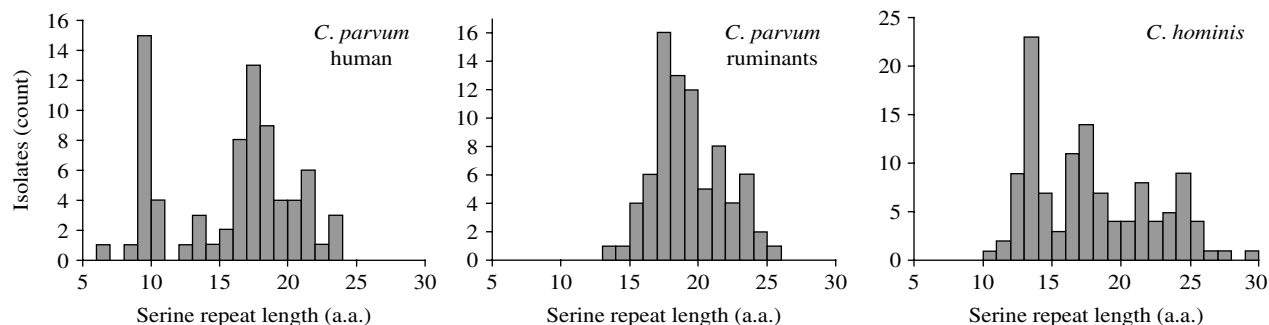
Fig. 1. Distribution of serine repeat length according to *Cryptosporidium* species and host. Short repeats are more prevalent in *C. parvum* from human infections, where 9-residue alleles are particularly common.

being based on the amino-acid sequence, silent substitutions were ignored. Coleman rarefaction numbers and their analytical standard deviations were plotted against the total number of alleles included in the analysis.

To estimate the geographical diversity of individual gp60 genotypes, Simpson's index [33], expressed as $1/D$, and Shannon's H' diversity index [34] were calculated using EstimateS. Phylogenetic trees were drawn using the Neighbour Joining clustering method [35] with Mega 4.0 software [36]. The percentage of replicate trees in which a specific branch occurs was determined by bootstrapping over 500 replicates. The number of non-synonymous substitutions per non-synonymous site (K_a), and synonymous substitutions per synonymous site (K_s) in pairwise sequence comparisons was calculated with DnaSP [37] according to Nei & Gojobori [38] using *C. parvum* sequences AY149616, DQ871348, AF440631, AY382674, AY738189, EF073049, and *C. hominis* sequences EU161648, EU140505, EF576980, EU146136, AY166808, DQ192509.

RESULTS

Analysis of serine repeats

The number of contiguous serine residues in the homopolymeric tract, which in entry AAF78281 initiates at amino-acid position 37, was tabulated. This analysis excluded any serine which was not part of this continuous repeat. In 118 *C. hominis* sequences, the length of the serine repeat ranged from a minimum of 10 to a maximum of 29 residues (Fig. 1). In 155 *C. parvum* sequences, the range was 6–25 residues. The median repeat length for *C. hominis* and *C. parvum* was 17 and 17·5, respectively. The species did not differ significantly with respect to repeat length

(Mann–Whitney rank sum test, $P = 0·55$). In *C. parvum* of ruminant origin (cattle, sheep) ($n = 79$) the range was 13–25, whereas in *C. parvum* of human origin ($n = 76$) repeat length ranged from 6–23 serine residues. The median residue number for the human *C. parvum* (17 residues) was one less than that of isolates originating from ruminants (18 residues). However, serine repeat lengths found in human and ruminant isolates were significantly different (Mann–Whitney rank sum test, $P < 0·01$), as repeats shorter than 13 residues were absent from animal isolates, but were relatively common in human isolates. Contributing to this difference was the frequent occurrence of 9-residue repeats, which was the most abundant repeat length in human *C. parvum* (15/76) (Fig. 1). The frequent occurrence of alleles with 9-residue repeats did not result from over-sampling in a specific region, as these repeats were identified in isolates originating from 10 of 22 regions from which human *C. parvum* gp60 sequences were available. Underscoring the wide geographical distribution of this repeat length, these 10 regions were located on five continents. Five of these 10 countries contributed gp60 sequences from human and ruminant *C. parvum*, which further reduces the possibility that the absence of short repeats in gp60 alleles from ruminants is a result of sampling bias.

