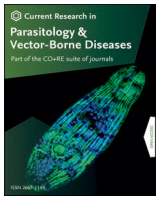
Current Research in Parasitology & Vector-Borne Diseases 7 (2025) 100257

Contents lists available at ScienceDirect

Current Research in Parasitology & Vector-Borne Diseases

journal homepage: www.sciencedirect.com/journal/current-research-in-parasitology-and-vector-borne-diseases



Deciphering a cryptic minefield: A guide to *Cryptosporidium gp60* subtyping

Guy Robinson [a,b,*], Rachel M. Chalmers [a,b], Kristin Elwin [a], Rebecca A. Guy [c], Kyrylo Bessonov [c], Karin Troell [d,e], Lihua Xiao [f]

[a] *Cryptosporidium Reference Unit, Public Health Wales Microbiology and Health Protection, Singleton Hospital, Swansea, SA2 8QA, UK*
[b] *Swansea University Medical School, Singleton Park, Swansea, SA2 8PP, UK*
[c] *Division of Enteric Diseases, National Microbiology Laboratory, Public Health Agency Canada, Guelph, Ontario, N1G 3W4, Canada*
[d] *Norwegian Veterinary Institute, Elizabeth Stephansens vei 1, 1433, Ås, Norway*
[e] *Swedish Veterinary Agency, Ulls väg 2, 75189, Uppsala, Sweden*
[f] *College of Veterinary Medicine, South China Agricultural University, 483 Wushan Road, Guangzhou, 510642, China*

ARTICLE INFO

Keywords:
Cryptosporidium
gp60
Nomenclature
Subtyping
CryptoGenotyper

ABSTRACT

For 25 years, analysis of the *gp60* gene has been the cornerstone of *Cryptosporidium* subtyping, particularly for *Cryptosporidium hominis* and *Cryptosporidium parvum*, during population-based and epidemiological studies. This gene, which encodes a 60 kDa glycoprotein, is highly polymorphic with several variable features that make it particularly useful for differentiating within *Cryptosporidium* species. However, while this variability has proven useful for subtyping, it has on occasion resulted in alternative interpretations, and descriptions of novel and unusual features have been added to the nomenclature system, resulting in inconsistency and confusion. The components of the *gp60* gene sequence used in the nomenclature that are discussed here include "R" repeats, "r" repeats, alphabetical suffixes, "variant" designations, and the use of the Greek alphabet as a family designation. As the subtyping scheme has expanded over the years, its application to different *Cryptosporidium* species has also made the scheme more complex. For example, key features may be absent, such as the typical TCA/TCG/TCT serine microsatellite that forms a major part of the nomenclature in *C. hominis* and *C. parvum*. As is to be expected in such a variable gene, different primer sets have been developed for the amplification of the *gp60* in various species and these have been collated. Here we bring together all the current components of *gp60*, including a guide to the nomenclature in various species, software to assist in analysing sequences, and links to useful reference resources with an aim to promote standardisation of this subtyping tool.

1. Introduction

Species of the protozoan genus *Cryptosporidium* are globally important One Health pathogens, causing the gastrointestinal illness cryptosporidiosis in both humans and animals (Innes et al., 2020). There are unequal impacts of this disease, for example, moderate-to-severe diarrhoea, growth faltering and increased mortality in young children in low- and middle-income countries, a burden of losses in intensive agricultural systems, and widespread human outbreaks of public health importance in all ages recognised in high-income countries (Innes et al., 2020; Levine et al., 2020; Yang et al., 2021; Zahedi and Ryan, 2020).

Investigations into the molecular epidemiology of *Cryptosporidium* have relied heavily over the past 25 years on PCR-based genotyping, particularly at the small subunit rRNA (SSU rRNA) gene, to determine the species causing infection, and within-species subtyping by sequencing part of a highly polymorphic gene encoding a 60 kDa glycoprotein (*gp60*) (Xiao and Feng, 2017). Over this time, the number of species identified has increased to more than 40, and many more host-adapted genotypes are awaiting valid species status (Ryan et al., 2021a). The *gp60* subtyping has mainly focused on the two species that cause most human infections, *C. hominis* (predominantly anthroponotic) and *C. parvum* (zoonotic with a wide host range), and the latter also impacting livestock (Feng et al., 2018).

The *gp60* gene codes for an immunodominant surface glycoprotein that mediates host cell invasion and thus plays a role in host infectivity and pathogenicity, consistent with its highly polymorphic nature (Li et al., 2024). The DNA, and corresponding amino acid, sequence variation seen at this locus both within and between *Cryptosporidium* species has provided multiple features for discrimination that have been included in the nomenclature of the *gp60* subtyping scheme. These

* Corresponding author. Cryptosporidium Reference Unit, Public Health Wales Microbiology and Health Protection, Singleton Hospital, Swansea, SA2 8QA, UK.
E-mail address: Guy.Robinson@wales.nhs.uk (G. Robinson).

https://doi.org/10.1016/j.crpvbd.2025.100257
Received 7 February 2025; Received in revised form 11 March 2025; Accepted 18 March 2025
Available online 27 March 2025

features include DNA sequence polymorphisms and mini- and microsatellite repeats. While the *gp60* typing scheme started with a simple nomenclature (i.e. Roman numeral for species and alphabetical family designation) (Strong et al., 2000), it has evolved as new features have been discovered by different investigators and has become increasingly complicated. The evolution of these new features and contradictions in their use has led to the nomenclature becoming confusing. This has led to errors, mostly incorrect naming or misinterpretation of the nomenclature, which differs depending on the species or allelic family in question, making it difficult to compare subtypes. This highlights the need for a comprehensive guide to standardise and promote a harmonised nomenclature for *gp60* subtyping, including online resources and tools to aid the community in the future evolution of this scheme. Here, we describe the various features underpinning a standardised nomenclature, highlighting pitfalls to avoid errors in interpretation, and provide useful reference resources for those analysing this locus.

2. The evolution of *gp60* subtyping and some of the problems encountered along the way

Since the *Cryptosporidium gp60* gene was first described by Strong et al. (2000), its use as a target for subtyping has been widely accepted by the scientific community and globally used in population studies as well as investigations into human and animal cryptosporidiosis, especially outbreaks (Xiao and Feng, 2017; Feng et al., 2018; Chalmers et al., 2019; Ryan et al., 2021b; Yang et al., 2021). The majority of *gp60* studies, particularly in the early days, focused on the two main species of veterinary (*C. parvum*) and public health (*C. parvum* and *C. hominis*) importance. Over the years the gene has been explored and described in many more *Cryptosporidium* species, each of which is provided with a *gp60* designation as the first part of the nomenclature (Table 1).

2.1. Basic gp60 nomenclature developed for C. parvum and C. hominis

The initial *gp60* subtyping nomenclature was developed by Strong et al. (2000) for *C. hominis* and *C. parvum* (at that time referred to as *C. parvum* genotype I or "human" genotype and genotype II or "bovine" genotype, respectively). Distinct allelic groupings were identified by sequence polymorphisms extending the species designations I and II to define "*gp60* families" named Ia, Ib, Ic and Id for *C. hominis* and II for *C. parvum*. The family Ic was later corrected to *C. parvum* allelic family IIc, and the name "Ic" retired to avoid future confusion (Sulaiman et al., 2005). This *gp60* family differs from other *C. parvum* families as it appears to be primarily host-adapted to humans, whereas other families such as IIa and IId are considered zoonotic.

Sulaiman et al. (2005) also introduced the second main element of the *gp60* subtype nomenclature, which describes the composition of the serine repeat located upstream of the family-defining region. The serine microsatellite is composed of TCA, TCG or TCT codons and the number of each is counted and recorded, prefixed by A, G or T, following the subtype family name (Sulaiman et al., 2005). For example, IeA11G3T3 is a subtype of *C. hominis* belonging to the Ie allelic family with 11 TCA, 3 TCG and 3 TCT codons in the serine repeat, and IIdA23G1 is a *C. parvum* subtype belonging to the IId allelic family with 23 TCA and 1 TCG codons in the serine repeat.

The count of serine codons recorded in this way terminates at the first non-serine codon. However, the following R repeat in some subtype families often includes a single serine codon, which occasionally was added to the count of serine codons and resulted in inconsistencies in the nomenclature as discussed below in *Section* 2.2 on "R" repeats. However, as with many subtyping schemes, particularly one as nuanced as this, there are some uncommon exceptions to this rule, as discussed below in *Section* 2.3 on "r" repeats.

2.2. The R repeat designation

In addition to the polyserine repeat, occasional mini- and microsatellites that define different subtypes are also incorporated in the nomenclature of a few *gp60* subtype families (e.g. *C. parvum* IIa, *C. hominis* Ia and If, and *C. cuniculus* Vb) as an "R" repeat (see Table 2). These R repeats, which follow different nomenclature rules depending on the species and *gp60* family, have been the source of some confusion leading to errors and misnaming of isolates, which often remain on NCBI's GenBank and in turn exacerbate the problem during new BLAST searches. The R repeats can be different sequences between species and subtypes and in alternative locations within the gene. For example, the R repeats in *C. parvum* IIa (ACATCA) and *C. cuniculus* Vb (ACA), are

Table 1
A list of *Cryptosporidium* species with described *gp60* genes.

