Received: 29 June 2023 | Revised: 16 August 2023 | Accepted: 21 August 2023

DOI: 10.1111/jeu.12998

The Journal of Eukaryotic Microbiology ISOP Published by the International Society of Protistologists

ORIGINAL ARTICLE

Division of *Blastocystis* ST10 into three new subtypes: ST42-ST44

Monica Santin[1] | **Ana Figueiredo**[2,3] | **Aleksey Molokin**[1] | **Nadja S George**[1] | **Pamela C. Köster**[4] | **Alejandro Dashti**[4] | **David González-Barrio**[4] | **David Carmena**[4,5] | **Jenny G. Maloney**[1]

[1]Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland, USA

[2]Department of Biology and CESAM, University of Aveiro, Aveiro, Portugal

[3]Department of Bioscience, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

[4]Parasitology Reference and Research Laboratory, Spanish National Centre for Microbiology, Madrid, Spain

[5]CIBER Infectious Diseases (CIBERINFEC), Health Institute Carlos III, Madrid, Spain

Correspondence
Jenny G. Maloney and Monica Santin, Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705, USA.
Email: jenny.maloney@usda.gov; monica.santin-duran@usda.gov

Funding information
Centre for Environmental and Marine Studies, Grant/Award Number: UIDP/50017/2020+UIDB/50017/2020+LA/P/0094/2020; COST Action, Grant/Award Number: CA21105; Fundação para a Ciência e Tecnologia, Grant/Award Number: SFRH/BD/144582/2019; Spanish Ministry of Science, Innovation and Universities, Grant/Award Number: CD19CIII/00011; USDA-ARS, Grant/Award Number: 8042-32000-112-00-D

Abstract

The *Blastocystis* subtype ST10 has been recognized to contain a great deal of diversity at the sequence level, potentially indicating the presence of multiple new STs within the clade. However, the data needed to validate these new STs were not available. To help resolve this diversity, full-length small subunit (*SSU*) rRNA gene reference sequences were generated using Oxford Nanopore MinION long-read sequencing from 21 samples representing multiple domestic and wild hosts and geographic regions and covering the sequence diversity previously described using fragments of the *SSU* rRNA gene. Phylogenetic and pairwise distance analyses were used to compare full-length sequences of the *SSU* rRNA gene generated in this study with all other valid STs of *Blastocystis*. We present data supporting the division of ST10/ST23 cluster into five subtypes, ST10, ST23, and three new subtypes with the proposed ST designations of ST42, ST43, and ST44. As the host range of *Blastocystis* continues to expand with new subtypes and new hosts being frequently identified, the reference sequences provided in this study will assist in accurate sequence classification and help to clarify the epidemiology of this common intestinal microeukaryote.

KEYWORDS
Blastocystis, MinION, small subunit ribosomal RNA gene, ST10, ST23, ST42, ST43, ST44, subtype

INTRODUCTION

BLASTOCYSTIS is one of the most common intestinal microeukaryotes in humans and animals worldwide. It has a fecal-oral transmission route and may be spread through contaminated food and water (Hublin et al., 2021). Infection with *Blastocystis* has been associated with a variety of symptoms in humans, including diarrhea, abdominal pain, urticaria (hives), irritable bowel syndrome, and in rare instances, invasive infection (Bahrami et al., 2020; Cifre et al., 2018; Fréalle et al., 2015; Tan, 2008). Asymptomatic infections are common and may be either acute or chronic in both humans and other animals (Hernández-Castro, Dashti, et al., 2023; Santin et al., 2023). However, factors that determine infection outcomes are not yet defined.

Blastocystis is divided into subtypes (STs) based on full-length sequences of the small subunit ribosomal

---

RNA (*SSU* rRNA) gene; genetic diversity within the *Blastocystis* species complex at this locus is high. Currently, 37 subtypes have been established from humans and other avian and mammalian species with ST1-ST17, ST21, and ST23-ST41 considered valid STs (Hernández-Castro, Maloney, et al., 2023; Maloney et al., 2023; Maloney & Santin, 2021; Stensvold et al., 2023; Stensvold & Clark, 2020; Yu et al., 2023). *Blastocystis SSU* rRNA gene sequence variants isolated from poikilothermic species are not currently included in the *Blastocystis* subtyping system (Stensvold et al., 2007).

As full-length *Blastocystis SSU* rRNA gene sequences are produced from more hosts around the world, it has assisted in clarifying the status of sequence variants that were originally classified using partial gene regions or incomplete data (Maloney & Santin, 2021). For example, the availability of a full-length *SSU* rRNA gene sequence for ST24 has assisted in identifying a sequence from a mouflon (*Ovis orientalis musimon*) previously identified as ST14 that should, in fact, be classified as ST24 (Stensvold et al., 2023). Such misclassifications appear to be common in GenBank and occur when partial gene sequences are originally incorrectly assigned to a given ST and then used in future reports in the classification of new gene sequences via BLAST, perpetuating misclassifications. Thus, full-length reference sequences are proving essential in the accurate naming and identification of *Blastocystis* STs.

In recent studies of *Blastocystis* ST diversity among wild and domestic ruminant species, it has been observed that wide genetic variation exists within ST10 (Higuera et al., 2021; Maloney et al., 2021; Maloney, Lombard, et al., 2019; Maloney, Molokin, & Santin, 2019; Santin et al., 2023; Stensvold et al., 2023). In fact, current data suggest that ST10 likely should be divided into at least two different subtypes (Santin et al., 2023; Stensvold et al., 2023). However, a comprehensive analysis of full-length reference sequences to confirm the validity of potential new subtypes has not been performed within the ST10 clade. In the present study, Oxford Nanopore sequencing was used to produce full-length *SSU* rRNA gene reference sequences from 21 isolates known to contain sequence variants representing potentially novel subtypes within the ST10 clade. Then, full-length sequences produced in this study were compared to other full-length reference sequences classified as ST10 to determine the support for the division of ST10 into multiple valid subtypes and subgroups.