Individual-based rarefaction analysis was used to compare the diversity in length polymorphism of the serine repeat between species and between human and animal *C. parvum*. This approach enables a direct comparison of allele richness in different samples, regardless of sample size. By 'rarefying' the large population, in this case *C. parvum*, to the size of the smaller *C. hominis*, repeat length diversity in both species was found to be essentially the same. In *C. hominis* 19 different repeat lengths were observed, whereas *C. parvum*, rarefied from 155 sequences to the
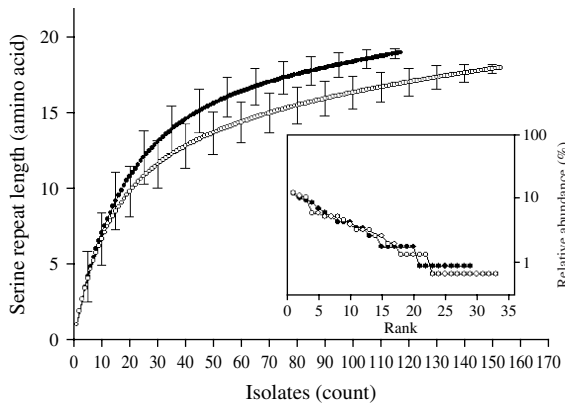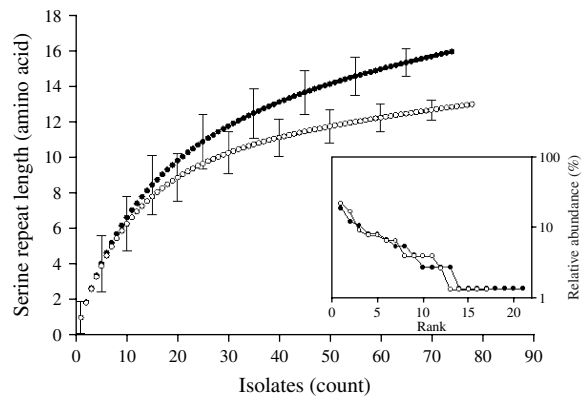
Fig. 2. Rarefaction analysis of serine repeat diversity in *C. parvum* (○) and *C. hominis* (●). Rarefied to the *C. hominis* sample size of 118, *C. parvum* gp60 is almost as diverse as *C. hominis* gp60. Error bars indicate standard deviation. Inset shows rank/abundance plots for the geographical regions represented in the analysis. Regions are ranked from left to right according to the number of isolates. Rank/abundance plots demonstrate that geographical diversity of both species is very similar.


Fig. 3. Rarefaction analysis of serine repeat diversity in *C. parvum* of human (●) and ruminant (○) origin. Allele diversity in *C. parvum* of human origin exceeds that of gp60 from animal isolates. Rank/abundance of geographical region for the two *C. parvum* populations shown in the inset demonstrates a similar geographical diversity for both *C. parvum* populations.

C. hominis sample size of 118, is estimated to have a diversity of 17·0 [95% confidence interval (CI) 15·2–18·8] (Fig. 2). Because unequal sampling could affect the results, region rank/abundance curves were plotted to visualize the geographical diversity of each species. In this analysis, geographical regions were ranked according the number of sequences each region contributed. The curves (Fig. 2, inset) are very similar, with the region ranked no. 1 for *C. parvum* (Holland) contributing 12·1% of the sequences, and the *C. hominis* no. 1 region (South Africa) contributing 11·9% of the sequences (see Supplementary Table 1, available online). This analysis does not imply that each region contributed a similar proportion of samples (a distribution which would generate horizontal rank/abundance plots), but is indicative of a similar geographical diversity for each species. When comparing *C. parvum* of human and ruminant origin with the same approach, more diversity in repeat length was observed in the human sample (16 alleles) than estimated for the animal sample (12·8, 95% CI 11·9–13·7 alleles) (Fig. 3). Rank/abundance curves again demonstrate a similar geographical diversity in these samples (inset).