Species	*gp60* designation	Key references
C. hominis	I	Strong et al. (2000); Sulaiman et al. (2005)
C. parvum	II	Strong et al. (2000); Sulaiman et al. (2005)
C. meleagridis	III	Glaberman et al. (2001); Stensvold et al. (2014)
C. fayeri	IV	Power et al. (2009)
C. cuniculus	V	Chalmers et al. (2009)
C. equi	VI	Xiao et al. (2009)
C. wrairi	VII	Lv et al. (2009)
C. sciurinum	VIII	Lv et al. (2009)
C. tyzzeri	IX	Lv et al. (2009)
Mink genotype	X	Feng et al. (2011)
Opossum genotype I	XI	Feng et al. (2011)
C. ubiquitum	XII	Li et al. (2014)
C. erinacei	XIII	Laatamna et al. (2013)
C. mortiferum	XIV	Guo et al. (2015)
C. viatorum	XV	Stensvold et al. (2015)
Skunk genotype	XVI	Yan et al. (2017)
Apodemus genotype I	XVII	Čondlová et al. (2019)
Apodemus genotype II	XVIII	Čondlová et al. (2019)
C. felis	XIX	Rojas-Lopez et al. (2020); Jiang et al. (2020)
C. canis	XX	Jiang et al. (2021)
C. ryanae	XXI	Yang et al. (2020)
C. myocastoris	XXII	Ježková et al. (2021)
C. xiaoi	XXIII	Fan et al. (2021)
C. occultus	XXIV	Unpublished: PV067136-PV067140
C. suis	XXV	Lebbad et al. (2024)
C. bovis	XXVI	Wang et al. (2021)

Table 2
The features of standardised *gp60* nomenclature for *Cryptosporidium* spp. For full details, see Supplementary file S3. An up-to-date, contemporary version with the current accepted families and features is available online on the *Cryptosporidium gp60* Resource Page at https://cryptodb.org/cryptodb/app/static-content/gp60.html.

	Roman numeral designation	Subtype family range	TCA/TCG/TCT serine trinucleotide repeats present	Families in which R repeats occur	Families in which r repeats occur	Families in which alphabetical suffixes occur[a]	Numerical designation representing sequence variations	Example subtypes (GenBank ID)
C. hominis	I	Ia-Io	Yes	Ia, If	No	Ib[b]	No	IaA12R3 (EU095263), IbA9G3b (JF927195), IeA11G3T3 (AY738184), IfA19G1R5 (AF440638)
C. parvum	II	IIa-IIgamma	Yes	IIa, IIl	IIa	IIc, IId	No	IIaA14G1R1r1 (AB777446), IIcA5G3j (HQ005749), IIdA20G1d (AY738186), IIlA19R3 (KP997147), IIgammaA11 (PP157591)
C. meleagridis	III	IIIa-IIIk	Yes	No	No	Yes	No	IIIaA24G3a (AF401499), IIIfA16G2 (EU164813)
C. fayeri	IV	IVa-IVg	Yes	No	No	No	No	IVaA11G3T1 (FJ490060), IVfA12G1T1 (FJ490076)
C. cuniculus	V	Va-Vb	Yes	Vb	No	Vb	No	VaA18 (FJ262730), VbA20R2b (KU852735),
C. equi	VI	VIa-VId	Yes	No	No	No	No	VIaA11G3 (FJ435960), VIcA16 (KU852738)
C. wrairi	VII	VIIa-VIIb	Yes	No	No	No	No	VIIaA17T1 (GQ121028), VIIbA13 (PP454709)
C. sciurinum	VIII	VIIIa-VIIIc	Yes	No	No	No	No	VIIIaA5G2 (GQ121029), VIIIcA12G2 (MF411081)
C. tyzzeri	IX	IXa-IXc	Yes	Yes[c]	No	No	No	IXaA6R3 (GQ121030), IXbA6 (HM234177)
Mink genotype	X	Xa-Xg	Yes	No	No	No	No	XaA5G1 (HM234174), XdA4G1 (KU608304)
Opossum genotype I	XI	XIa	Yes	No	No	No	No	XIaA4G1T1 (HM234181)
C. ubiquitum	XII	XIIa-XIId	No	No	No	No	XIIa	XIIa3 (JX412915), XIId (JX412922)
C. erinacei	XIII	XIIIa-XIIIb	Yes	Yes	No	No	No	XIIIaA20R10 (KF055453)
C. mortiferum	XIV	XIVa	Yes	No	No	Yes	No	XIVaA18G2T1b (KP099086), XIVaA20G2T2 (KP099083)
C. viatorum	XV	XVa-XVd	Yes	No	No	Yes	No	XVaA3c (KP115938), XVcA2G1a (MK433562)
Skunk genotype	XVI	XVIa-XVId	Yes	No	No	Yes	No	XVIbA16G2b (KX698285), XVIdA29 (KX698301)
Apodemus genotype I	XVII	XVIIa	No	Yes	No	No	No	XVIIaR2 (MH912979), XVIIaR3 (MH912981)
Apodemus genotype II	XVIII	XVIIIa	No	No	No	No	Yes	XVIIIa1 (MH912970), XVIIIa2 (MH912974)
C. felis	XIX	XIXa-XIXe	No	No	No	No	Yes	XIXa15 (MT458679), XIXc2 (MT458668)
C. canis	XX	XXa-XXj	No	No	No	No	Yes	XXa4 (ON863565), XXg2 (ON863569)
C. ryanae	XXI	XXIa-XXIk	No	No	No	No	No	XXIa (MT588099), XXIf (MT588094)
C. myocastoris	XXII	XXIIa-XXIIb	No	No	No	No	No	XXIIa (MW280995), XXIIb (MW280993)
C. xiaoi	XXIII	XXIIIa-XXIIIl	No	No	No	No	No	XXIIIa (MW815183), XXIIIk (MW815253)
C. occultus	XXIV	XXIVa-XXIVb	No	Yes	No	No	Yes	XXIVa (PV067136), XXIVb (PV067140)
C. suis	XXV	XXVa	No	Yes[d]	No	No	Yes[d]	XXVaR37 (MH187875), XXVa3 (PP467568)
C. bovis	XXVI	XXVIa-XXVIf	No	No	No	No	No	XXVIa (MZ977200), XXVIf (MZ977143)

[a] Alphabetical suffixes are used to differentiate members of the same subtype. As such, they are not used for all members of that subtype family, and there can be sequence variation in the downstream region between different subtypes using the same suffix. For example, the downstream sequence is different between IIIbA21G1R1b and IIIbA22G1R1b (Stensvold et al., 2014), and between XIVaA18G2T1b and XIVaA19G2T2b (Guo et al., 2015).

[b] Alphabetical suffixes in family Ib should only be used for those subtypes that are more variable, such as IbA9G3 and not highly conserved subtypes, such as IbA10G2 that may have a few rare variants (Supplementary file S1 and Supplementary file S2).

[c] The R regions in *C. tyzzeri* are still being fully characterised as there are multiple longer repeats. Therefore, it is currently better to describe each of these longer repeats than assign an R count.

[d] In *C. suis* a 21 bp minisatellite R repeat is present, but difficult to amplify and is normally not included in the nomenclature which instead uses a numerical count to represent sequence variation in the downstream region (Lebbad et al., 2024).

located immediately after the polyserine repeat (Sulaiman et al., 2005; Lebbad et al., 2021), whereas *C. hominis* has R repeats in families Ia (AAGACGGTGGTAAGG) and If (AAGAAGGCAAAGAAG) that differ not only in their repeat motifs, but also in their distances from the end of the serine microsatellite (95 bp downstream for Ia and 25 bp for If) (Cama et al., 2007; Xiao and Feng, 2017). Although some other species and families have these motifs, particularly ACATCA at the end of the serine repeat, they should not be included in the nomenclature, unless they have been shown to display different numbers of repeats within subtype families (see Table 2). For example, Bujila et al. (2024b) reported the new subtype families IIy and IIz, both of which contain the ACATCA repeat straight after the polyserine repeat and proposed IIyA23G1R1 and IIzA14R2, but no variation in the number of this R repeat has yet been identified in those families and therefore these subtypes should be named IIyA23G1 and IIzA14 respectively. Should variation be subsequently identified in future samples, comparison of the sequences would identify the original R repeat(s).

R repeats are also used for some other species including *C. tyzzeri*, *C. erinacei* and *C. suis* (Xiao and Feng, 2017; Lebbad et al., 2024) and are described below in *Section* 2.8.

Another area of confusion with R repeats is their DNA sequence variability making it particularly difficult for users to identify them as repeats and hence the correct number. This is due to how the repeats have been described, based on a repeat identified from the nucleotide sequence, often using Tandem Repeat Finder (Benson, 1999), whereas investigating the resulting amino acid sequence can simplify the identification. Two of the R repeats that cause the most difficulty are in *C. hominis* subtype families Ia and If.

The Ia R repeat is defined as a series of 15 bp repeats of AAGACGGTGGTAAGG (Cama et al., 2007; Xiao and Feng, 2017) and a final repeat which has a G to A substitution at nucleotide 3 and three base-pair deletions at nucleotides 10, 11 and 15 (AAAACGGTGAAG). However, the original Ia repeat stated by Strong et al. (2000) was described as a tandemly repeated DGGKE amino acid sequence (translated from GACGGTGGTAAGGAA) and would have been simpler to follow with no deviations, but unfortunately this was overlooked in the nomenclature and historical precedence makes it impossible to correct.