MATERIALS AND METHODS

Source of *Blastocystis* isolates

Twenty-one DNA samples containing *Blastocystis* obtained from livestock, cattle (*Bos taurus*), goat (*Capra hircus*), and sheep (*Ovis aries*) and wildlife, red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) fecal isolates were used in this study (Table 1). Prior to inclusion in the study, all samples were screened for the presence of sequence variants within the ST10 clade using a next generation sequencing protocol as previously described (Maloney, Molokin, & Santin, 2019).

PCR amplification, sequencing, and bioinformatic analysis to generate *Blastocystis* full-length *SSU* rRNA gene sequences

Samples containing ST10 and potentially novel genetic variants of *Blastocystis* were subjected to a Nanopore sequencing strategy previously described, with minor modifications, to obtain the full-length nucleotide sequence of the *SSU* rRNA gene (Maloney et al., 2020). Briefly, PCR was performed using the MinION-tailed primers forward (5′-TTT CTG TTG GTG CTG ATA TTG C AAC CTG GTT GAT CCT GCC AGT AGT C-3′) and reverse (5′-ACT TGC CTG TCG CTC TAT CTT C TGA TCC TTC TGC AGG TTC ACC TAC G-3′) (MinION adapter nucleotide sequences underlined), which amplify most eukaryotic organisms' full-length *SSU* rRNA gene sequence, and the high-fidelity proofreading polymerase contained in KAPA HiFi HotStart ReadyMix (KAPA BioSystems). Cycling conditions included initial denaturation at 98°C for 5 min, followed by 35 cycles of amplification (20 s at 98°C, 45 s at 60°C, and 90 s at 72°C), and a final extension at 72°C for 5 min. PCR amplicons were purified using a 0.5× AMPure XP beads (Beckman Coulter) to sample ratio and quantified on a Qubit fluorimeter (ThermoFisher Scientific). Nanopore sequencing libraries were prepared using the Oxford Nanopore Technologies (ONT) SQK-LSK112/114 Ligation Sequencing Kits and EXP-PBC001 PCR Barcoding Kit (ONT) following the manufacturer's protocol for Ligation Sequencing Amplicons—PCR Barcoding (SQK-LSK112/114 with EXP-PBC001). Final libraries were loaded onto FLO-MIN112/114) flow cells and sequenced on a MinION Mk1C.

Basecalling was performed using Guppy v6.1.2 and the super accuracy model. FASTQ reads were length filtered to retain reads between 1600 and 2000 nucleotides and then aligned against NCBI's nt database using megablast (blast+ v2.13.0). Reads with best hits to *Blastocystis* references were extracted. MinION adapter sequences were trimmed, and only reads containing intact forward and reverse primers were kept. Reads were dereplicated and clustered using VSEARCH (vsearch v2.15.1) at a 98%–99% identity threshold and checked for chimeras. Clusters were then polished with Medaka v1.6.0 and reclustered once more. The full-length nucleotide sequences of the *SSU* rRNA gene generated were deposited in GenBank under the accession numbers OR117623-OR117713.

TABLE 1 *Blastocystis* specimens used in this study including host, geographic origin, and subtype(s) identified.

Isolate ID	Host	Location	Subtype(s) identified (GenBank accession number/s)[a]
2691 16m	Cattle	USA	ST10a (OR117623)/ST25 (OR117685)/ST26 (OR117694)/ST42a (OR117628)
2696 8m	Cattle	USA	ST14 (OR117668)/ST24a (OR117675)/ST42a (OR117627)/ST44 (OR117624)
2818 5m1	Cattle	USA	ST44 (OR117625)
2696 18m	Cattle	USA	ST25 (OR117686)/ST26 (OR117695)/ST42a (OR117626)
2691 24m	Cattle	USA	ST42a (OR117629)
2695 17m	Cattle	USA	ST26 (OR117696)/ST42a (OR117630)
2695 18m	Cattle	USA	ST25 (OR117687)/ST26 (OR117697)/ST42b (OR117631)
2697 24m	Cattle	USA	ST25 (OR117688)/ST26 (OR117698)/ST42b (OR117632)
P423	Cattle	Portugal	ST25 (OR117705)/ST26 (OR117706)/ST42b (OR117655)
P428	Cattle	Portugal	ST42b (OR117656)
P431	Cattle	Portugal	ST42b (OR117637)/ST21 (OR117707)
P529	Cattle	Portugal	ST42a (OR117654)
P445	Goat	Portugal	ST10a (OR117633)/ST10b (OR117640)/ST14 (OR117669)/ST21 (OR117672)/ST23 (OR117666)/ST24b (OR117681)/ST26 (OR117699)/ST43 (OR117658)/ST44 (OR117648)
P446	Goat	Portugal	ST10a (OR117634)/ST10b (OR117641)/ST23 (OR117667)/ST24a (OR117676)/ST24b (OR117682)/ST24c (OR117679)/ST25 (OR117689)/ST30 (OR117702)/ST43 (OR117659)/ ST44 (OR117649)
S282	Red deer	Spain	ST25 (OR117690)/ST42a (OR117653)
P458	Roe deer	Portugal	ST10b (OR117642)
P508	Sheep	Portugal	ST10a (OR117635)/ST10b (OR117643)/ST21 (OR117691)
P510	Sheep	Portugal	ST10a (OR117636)/ST10b (OR117644)/ST14 (OR117670)/ST21 (OR117673)/ST24a (OR117677)/ST24b (OR117684)/ST23 (OR117665)/ST25 (OR117692)/ST26 (OR117700)/ ST30 (OR117703)/ST43 (OR117660)/ST44 (OR117650)
P516	Sheep	Portugal	ST10a (OR117637)/ST10b (OR117645)/ST14 (OR117671)/ST21 (OR117674)/ST23 (OR117664)/ST24a (OR117678)/ST24c (OR117680)/ST25 (OR117693)/ST26 (OR117701)/ ST30 (OR117704)/ST43 (OR117661)/ST44 (OR117651)
P531	Sheep	Portugal	ST10a (OR117638)/ST10b (OR117646)/ST14 (OR117708)/ST24a (OR117710)/ST24b (OR117709)/ST43 (OR117662)/ST44 (OR117652)
P534	Sheep	Portugal	ST10a (OR117639)/ST10b (OR117647)/ST14 (OR117711)/ST24a (OR117713)/ST24b (OR117712)/ST43 (OR117663)

[a] ST42a, ST42b, and ST44 were previously reported as ST10d, ST10e, and ST10c, respectively.