Analysis of partial amino-acid sequence

Rarefaction analysis was used to compare the gp60 amino-acid sequence diversity in *C. hominis* and *C. parvum*. For both species, the slopes of the rarefaction curves were steep (Fig. 4a), indicating that much diversity remains to be sampled. This was not the case with the repeat length curves (Figs 2, 3), which level off. Rarefied from $n = 155$ to the *C. hominis* sample size of 118, the estimated *C. parvum* gp60 amino-acid sequence diversity is 56·2 (95% CI 50·2–62·2), clearly less than the observed *C. hominis* allele diversity of 70. *C. hominis* is thus more diverse than *C. parvum* at this locus.

Human and ruminant (cattle, sheep, goat) *C. parvum* gp60 allele diversity was also compared (Fig. 4b). The rarefaction analysis confirmed a higher gp60 diversity in human *C. parvum* sequences (46 observed alleles) compared to animal *C. parvum* (36·1 estimated alleles, 95% CI 33·5–38·7). This result is consistent with the wider range in the length of serine repeats in human *C. parvum* described above (see Fig. 1).
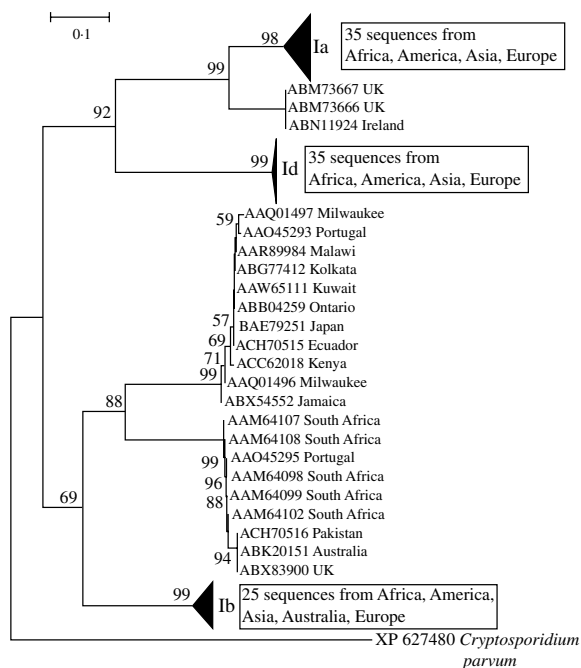
Global distribution of gp60 alleles

In light of the recently described geographical endemism of *C. parvum* and *C. hominis* MLGs [12], I was interested in exploring the global distribution of gp60 alleles. Contrary to MLGs, gp60 alleles showed no geographical partition in either species (Figs 5, 6). The *C. hominis* phylogeny created with the Neighbor Joining algorithm displayed five clearly defined branches, each comprising a relatively homogeneous group of sequences. These branches included alleles originating from widely different locations

(a)