The If R repeat is even more complicated and although originally defined as the 15 bp motif AAGAAGGCAAAGAAG (Xiao and Feng, 2017), the repeat motifs vary in size (12–16 bp) and sequence (see Fig. 1). These repeats were initially identified using the Tandem Repeats Finder (Benson, 1999), which does not consider reading frames and amino acid sequences. Therefore, many of the If R repeats are not identical to the defined consensus motif, and a more accurate description of the repeat sequences described would be [A/C]AGA[A/G]GGCA [A/G][A/T]GAAG, with additional notes: the first repeat contains a G insertion at position 11, the fifth repeat in an R6 subtype is 12 bp long as the first three nucleotides are deleted and the last repeat contains an A-to-G transition and A-to-T transversion at positions 10 and 11 (Fig. 1).

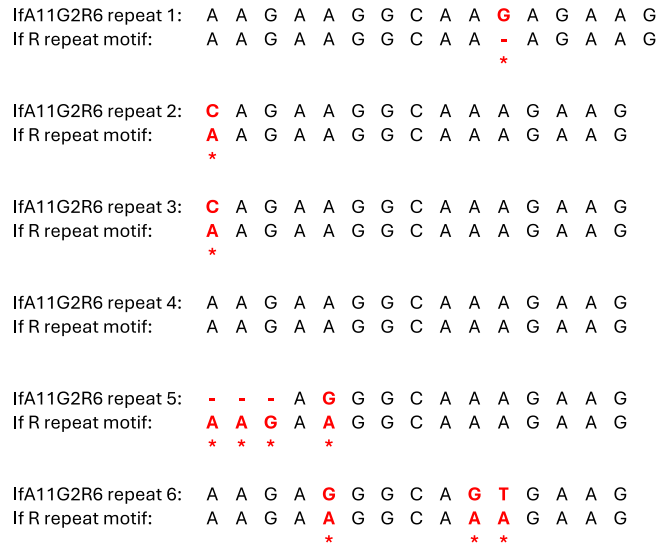


Fig. 1. Example of the variability in R repeats in a IfA11G2R6 (MH796379) sample and how they vary (*) from the published If R repeat sequence.

In some of the published literature R repeats have been defined differently resulting in an alternative subtype nomenclature, which unfortunately adds an additional complication when trying to compare isolates. Three publications and related GenBank entries redefined the R repeat in *C. parvum* and *C. hominis* as the trinucleotide codon for threonine 'ACA' or 'ACG' and included the subsequent TCA/TCG/TCT serine codons in the microsatellite count (Jex et al., 2007; Jex and Gasser, 2008, 2010). This resulted in *C. parvum* IIa subtypes with additional As and Gs (e.g. IIaA18G3R1 misnamed as IIaA19G3R1). Furthermore, *C. hominis* Ib and *C. parvum* IIc subtypes were allocated R repeats because their serine repeat ended with ACAACA (e.g. IbA10G2R2 and IIcA5G3R2 instead of IbA10G2 and IIcA5G3, respectively). This continues to impact the correct naming of subtypes (e.g. Banda et al., 2022).

2.3. The r repeat designation in IIa

In *C. parvum* family IIa, the very rare occurrence of an ACA threonine codon within the polyserine repeat region has been designated as an "r" repeat with a digit to represent the number of repeats and an alphabetical suffix to signify its different positions within the serine tract. This nomenclature stems from Amer et al. (2013), who described an unusual IIa sequence from a calf in Egypt, IIaA14G1R1r1b, which included an additional ACA codon, sandwiched between the end of the serine repeat and the ACATCA R repeat (GenBank: AB777446). They hypothesised that this extremely rare repeat could be due to either a T to A substitution in the last serine repeat or a deletion of the TCA codon in the first of two R repeats (Amer et al., 2013). However, the position of this ACA r repeat is variable as seen in GenBank: EU078321 from a sheep in Germany and GenBank: JX183796 (Silverlås et al., 2013) from a calf in Sweden where the position of the ACA r motif is part way along the serine tract with 11 repeats before and four after. All three of these r repeat subtypes had 14 TCA repeats and 1 TCG repeat, and it was proposed to differentiate between them using an alphabetical suffix. Thus, both EU078321 and JX183796 are IIaA14G1R1r1a, and AB777446 is IIaA14G1R1r1b (Amer et al., 2013). Subsequently, IIaA14G1R1r1a was reported in six human cases in Sweden, of which two reported foreign travel - one to Spain and one to Uzbekistan (Lebbad et al., 2021).

The presence of this r sequence interspersed in the serine repeat has resulted in naming errors that are still present in GenBank entries EU078321 and JX183796. Although the r group nomenclature was not published until the third finding (GenBank: AB777446), the rest of the reported subtype names were also incorrect as the rules for R repeats had not been followed by Silverlås et al. (2013). EU078321 was reported as IIaA14G1R2, where the authors correctly counted the number of TCA and TCG repeats but included the ACA motif as an extra ACATCA R repeat. JX183796 was published as IIaA13G1R2, as the "ACATCA" within the serine repeat was included as an additional R repeat, resulting in 13 TCAs reported instead of 14 (Silverlås et al., 2013).

2.4. Alphabetical suffix designation

The use of alphabetical suffixes in *gp60* nomenclature is not restricted to variations in r motifs. Traditionally, alphabetical suffixes in *gp60* subtyping are used to differentiate variants of the same subtype. The most common use of these suffixes is within the IIc *C. parvum gp60* subtype family. The vast majority of IIc subtypes share the same polyserine repeat, which contains five TCA and three TCG repeats, and rely upon single nucleotide polymorphisms (SNPs) in the downstream family-defining region to differentiate, each of which is designated an

alphabetical suffix (Cama et al., 2007). Some of these vary by only a single SNP, such as IIcA5G3p and IIcA5G3s, some by a few SNPs, such as IIcA5G3a, IIcA5G3j and IIcA5G3m, and others differ by more than 30 SNPs, such as IIcA5G3b.

It is not only IIc isolates that have alphabetical suffixes assigned to further differentiate subtypes, but also *C. parvum* IId due to individual SNPs in the post-repeat region (Sulaiman et al., 2005; Silverlås et al., 2010; Gilchrist et al., 2018; Lebbad et al., 2021). Sulaiman et al. (2005) identified SNPs in the downstream region of four IIdA20G1 sequences and named them a-d. Silverlås et al. (2010) also noted further variation (IIdA20G1e) as well as SNP differences in IIdA22G1 isolates. Here, alphabetical suffix designation was based on individual microsatellite subtypes rather than standardising the SNPs for subtype family sequences regardless of the serine repeat. Therefore, the downstream SNPs for IIdA20G1c differ from those in IIdA22G1c (Sulaiman et al., 2005; Silverlås et al., 2010). This has also been applied to other species (e.g. *C. meleagridis*) and families (e.g. Ib and IIa) (Misic and Abe, 2007; Silverlås et al., 2013; Alsmark et al., 2018; Lebbad et al., 2021). As mentioned above in *Section* 2.3, family IIa has the added complication of alphabetical suffixes to indicate the position of an ACA r repeat within the polyserine repeat (Amer et al., 2013).

Gilchrist et al. (2018) described two suffixes (a and b) for IbA9G3 isolates in Bangladesh based on a SNP in the downstream region, with the two subtypes showing different temporal distribution. There are several groups of downstream SNPs in Ib sequences that can be used similarly to further differentiate subtypes (Supplementary file S1 and Supplementary file S2). However, assigning alphabetical suffixes is only useful in more variable subtypes such as IbA9G3, and application to rare variants in highly prevalent subtypes such as IbA10G2 may be confusing and is not recommended (Table 2).

The practice of alphabetical suffixes is also adopted for some other species. *Cryptosporidium viatorum* subtypes (e.g. XVaA3a-g) mostly contain three TCA repeats and are differentiated based on post-repetitive region SNPs (Stensvold et al., 2015). However, for *C. mortiferum* the suffix use differs, designating the order of TCA/TCG/TCT repeats and post-repeat SNPs, but these designations are not standardised. For example, within subtype XIVaA18G2T1a there is a tandem repeat of 4 serine codons TCTTCATCGTCA in the microsatellite, whereas in XIVaA18G2T1b this feature is rearranged to TCATCTTCATCG and there is a G-to-A transition in the downstream region. In contrast, XIVaA19G2T2a and XIVaA19G2T2b differ only by a single differently located G-to-A transition (Guo et al., 2015). The *C. mortiferum* suffixes for downstream variations are also not based on defined sequences, unlike those in *C. parvum* IIc. For example, subtype XIVaA18G2T1b has the same downstream sequence as XIVaA19G2T2a, while XIVaA18G2T1a and XIVaA19G2T2b have other, different downstream regions (Guo et al., 2015).