Phylogenetic analysis

A phylogenetic tree was generated using (1) full-length *SSU* rRNA gene nucleotide sequences obtained in this study; (2) full-length *Blastocystis* reference nucleotide sequences for all currently validated STs obtained from GenBank; and (3) full-length *Proteromonas lacertae*, a Stramenopile (which is closely related to *Blastocystis*) as an outgroup (Table 2). Sequences were aligned with the Clustal W algorithm using MEGA 11, and the phylogenetic tree was artificially rooted using *P. lacertae* as an outgroup. Phylogenetic analyses were performed using the neighbor-joining (NJ) and Maximum Likelihood (ML) methods, and pairwise distances calculated with the Kimura 2-parameter model using MEGA 11. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1982 positions in the final dataset. Bootstrapping with 1000 replicates was used to determine support for the clades generated.

RESULTS

To determine if full-length *SSU* rRNA gene sequences supported the reclassification of sequences classified as ST10 based on partial gene sequence, a nanopore sequencing strategy was used to obtain full-length sequences of the *SSU* rRNA gene from 21 samples that contained ST10 subgroup sequences based on Illumina sequencing. Detection of mixed subtypes was common among the samples. A total of 91 full-length sequences were obtained from the 21 *Blastocystis*-positive samples representing ST10 subgroups previously described as ST10a (8 sequences), ST10b (8 sequences), ST10c (7 sequences), ST10d (7 sequences), and ST10e (5 sequences), an unpublished subgroup ST10f (6 sequences), ST14 (6 sequences), ST21 (5 sequences), ST23 (4 sequences), ST24 (ST24a, 6 sequences; ST24b, 6 sequences; ST24c, 2 sequences), ST25 (9 sequences), ST26 (9 sequences), and ST30 (3 sequences) (Table 1). Additionally, all previously published full-length references of ST10a, ST10b, ST10c,

TABLE 2 *Blastocystis* full-length *SSU* rRNA gene nucleotide sequences used in the construction of phylogenetic trees including information on host, country of origin, and GenBank accession number associated with each sequence. Bold denotes sequences obtained in this study.

Subtype	Subgroup	Host	Country	GenBank accession no.
ST1		Human	USA	MT898451
ST2		NHP	Japan	AB070997
ST3		Human	N/A	AB091234
ST4		Human	Spain	MT898453
ST5		Pig	Japan	AB070998
ST6		Chicken	Brazil	MW538478
ST7		Chicken	Brazil	MW538476
ST8		NHP	Japan	AB107970
ST9		Human	Denmark	KC138681
ST10	a	Goat	Colombia	MZ265404
		Muskox	Greenland	OM883870
		Sheep	Greenland	OM883871
		Cattle	**USA**	**OR117623**
		Goat	**Portugal**	**OR117633**
		Goat	**Portugal**	**OR117634**
		Sheep	**Portugal**	**OR117635**
		Sheep	**Portugal**	**OR117636**
		Sheep	**Portugal**	**OR117637**
		Sheep	**Portugal**	**OR117638**
		Sheep	**Portugal**	**OR117639**
	b	Cattle	USA	MT898456
		Cattle	USA	MT898457
		Dromedary	Libya	KC148207
		Goat	**Portugal**	**OR117640**
		Goat	**Portugal**	**OR117641**
		Roe deer	**Portugal**	**OR117642**
		Sheep	**Portugal**	**OR117643**
		Sheep	**Portugal**	**OR117644**
		Sheep	**Portugal**	**OR117645**
		Sheep	**Portugal**	**OR117646**
		Sheep	**Portugal**	**OR117647**
ST11		Elephant	USA	MT898454
ST12		Marsupial	N/A	EU427515
ST13		Mouse deer	UK	KC148209
ST14		Cattle	USA	MT898458
ST15		NHP	UK	KC148211
ST16		Marsupial	N/A	EU427514
ST17		Gundi	Libya	KC148208
ST21		White-tailed deer	USA	MW887929
ST23		Cattle	USA	MW887931
		Goat	Colombia	MZ265406
		Sheep	**Portugal**	**OR117664**
		Sheep	**Portugal**	**OR117665**
		Goat	**Portugal**	**OR117666**
		Goat	**Portugal**	**OR117667**
ST24		White-tailed deer	USA	MW887928
ST25		Cattle	USA	MW887933

TABLE 2 (Continued)