(b)
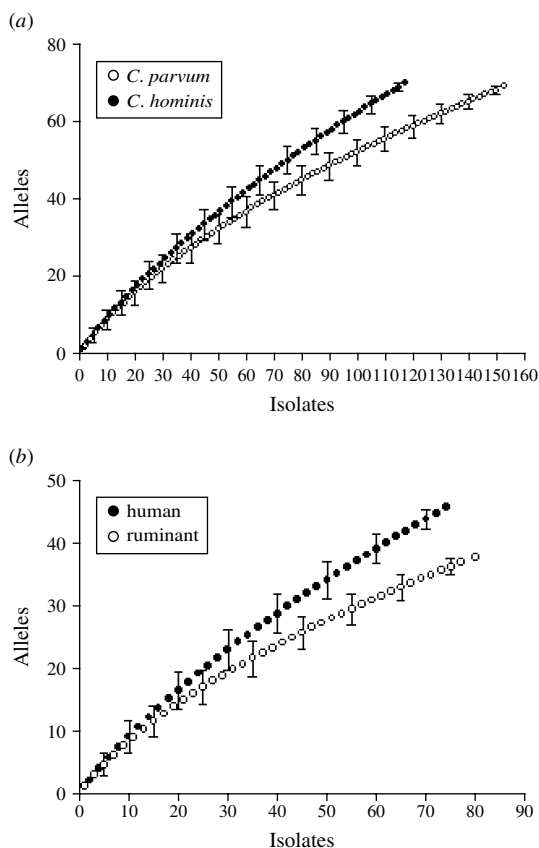
Fig. 4. Amino-acid sequence richness in a 98-amino-acid fragment of the gp60 gene. (*a*) *C. parvum* vs. *C. hominis*; (*b*) *C. parvum* from humans vs. *C. parvum* from ruminants. Allelic diversity in *C. hominis* and *C. parvum* of human origin exceeds that of *C. parvum* and *C. parvum* from ruminants, respectively. Note the steep increase in the number of alleles with increasing sample size, indicating that much of the amino-acid diversity remains to be sampled. For a comparison of geographical diversity see insets in Figures 2 and 3.

Fig. 5. Global phylogeny of *C. hominis* gp60 amino-acid sequences based on the Neighbour Joining method. Bootstrap values based on 500 replicates are shown if > 50 %. Scale indicates number of amino-acid substitutions per site. Triangles represent collapsed groups. Note the complete absence of geographical endemism of allelic groups. Groups are labelled according to Strong & Nelson [7] and Sulaiman *et al.* [9]. *C. parvum* gp60 was used as outgroup.