2.5. "Variant" designation

While some publications have designated alphabetic suffixes to sequences containing variations in the position of a TCG repeat within the polyserine microsatellite region (Alsmark et al., 2018), Lebbad et al. (2021) instead reserved the alphabetic suffixes for the 3′ SNPs (e.g. IIaA16G1R1b) and added the word "variant" where this TCG variation was present (e.g. IIaA15G1R1_variant and IIaA16G1R1b_variant). In their study of *Cryptosporidium* diversity in infections acquired in Sweden and abroad, they identified several *C. parvum* IIa variant samples that had TCA and TCG in the 3rd and 4th serine positions instead of TCG and TCA, respectively. The problem is that the word "variant" is not quantifiable. It does not provide information as to the position of the TCG repeat or an individual designation that can be labelled in a reference sequence (unlike those using alpha-numeric designations), and if a third or fourth position for the TCG repeat is identified, how would this be represented? So, although we do not recommend the use of this terminology as part of the standard nomenclature, it is useful that these variations are detected and compared, and we propose v1, v2, etc. for such purposes.

2.6. Use of Greek alphabet

With the recognition of the large number of *C. parvum gp60* families, the allocation of Latin alphabet letters reached its end with the submission of IIy and IIz to GenBank in 2021 (Bujila et al., 2024b). This created a dilemma for the future naming of families. In 2022, the next new family was published on GenBank using the Greek alphabet and named IIβ (GenBank: ON646229), although "beta" was spelled out in the GenBank record rather than using the symbol. This has been followed by IIγ (GenBank: PP157591) identified in a human in Sweden (Bujila et al., 2024a). However, the adoption of the Greek (or any) alphabet to continue the allocation of *C. parvum* families comes with limitations. The first letter used was β rather than α, due to the similarity to the Latin letter a in some fonts. This will also be the case with several of the other letters in the Greek alphabet, including γ, η, κ, ν, ο, ρ, υ, χ and ω. To overcome this problem, we recommend that the Greek alphabet letter name is written in full (alpha, beta, gamma, etc.) rather than using the symbol. If this approach is taken then all of the Greek letters can be applied, adding an additional 24 potential family designations. When the Greek alphabet is exhausted, another extension of the nomenclatural system will have to be considered.

2.7. Notable unusual gp60 subtypes

Some *gp60* subtypes are more common than others in different hosts and settings as has been reviewed extensively in previous publications (Xiao and Feng, 2017; Ryan et al., 2021a, 2021b; Oladele et al., unpublished), however there are some unusual subtypes that are worth noting for their interesting epidemiology.

The anthroponotic *C. parvum* IIc subtypes are consistently described as having the same serine A5G3 repeat units, with the only differences being SNPs within the 3′ region. However, there are exceptions, for example during examination of IIc isolates causing human infections in England and Wales in 2013 and 2014, a cluster of five cases infected with subtype IIcA3G2a (represented by GenBank accession PV013404) were identified, all of which were from adults (four 50–59 year-old males with HIV and a 27 year-old female) in southern England. During the same investigation we also identified a single IIcA6G2j isolate (PV013405) in a specimen from a 29 year-old female from the East Midlands in England.

While *C. hominis* is considered an anthroponotic species with infections largely restricted to humans, there are reports of some specific *C. hominis gp60* families causing infections in non-human primates such as Ii, Ij, Im and In (Chen et al., 2019) and equids with subtype family Ik (Ryan et al., 2021b). However, these subtypes of *C. hominis* have very divergent genomes (Huang et al., 2024). Although not regularly identified in humans, they appear to be zoonotic with some human cases reported, particularly from humans that have had potential exposure to non-human primates or their environment (Lebbad et al., 2018; Toriro et al., 2024; PV013406).

2.8. gp60 in different Cryptosporidium species

The majority of *Cryptosporidium* species contain the characteristic *gp60* serine microsatellite region that is used in the standard nomenclature; however there are some exceptions in which it is lacking, including *C. ubiquitum* (XII), *C. felis* (XIX), *C. canis* (XX), *C. ryanae* (XXI), *C. xiaoi* (XXIII) and *C. bovis* (XXVI) (Li et al., 2014; Rojas-Lopez et al., 2020; Yang et al., 2020; Jiang et al., 2021; Fan et al., 2021; Wang et al., 2021). Consequently, their nomenclature consists only of the species designation (e.g. XII for *C. ubiquitum*), the subtype family determined by sequence variations/SNPs (a, b, c, etc.), followed by a numerical count to differentiate between variations. For example, *gp60* subtype XXb2

refers to the 2nd subtype found within the *C. canis* subtype family b (Jiang et al., 2021).

In addition to the sequence variations in the *C. felis gp60* gene that differentiate between the families (currently a-e), there are several tandem repeats and indels, including two long minisatellites (33 bp and 39 bp) a 3 bp GGT microsatellite, that differentiate subtypes within the families that are represented by a numerical count, particularly within XIXa (e.g. XIXa1 to XIXa18) (Jiang et al., 2020).

Some species, such as *C. erinacei*, show the presence of the ACATCA R repeat that occurs in some of the *C. parvum gp60* subtype families, while there are other species that display longer or multiple R repeats. For example, *C. tyzzeri* (IX) tends to have a short, less variable serine repeat, typically containing six TCA codons, but in subtype families IXa and IXb there are two blocks of additional minisatellites (12 bp and 18 bp repeats) that can differentiate the subtypes further. The 18 bp repeat differs slightly in sequence between the two families (IXa: ATTCTGGTACTGAAGATA and IXb: GGTACTGAAAATAATTCT) while the 12 bp repeat for both subtype families is GGTACTCAAGGA (Kváč et al., 2013). Multiple R repeat regions make it difficult to name, and currently, there is no standardised approach to include this in the nomenclature.

Recently, the *gp60* gene has been described in *C. suis*, and although a relatively short serine repeat unit is present, it comprises AGT/AGC repeats instead of the typical TCA/TCG/TCT (Lebbad et al., 2024). In addition to the atypical polyserine repeat, the *C. suis gp60* gene is unusual in that it is one of the longest *gp60* genes identified to date (sometimes more than 1.6 kbp, compared to about 1 kbp for *C. hominis* and *C. parvum*). This is due to the presence of a 21 bp minisatellite R repeat (GGTG[A/G/T]TCAAG[A/G]GAATGC[C/T]CAG), with examples identified in Sweden containing 37 (XXVaR37) and 39 (XXVaR39) repeats equating to 777 and 819 bp, respectively (Lebbad et al., 2024). This exceptionally long minisatellite with its high number of repeats made it difficult to design a PCR assay to cover the entire variable region. The solution was a shorter product covering the sequence flanking the minisatellite, but with some loss of discrimination, and nomenclature similar to that of *C. canis* (Jiang et al., 2021), reflecting the species (XXV), subtype family (a) and numbered variations (1–3) (e.g. XXVa1, XXVa2 and XXVa3) (Lebbad et al., 2024).

3. The standardised *gp60* nomenclature

As described, there are many features of the *gp60* nomenclature which have evolved over time and led in some instances to alternative interpretations and inconsistencies. In order to avoid misinterpretation, we propose a standardisation of the naming of newly identified subtypes that is intuitive and easy to understand. This is challenging with *gp60* and the multiple types of variation in the gene both across and within *Cryptosporidium* species. Table 2 provides a summary of the different features used in the standardised *gp60* subtyping nomenclature.

4. Analytical workflows and tools for *gp60* investigations

4.1. PCR assays and recommended primer sets

Unsurprisingly in such a highly variable gene, many primer sets have been developed for amplification from various *Cryptosporidium* species (Table 3). Some, such as those published by Alves et al. (2003), have been widely used for the past couple of decades, particularly due to their broad applicability across many species. Most of these primer sets are designed for nested PCR, which has the advantage of increasing sensitivity when amplifying a single copy gene from a parasite that can be scanty in samples, particularly as an infection begins to wane or is asymptomatic. Nevertheless, samples often escape *gp60* typing. This is likely due, not only to assay sensitivity, but also to the highly variable nature of the gene resulting in sequence variation within the primer binding regions. The disadvantages of a nested PCR approach, particularly during public health investigations or screening large numbers of samples, are the increased time, cost and human resource required to perform two rounds of PCR and gel electrophoresis for the detection of amplicons prior to sequencing. To reduce time, cost and human resources in these instances, a real-time PCR assay with good sensitivity and broad specificity across several species was developed. The assay, using a fluorescent dye-based master mix to detect positive amplification prior to Sanger sequencing, has been described and is used routinely for *gp60* subtyping of clinical specimens in England and Wales (Robinson et al., 2025).

Following successful amplification with an appropriate primer set, products are Sanger sequenced, ideally in both directions although good quality single direction sequencing can be sufficient for economical subtype characterisation. However, all sequences generated for submission to repositories such as NCBI's GenBank should be bi-directional to give a good quality consensus sequence that fully spans between the primers. Any sequences submitted to these repositories should have the primer sequences deleted, this is because mismatched nucleotides between the primers and original DNA sequence may be introduced into the consensus sequence during the PCR process.

The use of phylogenetic trees for the comparison of *gp60* sequences is useful in certain instances, particularly for species that do not have the highly variable polyserine repeat and thus sequence comparison determines the subtypes. When creating *gp60* phylogenetic trees, thought must be given to the purpose. For example, if comparing subtype families and different species, it is often beneficial to align the sequences and remove the serine microsatellite prior to tree construction due to the extent of variation that region can introduce, whereas if the purpose is to investigate subtypes of the same family the microsatellite variation is particularly important.