Subtype	Subgroup	Host	Country	GenBank accession no.
ST26		Cattle	USA	MW887932
ST27		Indian peafowl	Brazil	MW887934
ST28		Indian peafowl	Brazil	MW887935
ST29		Chicken	Brazil	MW538473
ST30		White-tailed deer	USA	MZ267674
ST31		White-tailed deer	USA	MZ267676
ST32		Goat	Colombia	MZ265403
ST33		Horse	Colombia	ON932569
ST34		Horse	Colombia	ON932571
ST35		Human	Brazil	OP720870
ST36		Bat	Mexico	OP720871
ST37		Heteromyd	Mexico	OP720872
ST38		Water vole	UK	OP720869
ST39		NHP	China	OP563846
ST40		Muskox	Greenland	OM885776
ST41		Human	Colombia	OQ703608
ST42	a[a]	**Cattle**	**USA**	**OR117626**
		Cattle	**USA**	**OR117627**
		Cattle	**USA**	**OR117628**
		Cattle	**USA**	**OR117629**
		Cattle	**USA**	**OR117630**
		Red deer	**Spain**	**OR117653**
		Cattle	**Portugal**	**OR117654**
	b[b]	**Cattle**	**USA**	**OR117631**
		Cattle	**USA**	**OR117632**
		Cattle	**Portugal**	**OR117655**
		Cattle	**Portugal**	**OR117656**
		Cattle	**Portugal**	**OR117657**
ST43		**Goat**	**Portugal**	**OR117658**
		Goat	**Portugal**	**OR117659**
		Sheep	**Portugal**	**OR117660**
		Sheep	**Portugal**	**OR117661**
		Sheep	**Portugal**	**OR117662**
		Sheep	**Portugal**	**OR117663**
ST44[c]		White-tailed deer	USA	MZ267675
		White-tailed deer	USA	MZ267679
		Sheep	Greenland	OM883872
		Cattle	**USA**	**OR117624**
		Cattle	**USA**	**OR117625**
		Goat	**Portugal**	**OR117648**
		Goat	**Portugal**	**OR117649**
		Sheep	**Portugal**	**OR117650**
		Sheep	**Portugal**	**OR117651**
		Sheep	**Portugal**	**OR117652**

Note: N/A, not available; NHP, non-human primate.

[a] Previously reported as ST10d.

[b] Previously reported as ST10e.

[c] Previously reported as ST10c.

and ST23 currently available in GenBank were used in the analyses performed in the present study (Table 2).

To assess the relationship between sequences generated in this study and available references for all other accepted *Blastocystis* STs (ST1-ST17, ST21, and ST23-ST41), phylogenetic and pairwise distance analyses were performed for the full-length sequences of the *SSU* rRNA gene (Figures 1–3, and Figure S1). Sequences previously classified as ST10 formed a clade that includes ST23 (Figure 1 and Figure S1). Within the ST10/ST23 clade, several distinct clusters are formed, which included sequences previously classified as ST10 subgroups (Figure 2). ST10a and ST10b cluster together and share a branch with ST10f. ST10c and ST23 branch basely to the clade formed by ST10a, ST10b, and ST10f. ST10d and ST10e cluster together on a separate branch from any other ST10 subgroup or ST23 (Figure 2).

Pairwise distance analysis was performed to evaluate sequence similarity between sequences classified as ST10 subgroups and ST23 (Figure 3). Shared sequence identity within the ST10/ST23 clade ranged from 96% to 99%. Sequences previously classified as ST10d and ST10e shared 99% sequence identity with each other but differed from any other ST10 or ST23 sequence by 3%–4% across the full-length of the *SSU* rRNA gene (Figure 3). ST10a and ST10b shared 99% sequence identity with each other and 98% sequence identity with ST10c and ST10f. ST10a shared 97% sequence identity with ST23, while ST10b shared 98% sequence identity with ST23. ST23, ST10c, and ST10f also shared 98% sequence identity in the pairwise distance analysis of full-length reference sequences.

ST10a and ST10b share 99% sequence identity; thus, dividing these ST10 subgroups is not supported based on sequence level differences (Figure 3). Since the ST10a/ST10b clade contains the first full-length reference sequence to be classified as ST10, we suggest that this clade retain the subtype designation ST10. We also suggest the continued use of subgroup designations ST10a and ST10b to differentiate the sequence variants within this clade. Similarly, ST10d and ST10e share over 99% sequence identity across the full-length of the *SSU* rRNA gene; thus, dividing these ST10 subgroups is not supported based on sequence level differences. However, the ST10d and ST10e sequences vary from any other ST10 subgroup and ST23 by 3% or more, supporting the classification of these sequences as a new subtype. We propose assigning the subtype designation ST42 to the ST10d/ST10e clade (Figure 4). Additionally, we suggest the use of a subgrouping system within this new subtype. ST10d will be classified as ST42a and ST10e as ST42b (Figure 4). ST10c