(Fig. 5). The tree also confirmed the validity of the originally proposed and widely adopted genotype designation [7]. A clade of 35 sequences of the Ia genotype comprised isolates from South America, India, UK, Canada, USA, Africa and Europe. Similarly, in the Id group all five continents are represented. Alleles belonging to genotype Ib form a distinct and geographically equally diverse clade. Genotypes Ie, If and Ig were less common, but the former two were also geographically disperse. Similarly, in the *C. parvum* phylogeny allelic groups showed a wide and overlapping geographical distribution (Fig. 6). The geographical diversity of the most common genotypes was quantified using Simpson's (reciprocal) index $1/D$ and Shannon's index of diversity H' [34]. This analysis showed little difference in geographical diversity in the three main *C. hominis* alleles (Ib: $1/D = 20 \cdot 0$, $H' = 2 \cdot 6$; Id: $1/D = 15 \cdot 6$, $H' = 2 \cdot 7$; Ia: $1/D = 12 \cdot 4$, $H' = 2 \cdot 5$). To compare the geographical diversity of *C. hominis* with that from *C. parvum*, only sequences from human *C. parvum* were included. Animal *C. parvum* was excluded to ensure that different sampling strategies used for surveying humans and cattle would not bias the results. For 35 IIa sequences of human origin $1/D$ was $7 \cdot 0$ and H' was $2 \cdot 5$, and for 17 human IId sequences $1/D$ was $7 \cdot 2$ and H' $1 \cdot 8$. The geographical diversity of the other alleles was not analysed due to small sample sizes. Confidence intervals for the diversity estimates were not calculated, as replicate collections would be needed to generate confidence intervals by jackknifing. A comparison of $1/D$ and H' index values across species suggests that *C. parvum* alleles are geographically less diverse, but it is not clear whether the difference is statistically significant (*t* test, $P = 0 \cdot 052$ for $1/D$ and $P = 0 \cdot 48$ for H').
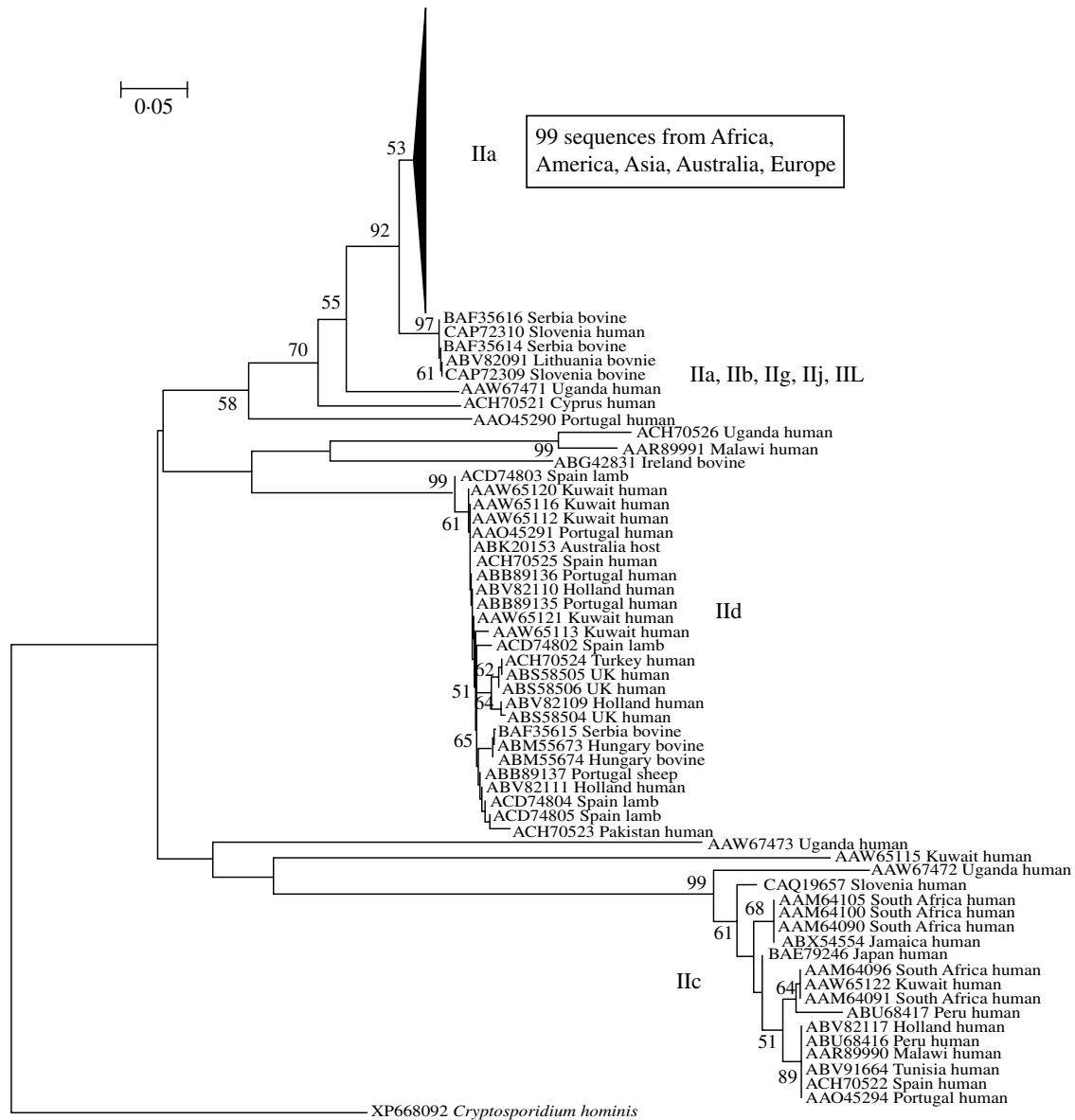
Fig. 6. Global phylogeny of *C. parvum* gp60 amino-acid sequences obtained as described for Fig. 5. As for *C. hominis*, individual clades are geographically diverse. A majority of sequences belonged to genotype IIa which was collapsed and is represented by a triangle.