4.2. Reference resources and gp60 resource webpage

As the range of *Cryptosporidium* species and genotypes grows, and our knowledge of them increases, designation of some of the *gp60* families and subtypes may be readjusted. For example, the original Ic family described by Strong et al. (2000) was corrected to IIc to fall under the *C. parvum* umbrella. More recently several *gp60* families have been re-designated due to new findings or clashes in nomenclature as different groups attempted to designate the next available family. For example, Laatamna et al. (2015) and Liu et al. (2015) both published different sequences for *C. hominis* Ik in 2015, but because Laatamna et al. (2015) was published first, their nomenclature remained (GenBank: KJ941148) and that of Liu et al. (2015) (GenBank: KP314263) was re-designated as *C. hominis* family Il. Likewise, due to findings of Čondlová et al. (2019), *C. ubiquitum* XIIe and XIIf were corrected to XVIIa and XVIIIa as the specimens were re-identified as *Apodemus* genotypes I and II, respectively. This highlights another problem for researchers who rely on NCBI's GenBank repository, as updates are not often made, leaving incorrect data within the online record. For example, both KJ941148 and KP314263 described above still state subtype Ik.

Lack of curation is one of the pitfalls of using a public repository like GenBank to determine and keep track of the current nomenclature status. The community relies upon published data in periodic reviews or through personal communication with Dr Lihua Xiao, who has historically kept a master-list of the current *gp60* designations. In this publication, we provide some useful resources to the *Cryptosporidium* community including the current master-list of *gp60* nomenclature at the time of this publication (Supplementary file S3), and a weblink to a *gp60* resource page containing a curated live version of the master-list, allowing researchers to access the contemporary nomenclature (https://cryptodb.org/cryptodb/app/static-content/gp60.html). In addition to the curated table, there is a contact link on the site to allow researchers to submit information for the allocation of a new *gp60* species or family designation, thereby reducing the chance of clashes and dual publication of different sequences under the same designation.

Table 3
Selected PCR primer sets for the amplification and Sanger sequencing of the *gp60* gene in *Cryptosporidium* species.

Species amplified	Assay type	Primers	Sequence (5′–3′)	Fragment size[a] (bp)	Reference
C. hominis, *C. parvum*, (others including: *C. cuniculus*, *C. equi*, *C. wrairi*, *C. sciurinum*, *C. tyzzeri*, *C. erinacei*, *C. myocastoris* and Mink, Opossum I and Skunk genotypes)	Nested	1°F: AL3531	ATAGTCTCCGCTGTATTC	1000	Alves et al. (2003)
		1°R: AL3535	GGAAGGAACGATGTATCT		
		2°F: AL3532	TCCGCTGTATTCTCAGCC	800–850	
		2°R: AL3534	GCAGAGGAACCAGCATC		
	Nested	1°F: F1	ATGAGATTGTCGCTCATTATCG	1000	Roellig and Xiao (2020)
		1°R: R1	TTACAACACGAATAAGGCTGC		
		2°F: AL3532	TCCGCTGTATTCTCAGCC	800–850	
		2°R: R2	GGAAGGAACGATGTATCT		
	Real-time	F: AL3532	TCCGCTGTATTCTCAGCC	400–600	Robinson et al. (2025)
		R: 500R	CCAGAGATATATCTTGGTGCGGG		
C. meleagridis	Nested	1°F: CRSout115F	GATGAGATTGTCGCTCGTTATC	1100	Stensvold et al. (2014)
		1°R: CRSout1328R	AACCTGCGGAACCTGTG		
		2°F: ATGFmod	GAGATTGTCGCTCGTTATCG	900	
		2°R: GATR2	GATTGCAAAAACGGAAGG		
C. fayeri	Nested	1°F: outF	CCACACATCTGTAGCGTCGTCA	Unknown	Power et al. (2009)
		1°R: Mar4	CAGTCGTCTTAATTCACGGT		
		2°F: ATGF	ATGAGATTGTCGCTCATTATCG	1000	
		2°R: MAR3	CGTGAGAACATTCTGGAAGCT		
C. ubiquitum and *Apodemus* genotypes I and II	Nested	1°F: Ubi-18S-F1	TTTACCCACACATCTGTAGCGTCG	1044	Li et al. (2014)
		1°R: Ubi-18S-R1	ACGGACGGAATGATGTATCTGA		
		2°F: Ubi-18S-F2	ATAGGTGATAATTAGTCAGTCTTTAAT	948	
		2°R: Ubi-18S-R2	TCCAAAAGCGGCTGAGTCAGCATC		
C. mortiferum	Nested	1°F: GP60-Chipmk-F1	TTTACCCACACATCTGTAACGTCG	1072	Bujila et al. (2021); Guo et al. (2015)
		1°R: GP60-Chipmk-R1	CCTGTGAGAATATTCTGGAAATTA		
		2°F: CrChigp60Gt1Fw3	GGAAAAATGAGATTAACGCTTATC	966	
		2°R: GP60-Chipmk-R2	TACTCTTAAAACGCTTAAACTCTTAA		
C. viatorum	Nested	1°F: CviatF2	TTCATTCTGACCCCTTCATAG	1192	Stensvold et al. (2015)
		1°R: CviatR5	GTCTCCTGAATCTCTGCTTACTC		
		2°F: CviatF3	GAGATTGTCACTCATCATCGTAC	950	
		2°R: CviatR8	CTACACGTAAAATAATTCGCGAC		
C. felis	Nested	1°F: GP60CF_F1	TTTCCGTTATTGTTGCAGTTGCA	1200	Rojas-Lopez et al. (2020)
		1°R: GP60CF_R1	ATCGGAATCCCACCATCGAAC		
		2°F: GP60CF_F2	GGGCGTTCTGAAGGATGTAA	900	
		2°R: GP60CF_R2	CGGTGGTCTCCTCAGTCTTC		
C. canis	Nested	1°F: GP60-Canis-F1	ATACTCTGGTCTCCCGTTT	750	Jiang et al. (2021)
		1°R: GP60-Canis-R1	GTACTCGGAAGCGGTGTA		
		2°F: GP60-Canis-R1	AAGGCGCCTCACTCATT	700	
		2°R: GP60-Canis-R2	TCAGTTAGATATCACCCATTAA		
	Nested	1°F: Canis-Fox-gp60-F1	TTCACAGTCTACTTTGATGG	Unknown	Wang et al. (2022)
		1°R: Canis-Fox-gp60-R1	GTACTCGGAAGCGGTGTA		
		2°F: Canis-Fox-gp60-R1	GACCCGGACGTTACATTTGATGG	850	
		2°R: Canis-Fox-gp60-R2	TCAGTTAGATATCACCCATTAA		
C. ryanae	Nested	1°F: Ry-gp60-F1	GCTCGAGTTCTGAGTCGA	1068	Yang et al. (2020)
		1°R: Ry-gp60-R1	ATACCGTTAAAATGAAGGCCAA		
		2°F: Ry-gp60-F2	CCTCAGATAATGAGCAGTCTA	1024	
		2°R: Ry-gp60-R2	GATGGGATAACATATCTATAACCAAA		
C. xiaoi	Nested	1°F: Xiaoi-gp60-F1	CCTCTCGGCACTTATTGCCCT	Unknown	Fan et al. (2021)
		1°R: Xiaoi-gp60-R1	ATACCTGAGATCAAATGCTGATGAA		

(*continued on next page*)

Table 3 (*continued*)

Species amplified	Assay type	Primers	Sequence (5′–3′)	Fragment size[a] (bp)	Reference
		2° F: Xiaoi-gp60-F2	CCTCTTAGGGGTTCATTGTCTA	711–1266	
		2° R: Xiaoi-gp60-R2	TACCTTCAAAGATGACATCAC		
C. occultus	Nested	1° F: Occ-gp60-F1	ATGAGATTATCATTCATTATCGTATTG	~918	Currently unpublished
		1° R: Occ-gp60-R1	GCGAAAATTGCAAATATGGATGG		
		2° F: Occ-gp60-F2	GCTTTCCGTTATCGTCTCTGT	~893	
		2° R: Occ-gp60-R1	GCGAAAATTGCAAATATGGATGG		
C. suis	Nested[b]	1° F: GP60-Chipmk-F1	TTTACCCACACATCTGTAACGTCG	Unknown	Lebbad et al. (2024)
		1° R: GP60-Chipmk-R1	CCTGTGAGAATATTCTGGAAATTA		
		2° F: GP60-Chipmk-F2	ATAGGTAATAATTACTCAGTATTTAAT	~2000	
		2° R: GP60-Chipmk-R2	TACTCTTAAAACGCTTAAACTCTTAA		
	Nested[c]	1° F: CsuisN1F	TGCTGTTGCTACTGAAGCTAGTGG	685	Lebbad et al. (2024)
		1° R: CsuisN1R	GAAGAACGCGGCGAAAATTG		
		2° F: CsuisN2F	CTACTGATGATACAAAGAGTGC	632	
		2° R: CsuisN2R	GGATGGAATGACATATCTAAG		
C. bovis	Nested	1° F: Bovis-gp60-F1	ATGCCACTTACGCTCTACATTACTCT	Unknown	Wang et al. (2021)
		1° R: Bovis-gp60-R1	GACAAAATGAAGGCTGAGATAGATGGGA		
		2° F: Bovis-gp60-F2	CCTCTCGGCATTTATTGCCCT	1162–1292	
		2° R: Bovis-gp60-R2	ATACCTAAGGCCAAATGCTGATGAA		

[a] Estimated from data published, but due to the polymorphic and repetitive nature of the gene, the fragment sizes can be highly variable.