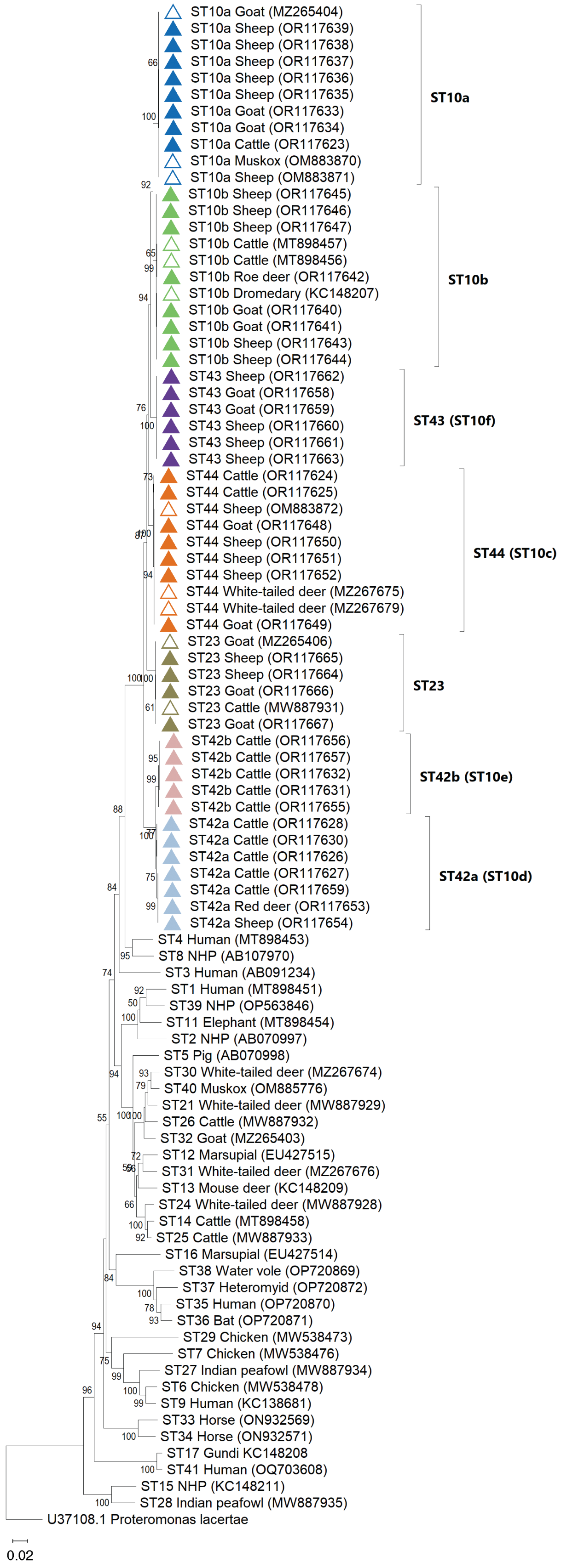


FIGURE 1 Phylogenetic relationships among *Blastocystis* full-length *SSU* rRNA gene nucleotide sequences generated in the present study (represented with a filled triangle), representative reference sequences of the accepted subtypes (Table 1), and *Proteromonas lacertae* used as outgroup taxon to artificially root the tree. Analysis was inferred using the neighbor-joining method with genetic distances calculated using the Kimura 2-parameter model. This analysis involved 92 nucleotide sequences, and there were a total of 1982 positions in the final dataset. Bootstrap values lower than 75% are not displayed.

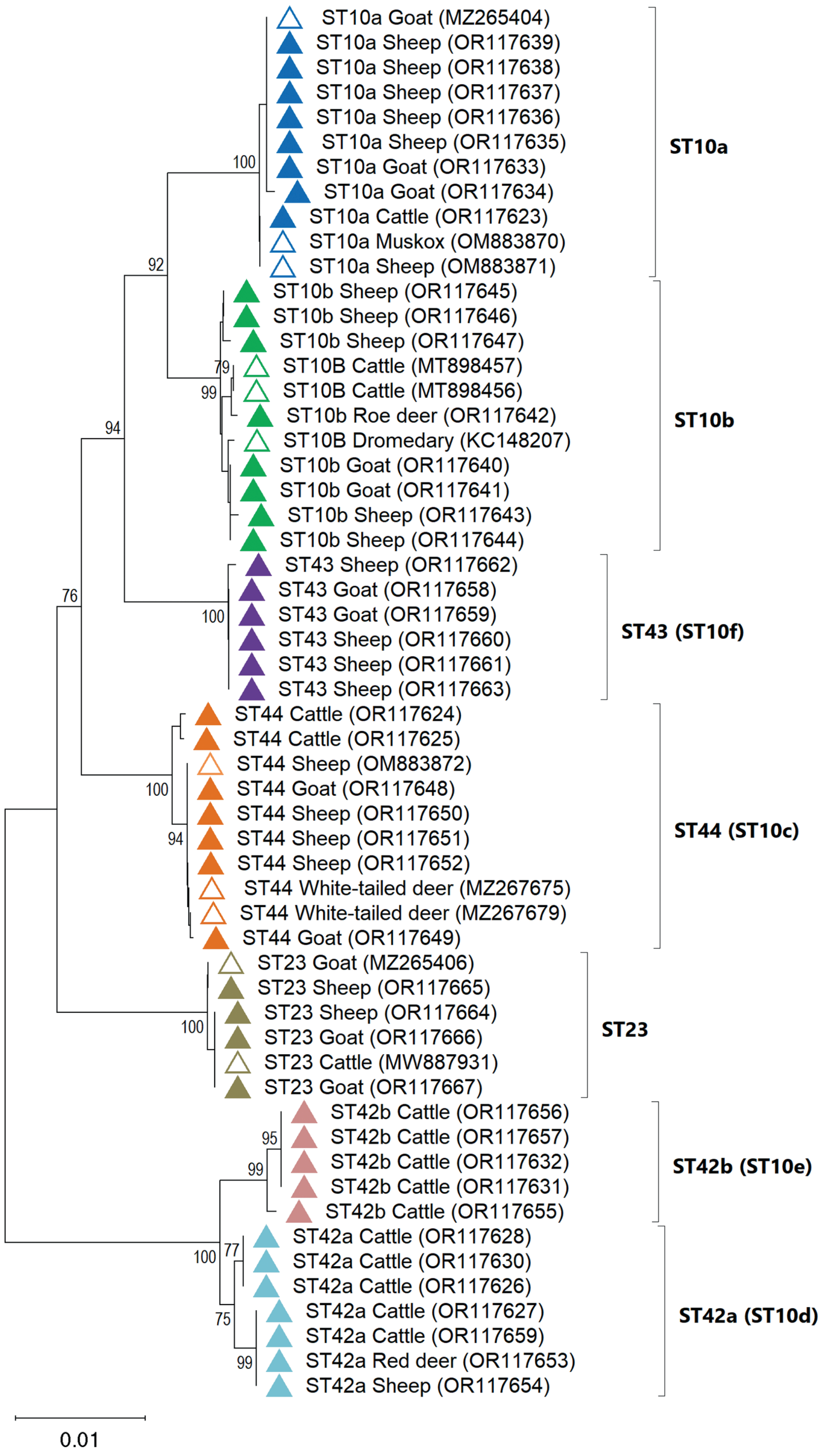
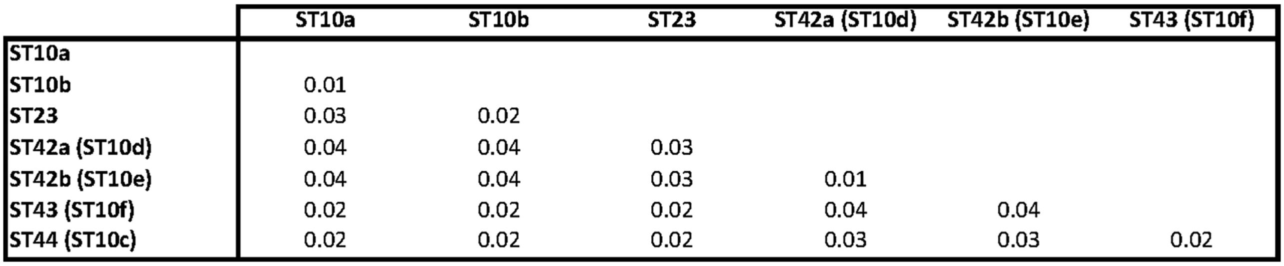