gp60 non-synonymous and synonymous mutations

Experimental evidence indicates that the protein encoded by the gp60 gene is strongly immunogenic [7, 39], suggesting that the extensive polymorphism may have resulted from selective pressure mediated by the immune response. To assess this possibility, the rate of synonymous and non-synonymous mutations was determined. In pairwise analyses of mutation rates in 12 gp60 sequences (six *C. hominis*, six *C. parvum*) 12/66 informative comparisons gave a $K_a/K_s > 1$. The mean K_a/K_s for 66 pairwise comparisons was 0·84 (s.d. = 0·18). In contrast the K_a/K_s ratio for two

C. parvum/*C. hominis* pairs of actin sequences and lactate dehydrogenase sequences, two genes likely to be under purifying selection, were 0·044 for actin, and 0·045 for lactate dehydrogenase.

DISCUSSION

This report is focused on gp60 polymorphisms and makes no inference on the genetic diversity of the species *C. parvum* and *C. hominis*. The lack of geographical sub-structuring of gp60 alleles is in contrast to the geographical endemism of *C. parvum*

and *C. hominis* MLGs [12]. The different pictures emerging from the wide distribution of gp60 alleles and the geographical endemism of MLGs demonstrate that a single locus, such as gp60, is not a reliable marker of *C. parvum* and *C. hominis* population structure. The discrepancy between single-locus genotypes and MLGs has been noted in a study of 26 human isolates from Jamaica [40]. In accord with the observations reported in the current study, Gatei *et al.* [40] found that *C. parvum* and *C. hominis* isolates sharing a gp60 allele were genetically distinct when other markers were included. Conversely, isolates with distinct gp60 sequences may have related MLGs. Together, these studies show that the gp60 genotype by itself is difficult to reconcile with the concept of *C. parvum* or *C. hominis* 'subtype' frequently used in the literature. The term 'subtype' invokes a genetically distinct population within a species, a model which does not seem to apply to gp60 genotypes.

The availability of a growing collection of hundreds of partial gp60 sequences has enabled a global analysis of the diversity of a biologically important surface glycoprotein which is intimately involved in host–parasite interaction. The gp60 glycoprotein, initially referred to as gp15, was first identified using monoclonal antibodies reacting with *C. parvum* sporozoites and with antigen shed by *C. parvum* sporozoites and merozoites [7, 8]. The protective nature of this antibody, and the fact that the gp60 glycoprotein is recognized by convalescent serum, indicates that this gene may be under positive selection, as observed for the merozoite surface protein family of *Plasmodium falciparum* [41]. The relative high proportion of non-synonymous gp60 substitutions is consistent with selective pressure, probably exerted by the host's immune response. An overlay of the wide geographical distribution of gp60 alleles onto the observed *C. parvum* and *C. hominis* endemic subpopulations [12] suggests that the same gp60 alleles may have emerged in different locations in response to selective pressure.

The current analysis shows an interesting contrast between the diversity in the length of the gp60 serine repeat and the diversity of the amino-acid sequence. Rarefaction curves indicate that most of the variation in serine repeat length has been sampled, whereas much of the amino-acid sequence diversity remains to be identified. In the first description of gp60 polymorphism, the high level of polymorphism in a region downstream of the serine repeat had already been observed [7]. Our analysis confirmed that much of

the *C. hominis* diversity lies outside the serine repeat. The rarefaction curves based on repeat length polymorphism do not show significant differences between *C. parvum* and *C. hominis*, but when observing the 98-residue sequence, *C. hominis* is significantly more diverse.