[b] Assay can struggle to cover the long 21 bp minisatellite repeat sequence that increases subtype discrimination.

[c] Assay limited in discriminatory power as only the post-repetitive region is amplified.

NCBI's GenBank is an immensely powerful resource, but due to the high volume of sequence data available (often multiple identical sequences from the same studies) and errors or alternative classifications within records, it can be difficult for researchers to correctly identify subtypes or have confidence in the generated results. The use of a library file containing selected known reference sequences to locally BLAST query sequences in freeware such as BioEdit (https://thalljiscience.github.io/) can provide greater confidence in manual identification. Sequences used to generate a local library should be trimmed to exclude the serine repeat, as the BLAST search is primarily to identify the subtype family. Enumeration and analysis of the serine repeat is preferably manual. Additionally, library sequences from closely related families and species should be of a similar length to ensure accurate identity matches. BLAST queries of longer sequences might more closely match a different species, due to the lower proportional difference in each nucleotide variation.

In addition to these BLAST issues, the complexities of determining unusual and variable R repeats can be challenging with many researchers seeking assistance in what, and how, to correctly identify and align these repeats. We have created a *gp60* resource page (https://cryptodb.org/cryptodb/app/static-content/gp60.html), which is hosted on the CryptoDB site, and aims to assist researchers by containing links to reference sequences and alignments demonstrating the various features identified in different *gp60* subtypes.

4.3. Software and web-based analysis

4.3.1. Software for gp60 analysis: CryptoGenotyper

To assist further, an automated tool called CryptoGenotyper was developed for analysing both *gp60* and SSU rRNA (single or mixed SSU) Sanger sequences to increase the accuracy and speed of genotyping (Yanta et al., 2021). The *gp60* part of the tool accepts ".ab1" input files and determines the sequence based on fluorescent intensities. The software includes a reference database against which the sequence is BLASTed and the number of repeat regions, species and subtypes are reported. The updated CryptoGenotyper v1.5.0 now has a new added feature for direct analysis of nucleotide sequences stored in text-based FASTA files in addition to Sanger sequencing chromatogram data files in ".ab1" format. The manually curated reference database for classifying the sequences has also been updated and now includes: the multitude of newly described species and families with distinct nomenclature; updated algorithms for repeat region detection; and the standardized nomenclature described in this paper. This tool is especially useful when processing large number of sequences.

The tool is available on several platforms, including the web-based platform Galaxy installable from the public ToolShed repository (https://toolshed.g2.bx.psu.edu/, tool id: cryptogenotyper, owner: nml), from the public Galaxy EU server (https://usegalaxy.eu/root?tool_id=CryptoGenotyper), or as a standalone Bioconda package (https://anaconda.org/bioconda/cryptogenotyper). The source code is available at https://github.com/phac-nml/CryptoGenotyper.

4.3.2. Software for detecting heterogenous populations: TIDE

TIDE (Tracking of InDels by Decomposition) is an algorithm originally developed to analyse DNA sequencing data for the identification and quantification of indels (insertions and deletions) in heterogeneous populations (Brinkman et al., 2014). Specifically, TIDE was designed to help resolve complex mixtures of DNA sequences, such as those that arise from CRISPR-Cas9 editing experiments. It decomposes the sequencing data into its constituent genotypes, allowing for precise quantification of the different alleles present in a sample, even when they are in mixed populations. The algorithm was shown to be useful when repurposed as a bioinformatics tool for interpreting *Cryptosporidium gp60* sequences, including those with only archived chromatograms available (Dettwiler et al., 2022). Based on chromatograms from the Sanger sequences of *Cryptosporidium gp60*, the tool identified and

quantified mixed infections and could thereby record mixes that may otherwise be missed. In addition, the tool also detected PCR stutter artifacts that may arise during the elongation step of the amplification. As expected, the stutter artifacts were particularly evident in sequences with longer repeat regions. This tool is useful if manual sequence analysis suggests that mixed populations may be present in your sample and is available online at https://tide.nki.nl.

4.3.3. Future scope for a Cryptosporidium gp60 database in PubMLST.org

A useful future addition to the tools used for analysis and comparison of *Cryptosporidium gp60* subtypes might be a database hosted on PubMLST.org. This site hosts public databases for molecular typing and microbial genome diversity that integrate sequence data with sample provenance and other metadata (such as patient demographics). Inclusion would enable the sharing of *gp60* data with public and animal health professionals and allow for data comparisons between countries, for example, as an overview of trends or if there are widespread outbreaks. As a curated database, the quality control of included data should be more reliable.

5. Recommendations

To assist in *gp60* genotyping and standardise the nomenclature we suggest the following recommendations.

- Select the appropriate assay and primers for the species, study or purpose in question (Table 3)
- Use the reference resources and tools available on the *gp60* Resource Page (https://cryptodb.org/cryptodb/app/static-content/gp60.html), including CryptoGenotyper (https://usegalaxy.eu/root?tool_id=CryptoGenotyper), to assist with *gp60* genotyping while ensuring correct contemporary use of nomenclature of isolates is used:
  - o Use the reference sequences and libraries provided for comparisons
  - o Use the reference alignments provided to identify complicated R repeat sequences
  - o Check the rules for different families or species, as they differ
  - o Write out in full Greek alphabet family names to avoid confusion with Latin letters in some fonts
- Use amino acid translations to assist with correct alignment and identification of new repeat units
- Refer to the *gp60* Master-list on the *gp60* Resource Page to identify the subtype family, or to determine new subtype families or species designations, and use the contact link to notify curators of any novel designations. This will ensure curation and avoid duplication thus preserving the nomenclature
- Submit updates of any additional features discovered, for example, the discovery of new R repeat variation, to the *gp60* Master-list by using the contact link on the *gp60* Resource page
- Update or correct erroneous records on NCBI's GenBank if new information becomes available

6. Conclusions

The *gp60* gene has been widely used for over 25 years to subtype *Cryptosporidium* species and has been useful in the epidemiological investigation of these parasites, particularly *C. hominis* and *C. parvum*. Although it is not a perfect subtyping scheme, partly due to the limitations of using a single locus to differentiate a sexually recombinant parasite, it continues to provide useful data for comparing isolates and characterising outbreaks. The *gp60* scheme has continued to evolve for an increasing number of species including those where the features used to name the subtypes differ substantially. This, along with different interpretations of the nomenclature and historical sequence errors remaining on the NCBI GenBank repository, has made characterising subtypes more complicated. Here we have attempted to outline the details and rules for *gp60* subtyping into a single place and provide useful resources to standardise the nomenclature, which will hopefully assist the community. Undoubtedly, the increased use of multi-locus and genomic methods will become more common place in epidemiological investigations, but the current limitations of these methods indicate that *gp60* subtyping will continue to be useful for the foreseeable future.

CRediT authorship contribution statement

Guy Robinson: Conceptualization, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Rachel M. Chalmers:** Writing – review & editing. **Kristin Elwin:** Writing – review & editing. **Rebecca A. Guy:** Resources, Visualization, Software, Writing – original draft, Writing – review & editing. **Kyrylo Bessonov:** Resources, Software, Writing – review & editing. **Karin Troell:** Data curation, Writing – original draft, Writing – review & editing. **Lihua Xiao:** Data curation, Visualization, Writing – review & editing.

Ethical approval

Not applicable.

Data availability

The data supporting the conclusions of this article are included within the article and its supplementary files.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Dr Omar Harb and Dr Nupur Kittur from the VEuPathDB project team for hosting and facilitating the set-up of the gp60 Resource web page within CryptoDB.org; Dr Harriet Risby from the Cryptosporidium Reference Unit for assistance in the generation of the graphical abstract; Harri Ahola from the Swedish Veterinary Agency for his work on the *Cryptosporidium* occultus primers and sequences.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crpvbd.2025.100257.

References

Alsmark, C., Nolskog, P., Angervall, A.L., Toepfer, M., Winiecka-Krusnell, J., Bouwmeester, J., et al., 2018. Two outbreaks of cryptosporidiosis associated with cattle spring pasture events. Vet. Parasitol. Reg. Stud. rep. 14, 71–74. https://doi.org/10.1016/j.vprsr.2018.09.003.

Alves, M., Xiao, L., Sulaiman, I., Lal, A.A., Matos, O., Antunes, F., 2003. Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal. J. Clin. Microbiol. 41, 2744–2747. https://doi.org/10.1128/JCM.41.6.2744-2747.2003.

Amer, S., Zidan, S., Adamu, H., Ye, J., Roellig, D., Xiao, L., Feng, Y., 2013. Prevalence and characterization of *Cryptosporidium* spp. in dairy cattle in Nile River Delta provinces, Egypt. Exp. Parasitol. 135, 518–523. https://doi.org/10.1016/j.exppara.2013.09.002.

Banda, B., Siwila, J., Mukubesa, A.N., Chitanga, S., Kaonga, P., Changula, K., et al., 2022. Cryptosporidiosis is predominantly an urban, anthroponotic infectious disease among Zambian children. Trans. R. Soc. Trop. Med. Hyg. 116, 270–277. https://doi.org/10.1093/trstmh/trab121.

Benson, G., 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. https://doi.org/10.1093/nar/27.2.573.