FIGURE 2 Subtree from Figure 1 showing phylogenetic relationships among *Blastocystis* full-length *SSU* rRNA gene for subtypes ST10a, ST10b, ST23, ST42a, ST42b, ST43, and ST44. Nucleotide sequences generated in the present study are represented with a filled dark blue, green, purple, orange, brown, pink, and light blue triangles for ST10a, ST10b, ST43, ST44, ST23, ST42a, and ST42b, respectively. All full-length reference sequences available in GenBank for subtypes ST10a (unfilled dark blue triangle), ST10b (unfilled green triangle), ST23 (unfilled brown triangle), and ST44 (unfilled orange) triangle are also represented. Bootstrap values lower than 75% are not displayed.



	ST10a	ST10b	ST23	ST42a (ST10d)	ST42b (ST10e)	ST43 (ST10f)
ST10a						
ST10b	0.01					
ST23	0.03	0.02				
ST42a (ST10d)	0.04	0.04	0.03			
ST42b (ST10e)	0.04	0.04	0.03	0.01		
ST43 (ST10f)	0.02	0.02	0.02	0.04	0.04	
ST44 (ST10c)	0.02	0.02	0.02	0.03	0.03	0.02

FIGURE 3 Pairwise distances between *Blastocystis* full-length *SSU* rRNA gene sequences for subtypes ST10a, ST10b, ST23, ST42a, ST42b, ST43, and ST44 showing the average number of base substitutions per site conducted using the Kimura 2-parameter model.

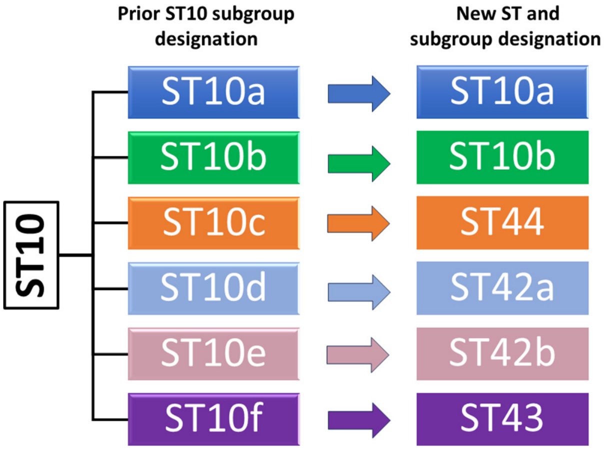


FIGURE 4 Depiction of the division of ST10 into ST10, ST42, ST43, and ST44 including previously used ST10 subgroup designations and current ST classifications.

and ST10f full-length sequences vary from any other ST by at least 2%, supporting the classification of two new subtypes. The subtype designation ST43 and ST44 will be assigned to ST10f and ST10c, respectively (Figure 4).

DISCUSSION

It has been noted that there is considerable diversity among the *SSU* rRNA sequences reported as *Blastocystis* ST10 (Higuera et al., 2021; Maloney et al., 2021; Maloney, Lombard, et al., 2019; Maloney, Molokin, & Santin, 2019; Santin et al., 2023; Stensvold et al., 2023). However, additional data were needed to validate these sequence variants as new subtypes. By applying a MinION-based sequencing strategy that produces near full-length sequences of the *SSU* rRNA gene in 21 samples selected from domestic and wild ruminants from different geographical origins, we have produced reference sequences of *Blastocystis* ST10 variants with the potential to be new subtypes. Furthermore, we have used this data to demonstrate that the ST10/ST23 clade contains five valid subtypes. These STs include the previously identified ST10 and ST23 and the newly named ST42, ST43, and ST44.

The division of the ST10/ST23 cluster, which includes three new STs (ST42-ST44), is an important step in clarifying the epidemiology of these STs both in humans and in other animals. For example, data showing distinct patterns of infection between ST10 sequence variants was recently reported in a longitudinal study of dairy calves (Santin et al., 2023). Additionally, recent publications have reported ST10 in samples from humans living in Thailand and Vietnam (Jinatham et al., 2021; McCain et al., 2023; Nguyen et al., 2023). When sequences from humans from Thailand classified as ST10 were compared to full-length references for ST10 and ST42-ST44, we observed that both ST10 and ST42 represented best matches with over 99% sequence similarity. Likewise, when sequences from humans from Vietnam that were classified as ST10 were compared to full-length references for ST10 and ST42-ST44, one sequence with 99.7% similarity with ST10b and two sequences with 100% similarity to ST44 were observed. Those findings indicate that ST10, ST42, and ST44 may be found in humans. These observations further demonstrate the need for full-length references of *Blastocystis* subtypes and point to the need to exercise caution in how subtype designations are assigned when using relatively small fragments of the *SSU* rRNA gene.