Of the observations reported here, the frequent occurrence of short serine repeats in *C. hominis* and *C. parvum* of human origin is intriguing. In *C. parvum*, the abundance of short repeats of 9-serine residues is due to the IIc allelic group (see Fig. 6), which appears to be completely absent from animals. Short repeats were also found in *C. hominis*, although none were shorter than 10 residues. Sampling bias was considered as a possible explanation for the absence of IIc in cattle, because many regions where IIc was found did not provide animal samples. However, given the wide geographical distribution of IIc, which was found on three continents, and the partial overlap in the geographical origin of human and animal *C. parvum* sequences, sampling bias does not seem to be a likely explanation for the absence of the IIc alleles in animals. Therefore, these observations suggest that alleles with short repeats may be selectively favoured in the human hosts. Assuming that the host's immune response is the main driver of gp60 diversification, the prevalence of short alleles in parasites infecting humans may indicate differences in selective pressure acting on gp60 in different host species.

NOTE

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/hyg).

DECLARATION OF INTEREST

None.

REFERENCES

1. **Abrahamsen MS, *et al.*** Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 2004; **304**: 441–445.

2. **Xu P, et al.** The genome of *Cryptosporidium hominis*. *Nature* 2004; **431**: 1107–1112.

3. **Mallon M, et al.** Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*. *Journal of Molecular Evolution* 2003; **56**: 407–417.

4. **Tanriverdi S, et al.** Emergence of distinct genotypes of *Cryptosporidium parvum* in structured host populations. *Applied and Environmental Microbiology* 2006; **72**: 2507–2513.

5. **Gatei W, et al.** Multilocus sequence typing and genetic structure of *Cryptosporidium hominis* from children in Kolkata, India. *Infection, Genetics and Evolution* 2007; **7**: 197–205.

6. **Caccio S, et al.** A microsatellite marker reveals population heterogeneity within human and animal genotypes of *Cryptosporidium parvum*. *Parasitology* 2000; **120**: 237–244.

7. **Strong WB, Gut J, Nelson RG.** Cloning and sequence analysis of a highly polymorphic *Cryptosporidium parvum* gene encoding a 60-kilodalton glycoprotein and characterization of its 15- and 45-kilodalton zoite surface antigen products. *Infection and Immunity* 2000; **68**: 4117–4134.

8. **Cevallos AM, et al.** Molecular cloning and expression of a gene encoding *Cryptosporidium parvum* glycoproteins gp40 and gp15. *Infection and Immunity* 2000; **68**: 4108–4116.

9. **Sulaiman IM, et al.** Unique endemicity of cryptosporidiosis in children in Kuwait. *Journal Clinical Microbiology* 2005; **43**: 2805–2809.

10. **Feng X, et al.** Experimental evidence for genetic recombination in the opportunistic pathogen *Cryptosporidium parvum*. *Molecular and Biochemical Parasitology* 2002; **119**: 55–62.

11. **Tanriverdi S, et al.** Genetic crosses in the apicomplexan parasite *Cryptosporidium parvum* define recombination parameters. *Molecular Microbiology* 2007; **63**: 1432–1439.

12. **Tanriverdi S, et al.** Inferences about the global population structures of *Cryptosporidium parvum* and Cryptosporidium hominis. *Applied and Environmental Microbiology* 2008; **74**: 7227–7234.

13. **Leav BA, et al.** Analysis of sequence diversity at the highly polymorphic Cpgp40/15 locus among *Cryptosporidium* isolates from human immunodeficiency virus-infected children in South Africa. *Infection and Immunity* 2002; **70**: 3881–3890.

14. **Ajjampur SS, et al.** Molecular and spatial epidemiology of cryptosporidiosis in children in a semiurban community in South India. *Journal of Clinical Microbiology* 2007; **45**: 915–920.

15. **Hunter PR, et al.** Subtypes of *Cryptosporidium parvum* in humans and disease risk. *Emerging Infectious Diseases* 2007; **13**: 82–88.