Brinkman, E.K., Chen, T., Amendola, M., Van Steensel, B., 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 42, e168. https://doi.org/10.1093/NAR/GKU936.

Bujila, I., Troell, K., Fischerström, K., Nordahl, M., Killander, G., Hansen, A., et al., 2021. *Cryptosporidium* chipmunk genotype I – an emerging cause of human cryptosporidiosis in Sweden. Infect. Genet. Evol. 92, 104895. https://doi.org/10.1016/J.MEEGID.2021.104895.

Bujila, I., Ohlson, A., Hansen, A., Agudelo, Lady, Kühlmann-Berenzon, S., Galanis, I., et al., 2024a. Outbreak of the novel *Cryptosporidium parvum* IIγA11 linked to salad bars in Sweden, December 2023. Epidemiol. Infect. 152, e140. https://doi.org/10.1017/S0950268824001432.

Bujila, I., Troell, K., Ögren, J., Hansen, A., Killander, G., Agudelo, L., et al., 2024b. *Cryptosporidium* species and subtypes identified in human domestic cases through the national microbiological surveillance programme in Sweden from 2018 to 2022. BMC Infect. Dis. 24, 146. https://doi.org/10.1186/s12879-024-09049-x.

Cama, V.A., Ross, J.M., Crawford, S., Kawai, V., Chavez-Valdez, R., Vargas, D., et al., 2007. Differences in clinical manifestations among *Cryptosporidium* species and subtypes in HIV-infected persons. J. Infect. Dis. 196, 684–691. https://doi.org/10.1086/519842.

Chalmers, R.M., Robinson, G., Elwin, K., Hadfield, S.J., Xiao, L., Ryan, U., et al., 2009. *Cryptosporidium* sp. rabbit genotype, a newly identified human pathogen. Emerg. Infect. Dis. 15, 829–830. https://doi.org/10.3201/eid1505.081419.

Chalmers, R.M., Robinson, G., Elwin, K., Elson, R., 2019. Analysis of the *Cryptosporidium* spp. and *gp60* subtypes linked to human outbreaks of cryptosporidiosis in England and Wales, 2009 to 2017. Parasites Vectors 12, 95. https://doi.org/10.1186/s13071-019-3354-6.

Chen, L., Hu, S., Jiang, W., Zhao, J., Li, N., Guo, Y., et al., 2019. *Cryptosporidium parvum* and *Cryptosporidium hominis* subtypes in crab-eating macaques. Parasites Vectors 12, 350. https://doi.org/10.1186/s13071-019-3604-7.

Čondlová, Š., Horčičková, M., Havrdová, N., Sak, B., Hlásková, L., Perec-Matysiak, A., et al., 2019. Diversity of *Cryptosporidium* spp. in *Apodemus* spp. in Europe. Eur. J. Protistol. 69, 1–13. https://doi.org/10.1016/J.EJOP.2019.02.005.

Dettwiler, I., Troell, K., Robinson, G., Chalmers, R.M., Basso, W., Rentería-Solís, Z.M., et al., 2022. TIDE analysis of *Cryptosporidium* infections by *gp60* typing reveals obscured mixed infections. J. Infect. Dis. 225, 686–695. https://doi.org/10.1093/INFDIS/JIAB417.

Fan, Y., Huang, X., Guo, S., Yang, F., Yang, X., Guo, Y., et al., 2021. Subtyping *Cryptosporidium xiaoi*, a common pathogen in sheep and goats. Pathogens 10, 800. https://doi.org/10.3390/pathogens10070800.

Feng, Y., Lal, A.A., Li, N., Xiao, L., 2011. Subtypes of *Cryptosporidium* spp. in mice and other small mammals. Exp. Parasitol. 127, 238–242. https://doi.org/10.1016/j.exppara.2010.08.002.

Feng, Y., Ryan, U.M., Xiao, L., 2018. Genetic diversity and population structure of *Cryptosporidium*. Trends Parasitol. 34, 997–1011. https://doi.org/10.1016/j.pt.2018.07.009.

Gilchrist, C.A., Cotton, J.A., Burkey, C., Arju, T., Gilmartin, A., Lin, Y., et al., 2018. Genetic diversity of *Cryptosporidium hominis* in a Bangladeshi community as revealed by whole-genome sequencing. J. Infect. Dis. 218, 259–264. https://doi.org/10.1093/infdis/jiy121.

Glaberman, S., Sulaiman, I.M., Bern, C., Levior, J., Peng, M.M., Morgan, U., et al., 2001. A multilocus genotypic analysis of *Cryptosporidium meleagridis*. J. Eukaryot. Microbiol. 19S–22S. https://doi.org/10.1111/J.1550-7408.2001.TB00439.X. Suppl.

Guo, Y., Cebelinski, E., Matusevich, C., Alderisio, K.A., Lebbad, M., McEvoy, J., et al., 2015. Subtyping novel zoonotic pathogen *Cryptosporidium* chipmunk genotype I. J. Clin. Microbiol. 53, 1648–1654. https://doi.org/10.1128/JCM.03436-14.

Huang, W., He, W., Huang, Y., Tang, Y., Chen, M., Sun, L., et al., 2024. Multicopy subtelomeric genes underlie animal infectivity of divergent *Cryptosporidium hominis* subtypes. Nat. Commun. 15, 10774. https://doi.org/10.1038/S41467-024-54995-4.

Innes, E.A., Chalmers, R.M., Wells, B., Pawlowic, M.C., 2020. A One Health approach to tackle cryptosporidiosis. Trends. Parasitol. 36, 290–303. https://doi.org/10.1016/j.pt.2019.12.016.

Jex, A.R., Whipp, M., Campbell, B.E., Cacciò, S.M., Stevens, M., Hogg, G., Gasser, R.B., 2007. A practical and cost-effective mutation scanning-based approach for investigating genetic variation in *Cryptosporidium*. Electrophoresis 28, 3875–3883. https://doi.org/10.1002/elps.200700279.

Jex, A.R., Gasser, R.B., 2008. Analysis of the genetic diversity within *Cryptosporidium hominis* and *Cryptosporidium parvum* from imported and autochthonous cases of human cryptosporidiosis by mutation scanning. Electrophoresis 29, 4119–4129. https://doi.org/10.1002/elps.200800422.

Jex, A.R., Gasser, R.B., 2010. Genetic richness and diversity in *Cryptosporidium hominis* and *C. parvum* reveals major knowledge gaps and a need for the application of "next generation" technologies - research review. Biotechnol. Adv. 28, 17–26. https://doi.org/10.1016/j.biotechadv.2009.08.003.

Ježková, J., Limpouchová, Z., Prediger, J., Holubová, N., Sak, B., Konečný, R., et al., 2021. *Cryptosporidium myocastoris* n. sp. (Apicomplexa: Cryptosporidiidae), the species adapted to the nutria (*Myocastor coypus*). Microorganisms 9, 813. https://doi.org/10.3390/MICROORGANISMS9040813.

Jiang, W., Roellig, D.M., Lebbad, M., Beser, J., Troell, K., Guo, Y., et al., 2020. Subtype distribution of zoonotic pathogen *Cryptosporidium felis* in humans and animals in several countries. Emerg. Microb. Infect. 9, 2446–2454. https://doi.org/10.1080/22221751.2020.1840312.

Jiang, W., Roellig, D.M., Guo, Y., Li, N., Feng, Y., Xiao, L., 2021. Development of a subtyping tool for zoonotic pathogen *Cryptosporidium canis*. J. Clin. Microbiol. 59, e02474. https://doi.org/10.1128/JCM.02474-20, 20.

Kváč, M., McEvoy, J., Loudová, M., Stenger, B., Sak, B., Květoňová, D., et al., 2013. Coevolution of *Cryptosporidium tyzzeri* and the house mouse (*Mus musculus*). Int. J. Parasitol. 43, 805–817. https://doi.org/10.1016/J.IJPARA.2013.04.007.

Laatamna, A.E., Wagnerová, P., Sak, B., Květoňová, D., Aissi, M., Rost, M., Kváč, M., 2013. Equine cryptosporidial infection associated with *Cryptosporidium* hedgehog genotype in Algeria. Vet. Parasitol. 197, 350–353. https://doi.org/10.1016/J.VETPAR.2013.04.041.

Laatamna, A.E., Wagnerová, P., Sak, B., Květoňová, D., Xiao, L., Rost, M., et al., 2015. Microsporidia and *Cryptosporidium* in horses and donkeys in Algeria: Detection of a novel *Cryptosporidium hominis* subtype family (Ik) in a horse. Vet. Parasitol. 208, 135–142. https://doi.org/10.1016/J.VETPAR.2015.01.007.

Lebbad, M., Winiecka-Krusnell, J., Insulander, M., Beser, J., 2018. Molecular characterization and epidemiological investigation of *Cryptosporidium hominis* IkA18G1 and *C. hominis* monkey genotype IiA17, two unusual subtypes diagnosed in Swedish patients. Exp. Parasitol. 188, 50–57. https://doi.org/10.1016/J.EXPPARA.2018.03.006.

Lebbad, M., Winiecka-Krusnell, J., Stensvold, C.R., Beser, J., 2021. High diversity of *Cryptosporidium* species and subtypes identified in cryptosporidiosis acquired in Sweden and abroad. Pathogens 10, 523. https://doi.org/10.3390/pathogens10050523.