An important consideration when classifying sequences of ST10, ST23, and ST42-ST44 is the region of the *SSU* rRNA gene being used for classification. Sequence variation within the first 600 bp (barcoding region) of the *SSU* rRNA gene is low for subtypes within this clade. This situation is not unique to the ST10/ST23/ST42-ST44 clade and has been similarly noted as problematic for classifying sequences within the ST14/ST24/ST25 clade, the ST5/ST12 clade, and the ST17/ST41 clade (Hernández-Castro, Dashti, et al., 2023; Hernández-Castro, Maloney, et al., 2023; Maloney & Santin, 2021). Thus, if subtype designations are being made using primers such as the so-called barcoding primers (Scicluna et al., 2006) that cover the first 600 bp region of the *SSU* rRNA gene or the commonly used qPCR primers (Poirier et al., 2011) that cover a 300 bp region within the barcoding region, definitive ST assignments may not be possible. In such situations, it is suggested that using clade-level designations as described above would prevent the misclassification of sequences. Sequencing of the hypervariable regions found in the 500–900 bp region of the *SSU* rRNA gene, which is amplified using the Santin primers (Santín et al., 2011) allows for discrimination of these subtypes. However, caution should still be used when choosing a BLAST match to assist with subtype determination of a sequence. Using a BLAST criterion of shared percent identity of at least 99% to a full-length reference sequence of a valid ST to classify *Blastocystis* sequences would help to limit the misclassification of sequences and the propagation of misclassifications in the literature. When shorter sequences, such as those generated using qPCR primers, are used to classify STs, it may be necessary to consider matches of 100% to established references for accurate sequence classification.

Surveys of *Blastocystis* subtype diversity in new hosts and regions of the world continue to expand the host range of previously identified subtypes and identify new subtypes (Baek et al., 2022; Hernández-Castro, Dashti, et al., 2023; Hernández-Castro, Maloney, et al., 2023; Jinatham et al., 2021; Nguyen et al., 2023; Stensvold et al., 2023; Yu et al., 2023). As interest in *Blastocystis* grows, the ability to correctly identify and distinguish sequences is more important than ever. The validation and designation of new subtypes ST42-ST44 provide an important step in helping to understand the epidemiology and host specificity of *Blastocystis*. This study also provides baseline evidence suggesting that ST subgrouping based on near full-length nucleotide sequences of the *SSU* rRNA gene might be a useful tool to better understand not only the actual host range and specificity of a given genetic variant but also potential links with pathogenicity and virulence.

ACKNOWLEDGMENTS

This research was funded by USDA-ARS Project No. 8042-32000-112-00-D. A.M.F. was supported by a PhD grant from Fundação para a Ciência e Tecnologia (SFRH/BD/144582/2019), co-financed by the European Social Fund POPH-QREN program. D.G.-B. was the recipient of a 'Sara Borrell' postdoctoral fellowship (CD-19CIII/00011) funded by the Spanish Ministry of Science, Innovation and Universities. A.D. was the recipient of a PFIS contract (FI20CIII/00002) funded by the Spanish Ministry of Science and Innovation and Universities. This work was supported by the Centre for Environmental and Marine Studies (CESAM) through FCT/MCTES (UIDP/50017/2020+UIDB/50017/2020+LA/P/0094/2020), and national funds. This article is based on work from COST Action *Blastocystis* under One Health, CA21105, supported by COST (European Cooperation in Science and Technology).

ORCID

Monica Santin https://orcid.org/0000-0002-1386-6255
Ana Figueiredo https://orcid.org/0000-0002-2623-6340
Aleksey Molokin https://orcid.org/0000-0002-4571-7641
Nadja S George https://orcid.org/0000-0001-8681-3544
Pamela C. Köster https://orcid.org/0000-0001-5963-8824
Alejandro Dashti https://orcid.org/0000-0001-8707-5731
David González-Barrio https://orcid.org/0000-0001-5083-8854
David Carmena https://orcid.org/0000-0002-4015-8553
Jenny G. Maloney https://orcid.org/0000-0002-6405-883X

REFERENCES

Baek, S., Maloney, J.G., Molokin, A., George, N.S., Cortés Vecino, J.A. & Santin, M. (2022) Diversity of *Blastocystis* subtypes in horses in Colombia and identification of two new subtypes. *Microorganisms*, 10, 1693.

Bahrami, F., Babaei, E., Badirzadeh, A., Riabi, T.R. & Abdoli, A. (2020) *Blastocystis*, urticaria, and skin disorders: review of the current evidences. *European Journal of Clinical Microbiology and Infectious Diseases*, 39, 1027–1042.

Cifre, S., Gozalbo, M., Ortiz, V., Soriano, J.M., Merino, J.F. & Trelis, M. (2018) *Blastocystis* subtypes and their association with irritable bowel syndrome. *Medical Hypotheses*, 116, 4–9.

Fréalle, E., El Safadi, D., Cian, A., Aubry, E., Certad, G., Osman, M. et al. (2015) Acute *Blastocystis*-associated appendicular peritonitis in a child, Casablanca, Morocco. *Emerging Infectious Diseases*, 21, 91–94.

Hernández-Castro, C., Dashti, A., Vusirikala, A., Balasegaram, S., Köster, P.C., Bailo, B. et al. (2023) Prevalence and temporal dynamics of *Cryptosporidium* spp., *Giardia duodenalis*, and *Blastocystis* sp. among toddlers attending day-care centres in Spain. A prospective molecular-based longitudinal study. *European Journal of Pediatrics*, 182, 213–223.