16. **Misic Z, Abe N.** Subtype analysis of *Cryptosporidium parvum* isolates from calves on farms around Belgrade, Serbia and Montenegro, using the 60 kDa glycoprotein gene sequences. *Parasitology* 2006; **134**: 351–358.

17. **Xiao L, et al.** Distribution of *Cryptosporidium parvum* subtypes in calves in eastern United States. *Parasitology Research* 2007; **100**: 701–706.

18. **Geurden T, et al.** Molecular epidemiology with subtype analysis of *Cryptosporidium* in calves in Belgium. *Parasitology* 2007; **134**: 1981–1987.

19. **Jex AR, et al.** Classification of *Cryptosporidium* species from patients with sporadic cryptosporidiosis by use of sequence-based multilocus analysis following mutation scanning. *Journal of Clinical Microbiology* 2008; **46**: 2252–2262.

20. **Alves M, et al.** Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. *Journal of Clinical Microbiology* 2003; **41**: 2744–2747.

21. **Zintl A, et al.** The prevalence of *Cryptosporidium* species and subtypes in human faecal samples in Ireland. *Epidemiology Infection* 2009; **137**: 270–277.

22. **Peng MM, et al.** Genetic polymorphism among *Cryptosporidium parvum* isolates: evidence of two distinct human transmission cycles. *Emerging Infectious Diseases* 1997; **3**: 567–573.

23. **Grinberg A, et al.** Genetic diversity and zoonotic potential of *Cryptosporidium parvum* causing foal diarrhea. *Journal of Clinical Microbiology* 2008; **46**: 2396–2398.

24. **Cama VA, et al.** *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. *Emerging Infectious Diseases* 2008; **14**: 1567–1574.

25. **Peng MM, et al.** Genetic diversity of *Cryptosporidium* spp. in cattle in Michigan: implications for understanding the transmission dynamics. *Parasitology Research* 2003; **90**: 175–180.

26. **Cohen S, et al.** Identification of Cpgp40/15 Type Ib as the predominant allele in isolates of *Cryptosporidium* spp. from a waterborne outbreak of gastroenteritis in South Burgundy, France. *Journal of Clinical Microbiology* 2006; **44**: 589–591.

27. **Al-Brikan FA, et al.** Multilocus genetic analysis of *Cryptosporidium* isolates from Saudi Arabia. *Journal of the Egyptian Society of Parasitology* 2008; **38**: 645–658.

28. **Morgan-Ryan UM, et al.** *Cryptosporidium hominis* n.sp. (Apicomplexa: Cryptosporidiidae) from *Homo sapiens*. *Journal of Eukaryotic Microbiology* 2002; **49**: 433–440.

29. **Hall T.** BioEdit, a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999; **41**: 95–98.

30. **Thompson JD, Higgins DG, Gibson TJ.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994; **22**: 4673–4680.

31. **Gotelli NJ, Colwell RK.** Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 2001; **4**: 379–391.

32. **Coleman BD.** On random placement and species-area relations. *Mathematical Biosciences* 1981; **54**: 191–215.

33. **Simpson EH.** Measurement of diversity. *Nature* 1949; **163**: 688–688.

34. **Magurran AE.** *Measuring Biolocial Diversity*. Oxford: Blackwell Publishing, 2004.

35. **Saitou N, Nei M.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 1987; **4**: 406–425.

36. **Tamura K, *et al.*** MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 2007; **24**: 1596–1599.

37. **Rozas J, *et al.*** DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003; **19**: 2496–2497.

38. **Nei M, Gojobori T.** Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 1986; **3**: 418–426.

39. **Cevallos AM, *et al.*** Mediation of *Cryptosporidium parvum* infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody. *Infection and Immunity* 2000; **68**: 5167–5175.

40. **Gatei W, *et al.*** Unique *Cryptosporidium* population in HIV-infected persons, Jamaica. *Emerging Infectious Diseases* 2008; **14**: 841–843.

41. **Escalante AA, Lal AA, Ayala FJ.** Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 1998; **149**: 189–202.