Lebbad, M., Grüttner, J., Beser, J., Lizana, V., Dea-Ayuela, M.A., Oropeza-Moe, M., et al., 2024. Complete sequencing of the *Cryptosporidium suis gp60* gene reveals a novel type of tandem repeats - implications for surveillance. Infect. Genet. Evol. 122, 105614. https://doi.org/10.1016/j.meegid.2024.105614.

Levine, M.M., Nasrin, D., Acácio, S., Bassat, Q., Powell, H., Tennant, S.M., et al., 2020. Diarrhoeal disease and subsequent risk of death in infants and children residing in low-income and middle-income countries: analysis of the GEMS case-control study and 12-month GEMS-1A follow-on study. Lancet Glob. Health 8, e204–e214. https://doi.org/10.1016/S2214-109X(19)30541-8.

Li, N., Xiao, L., Alderisio, K., Elwin, K., Cebelinski, E., Chalmers, R., et al., 2014. Subtyping *Cryptosporidium ubiquitum*, a zoonotic pathogen emerging in humans. Emerg. Infect. Dis. 20, 217–224. https://doi.org/10.3201/eid2002.121797.

Li, M., Yang, F., Hou, T., Gong, X., Li, N., Sibley, L.D., et al., 2024. Variant surface protein *GP60* contributes to host infectivity of *Cryptosporidium parvum*. Commun. Biol. 7, 1175. https://doi.org/10.1038/s42003-024-06885-0.

Liu, X., Xie, N., Li, W., Zhou, Z., Zhong, Z., Shen, L., et al., 2015. Emergence of *Cryptosporidium hominis* monkey genotype II and novel subtype family Ik in the squirrel monkey (*Saimiri sciureus*) in China. PLoS One 10, e0141450. https://doi.org/10.1371/JOURNAL.PONE.0141450.

Lv, C., Zhang, L., Wang, R., Jian, F., Zhang, S., Ning, C., et al., 2009. *Cryptosporidium* spp. in wild, laboratory, and pet rodents in China: Prevalence and molecular characterization. Appl. Environ. Microbiol. 75, 7692–7699. https://doi.org/10.1128/AEM.01386-09.

Misic, Z., Abe, N., 2007. Subtype analysis of *Cryptosporidium parvum* isolates from calves on farms around Belgrade, Serbia and Montenegro, using the 60 kDa glycoprotein gene sequences. Parasitology 134, 351–358. https://doi.org/10.1017/S0031182006001508.

Power, M.L., Cheung-Kwok-Sang, C., Slade, M., Williamson, S., 2009. *Cryptosporidium fayeri*: Diversity within the *GP60* locus of isolates from different marsupial hosts. Exp. Parasitol. 121, 219–223. https://doi.org/10.1016/J.EXPPARA.2008.10.016.

Robinson, G., Elwin, K., Chalmers, R.M., 2025. A real-time PCR based method for the sequence determination of *Cryptosporidium hominis* and *Cryptosporidium parvum gp60* subtypes during epidemiological investigations. In: Mead, J.R., Arrowood, M.J. (Eds.), *Cryptosporidium*. Methods in Molecular Biology, 2nd ed. Humana, New York (in press).

Roellig, D.M., Xiao, L., 2020. *Cryptosporidium* genotyping for epidemiology tracking. In: *Cryptosporidium*. In: Mead, J.R., Arrowood, M.J. (Eds.), Methods Mol. Biol, vol. 2052. Humana, New York, pp. 103–116. https://doi.org/10.1007/978-1-4939-9748-0_7.

Rojas-Lopez, L., Elwin, K., Chalmers, R.M., Enemark, H.L., Beser, J., Troell, K., 2020. Development of a *gp60*-subtyping method for *Cryptosporidium felis*. Parasites Vectors 13, 39. https://doi.org/10.1186/s13071-020-3906-9.

Ryan, U.M., Feng, Y., Fayer, R., Xiao, L., 2021a. Taxonomy and molecular epidemiology of *Cryptosporidium* and *Giardia* – a 50-year perspective (1971–2021). Int. J. Parasitol. 51, 1099–1119. https://doi.org/10.1016/j.ijpara.2021.08.007.

Ryan, U., Zahedi, A., Feng, Y., Xiao, L., 2021b. An update on zoonotic *Cryptosporidium* species and genotypes in humans. Animals 11, 3307. https://doi.org/10.3390/ani11113307.

Silverlås, C., Näslund, K., Björkman, C., Mattsson, J.G., 2010. Molecular characterisation of *Cryptosporidium* isolates from Swedish dairy cattle in relation to age, diarrhoea and region. Vet. Parasitol. 169, 289–295. https://doi.org/10.1016/j.vetpar.2010.01.003.

Silverlås, C., Bosaeus-Reineck, H., Näslund, K., Björkman, C., 2013. Is there a need for improved *Cryptosporidium* diagnostics in Swedish calves? Int. J. Parasitol. 43, 155–161. https://doi.org/10.1016/j.ijpara.2012.10.009.

Stensvold, C.R., Beser, J., Axén, C., Lebbad, M., 2014. High applicability of a novel method for *gp60*-based subtyping of *Cryptosporidium meleagridis*. J. Clin. Microbiol. 52, 2311–2319. https://doi.org/10.1128/JCM.00598-14.

Stensvold, C.R., Elwin, K., Winiecka-Krusnell, J., Chalmers, R.M., Xiao, L., Lebbad, M., 2015. Development and application of a *gp60*-based typing assay for *Cryptosporidium viatorum*. J. Clin. Microbiol. 53, 1891–1897. https://doi.org/10.1128/JCM.00313-15.

Strong, W.B., Gut, J., Nelson, R.G., 2000. 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. Infect. Immun. 68, 4117–4134. https://doi.org/10.1128/IAI.68.7.4117-4134.2000.

Sulaiman, I.M., Hira, P.R., Zhou, L., Al-Ali, F.M., Al-Shelahi, F.A., Shweiki, H.M., et al., 2005. Unique endemicity of cryptosporidiosis in children in Kuwait. J. Clin. Microbiol. 43, 2805–2809. https://doi.org/10.1128/JCM.43.6.2805-2809.2005.

Toriro, R., Pallett, S., Woolley, S., Bennett, C., Hale, I., Heylings, J., et al., 2024. Outbreak of diarrhea caused by a novel *Cryptosporidium hominis* subtype during British military training in Kenya. Open Forum Infect. Dis. 11, ofae001. https://doi.org/10.1093/ofid/ofae001.

Wang, W., Wan, M., Yang, F., Li, N., Xiao, L., Feng, Y., Guo, Y., 2021. Development and application of a *gp60*-based subtyping tool for *Cryptosporidium bovis*. Microorganisms 9, 2067. https://doi.org/10.3390/microorganisms9102067.

Wang, W., Wei, Y., Cao, S., Wu, W., Zhao, W., Guo, Y., Xiao, L., Feng, Y., Li, N., 2022. Divergent *Cryptosporidium* species and host-adapted *Cryptosporidium canis* subtypes in farmed minks, raccoon dogs and foxes in Shandong, China. Front. Cell. Infect. Microbiol. 12, 980917. https://doi.org/10.3389/FCIMB.2022.980917.

Xiao, L., Hlavsa, M.C., Yoder, J., Ewers, C., Dearen, T., Yang, W., et al., 2009. Subtype analysis of *Cryptosporidium* specimens from sporadic cases in Colorado, Idaho, New Mexico, and Iowa in 2007: Widespread occurrence of one *Cryptosporidium hominis* subtype and case history of an infection with the *Cryptosporidium* horse genotype. J. Clin. Microbiol. 47, 3017–3020. https://doi.org/10.1128/JCM.00226-09.

Xiao, L., Feng, Y., 2017. Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and *Giardia duodenalis*. Food Waterborne Parasitol. 8–9, 14–32. https://doi.org/10.1016/j.fawpar.2017.09.002.

Yan, W., Alderisio, K., Roellig, D.M., Elwin, K., Chalmers, R.M., Yang, F., et al., 2017. Subtype analysis of zoonotic pathogen *Cryptosporidium* skunk genotype. Infect. Genet. Evol. 55, 20–25. https://doi.org/10.1016/j.meegid.2017.08.023.

Yang, X., Huang, N., Jiang, W., Wang, X., Li, N., Guo, Y., et al., 2020. Subtyping *Cryptosporidium ryanae*: A common pathogen in bovine animals. Microorganisms 8, 1107. https://doi.org/10.3390/microorganisms8081107.

Yang, X., Guo, Y., Xiao, L., Feng, Y., 2021. Molecular epidemiology of human cryptosporidiosis in low- and middle-income countries. Clin. Microbiol. Rev. 34, e00087. https://doi.org/10.1128/CMR.00087-19, 19.

Yanta, C.A., Bessonov, K., Robinson, G., Troell, K., Guy, R.A., 2021. CryptoGenotyper: A new bioinformatics tool for rapid *Cryptosporidium* identification. Food Waterborne Parasitol. 23, e00115. https://doi.org/10.1016/j.fawpar.2021.e00115.

Zahedi, A., Ryan, U., 2020. *Cryptosporidium* – An update with an emphasis on foodborne and waterborne transmission. Res. Vet. Sci. 132, 500–512. https://doi.org/10.1016/j.rvsc.2020.08.002.