Hernández-Castro, C., Maloney, J.G., Agudelo-López, S.P., Toro-Londoño, M.A., Botero-Garcés, J.H., Orozco, M.C. et al. (2023) Identification and validation of novel *Blastocystis* subtype ST41 in a Colombian patient undergoing colorectal cancer screening. *Journal of Eukaryotic Microbiology*, e12978. Available from: https://doi.org/10.1111/jeu.12978

Higuera, A., Herrera, G., Jimenez, P., García-Corredor, D., Pulido-Medellín, M., Bulla-Castañeda, D.M. et al. (2021) Identification of multiple *Blastocystis* subtypes in domestic animals from Colombia using amplicon-based next generation sequencing. *Frontiers in Veterinary Science*, 8, 932.

Hublin, J.S.Y., Maloney, J.G. & Santin, M. (2021) *Blastocystis* in domesticated and wild mammals and birds. *Research in Veterinary Science*, 135, 260–282.

Jinatham, V., Maxamhud, S., Popluechai, S., Tsaousis, A.D. & Gentekaki, E. (2021) *Blastocystis* one health approach in a rural community of northern Thailand: prevalence, subtypes and novel transmission routes. *Frontiers in Microbiology*, 12, 3800.

Maloney, J.G. & Santin, M. (2021) Mind the gap: new full-length sequences of *Blastocystis* subtypes generated via Oxford nanopore MinION sequencing allow for comparisons between full-length and partial sequences of the small subunit of the ribosomal RNA gene. *Microorganisms*, 9, 997.

Maloney, J.G., Lombard, J.E., Shivley, C.B., Urie, N.J. & Santin, M. (2019) Zoonotic and genetically diverse subtypes of *Blastocystis* in US pre-weaned dairy heifer calves. *Parasitology Research*, 118, 575–582.

Maloney, J.G., Molokin, A. & Santin, M. (2019) Next generation amplicon sequencing improves detection of *Blastocystis* mixed subtype infections. *Infection, Genetics and Evolution*, 73, 119–125.

Maloney, J.G., Molokin, A. & Santin, M. (2020) Use of Oxford nanopore MinION to generate full-length sequences of the *Blastocystis* small subunit (*SSU*) rRNA gene. *Parasites and Vectors*, 13, 595.

Maloney, J.G., Jang, Y., Molokin, A., George, N.S. & Santin, M. (2021) Wide genetic diversity of *Blastocystis* in white-tailed deer (*Odocoileus virginianus*) from Maryland, USA. *Microorganisms*, 9, 1343.

Maloney, J.G., Molokin, A., Seguí, R., Maravilla, P., Martínez-Hernández, F., Villalobos, G. et al. (2023) Identification and molecular characterization of four new *Blastocystis* subtypes designated ST35-ST38. *Microorganisms*, 11, 46.

McCain, A., Gruneck, L., Popluechai, S., Tsaousis, A.D. & Gentekaki, E. (2023) Circulation and colonisation of *Blastocystis* subtypes in schoolchildren of various ethnicities in rural northern Thailand. *Epidemiology and Infection*, 151, e77.

Nguyen, L.D.N., Gantois, N., Hoang, T.T., Do, B.T., Desramaut, J., Naguib, D. et al. (2023) First epidemiological survey on the prevalence and subtypes distribution of the enteric parasite *Blastocystis* sp. in Vietnam. *Microorganisms*, 11, 731.

Poirier, P., Wawrzyniak, I., Albert, A., El Alaoui, H., Delbac, F. & Livrelli, V. (2011) Development and evaluation of a real-time PCR assay for detection and quantification of *Blastocystis* parasites in human stool samples: prospective study of patients with hematological malignancies. *Journal of Clinical Microbiology*, 49, 975–983.

Santín, M., Gómez-Muñoz, M.T., Solano-Aguilar, G. & Fayer, R. (2011) Development of a new PCR protocol to detect and subtype *Blastocystis* spp. from humans and animals. *Parasitology Research*, 109, 205–212.

Santin, M., Molokin, A. & Maloney, J.G. (2023) A longitudinal study of *Blastocystis* in dairy calves from birth through 24 months demonstrates dynamic shifts in infection rates and subtype prevalence and diversity by age. *Parasites & Vectors*, 16, 177.

Scicluna, S.M., Tawari, B. & Clark, C.G. (2006) DNA barcoding of *Blastocystis*. *Protist*, 157, 77–85.

Stensvold, C.R. & Clark, C.G. (2020) Pre-empting Pandora's box: *Blastocystis* subtypes revisited. *Trends in Parasitology*, 36, 229–232.

Stensvold, C.R., Suresh, G.K., Tan, K.S.W., Thompson, R.C.A., Traub, R.J., Viscogliosi, E. et al. (2007) Terminology for *Blastocystis* subtypes – a consensus. *Trends in Parasitology*, 23, 93–96.

Stensvold, C.R., Berg, R.P.K.D., Maloney, J.G., Molokin, A. & Santin, M. (2023) Molecular characterization of *Blastocystis* and *Entamoeba* of muskoxen and sheep in Greenland. *International Journal for Parasitology*. Available from: https://doi.org/10.1016/j.ijpara.2023.05.005

Tan, K.S.W. (2008) New insights on classification, identification, and clinical relevance of *Blastocystis* spp. *Clinical Microbiology Reviews*, 21, 639–665.

Yu, M., Yao, Y., Xiao, H., Xie, M., Xiong, Y., Yang, S. et al. (2023) Extensive prevalence and significant genetic differentiation of *Blastocystis* in high- and low-altitude populations of wild rhesus macaques in China. *Parasites & Vectors*, 16, 107.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Santin, M., Figueiredo, A., Molokin, A., George, N.S., Köster, P.C., Dashti, A. et al. (2024) Division of *Blastocystis* ST10 into three new subtypes: ST42-ST44. *Journal of Eukaryotic Microbiology*, 71, e12998. Available from: https://doi.org/10.1111/jeu.12998