Heliyon 6 (2020) e05387

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Succinate dehydrogenase gene as a marker for studying *Blastocystis* genetic diversity

Adriana Higuera [a], Marina Muñoz [a], Myriam Consuelo López [b], Patricia Reyes [b], Plutarco Urbano [c], Oswaldo Villalobos [d], Juan David Ramírez [a,*]

[a] Grupo de Investigaciones Microbiológicas-UR (GIMUR), Departamento de Biología, Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia
[b] Departamento de Salud Pública, Universidad Nacional de Colombia, Bogotá, Colombia
[c] Grupo de Investigaciones Biológicas de la Orinoquia, Fundación Universitaria Internacional del Trópico Americano - Unitrópico, Yopal, Colombia
[d] Hospital Local Santa María de Mompox, Programas Especiales (Lepra y TB), Mompox, Bolívar, Colombia

ARTICLE INFO

Keywords:
Microbiology
Molecular biology
Zoology
Epidemiology
Evolutionary biology
Blastocystis
ssu rRNA
Succinate dehydrogenase subunit A
Genetic diversity
Genetic population structure

ABSTRACT

Blastocystis has been reported as the most common eukaryotic microorganism residing in the intestines of both humans and animals, with a prevalence of up to 100% in some populations. Since this is a cryptic species, sequence polymorphism are the single strategy to analyses its genetic diversity, being traditionally used the analysis of *ssu rRNA* gene sequence to determine alleles and subtypes (STs) for this species. This multicopy gene has shown high diversity among different STs, making necessary to explore other genes to assess intraspecific diversity. This study evaluated the use of a novel genetic marker, succinate dehydrogenase (*SDHA*), for the typing and evaluation of the genetic diversity and genetic population structure of *Blastocystis*. In total, 375 human fecal samples were collected and subjected to PCR, subtyped using the *ssu rRNA* marker, and then the *SDHA* gene was amplified via PCR for 117 samples. We found some incongruences between tree topologies for both molecular markers. However, the clustering by ST previously established for *Blastocystis* was congruent in the concatenated sequence. *SDHA* showed lower reticulation (The origination of a lineage through the partial merging of two ancestor lineages) signals and better intra ST clustering ability. Clusters with geographical associations were observed intra ST. The genetic diversity was lower in the marker evaluated compared to that of the *ssu rRNA* gene (nucleotide diversity = 0.03344 and 0.16986, respectively) and the sequences analyzed showed population expansion with genetic differentiation principally among STs. The *ssu rRNA* gene was useful to explore inter-specific diversity but together with the *SDHA* gene the resolution power to evaluate intra ST diversity was higher. These results showed the potential of the *SDHA* marker for studying the intra ST genetic diversity of *Blastocystis* related with geographical location and the inter ST diversity using the concatenated sequences.

1. Introduction

Blastocystis spp., are anaerobic intestinal protists belonging to the phylum Heterokontophyta [1] of the Stramenopila group, which includes heterotrophic and photosynthetic protozoa [2]. *Blastocystis* has a cosmopolitan distribution [3, 4] and is the most common eukaryotic protozoan in the human intestine [5], with prevalence's up to 100% in a population of children in Senegal [6]. However, the role of this parasite at the intestinal level is still a matter of contention as it is present in both asymptomatic [7] and symptomatic patients, in the latter associated with inflammatory bowel disease (IBD), irritable bowel syndrome (IBS) [8] and chronic or acute urticarial lesions [9]. In some studies, it has been suggested that the potential pathogenic of *Blastocystis* could be related with the subtypes or STs [10, 11], however, this association remains in debate. Also, the STs have been related to its geographical distribution [4, 11, 12] and a relative specificity to different hosts [13]. For all these, it is important to increase studies in molecular epidemiology in areas where the STs are circulating, to determine the impact of *Blastocystis* diversity and associate it with both biological and clinical factors. Initially, the genetic diversity of this microorganism had been demonstrated through a variety of different techniques [14, 15, 16]; however, it was not until the establishment of phylogeny with the use of the complete small subunit ribosomal RNA (*ssu rRNA*) sequence [17] and elongation factor 1 alpha, that the clusters corresponding to the STs were observed.

* Corresponding author.
E-mail address: juand.ramirez@urosario.edu.co (J.D. Ramírez).

https://doi.org/10.1016/j.heliyon.2020.e05387
Received 3 August 2020; Received in revised form 1 October 2020; Accepted 27 October 2020
2405-8440/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Subsequently, with the use of genomic sequencing, that diversity and the differences between some STs with respect to their size, guanine-cytosine content, number of introns, and gene content were elucidated [18]. Currently, there are sequenced genomes for ST1, ST4 and ST7 [18, 19, 20] and some drafts genomes for ST2, ST3, ST6, ST8 and ST9 available on GenBank, with sizes ranging between 12 -18 Mb.

Due to the existence of genomic sequencing and multiple improvements in the ease and access of new generation sequencing techniques today, the sequencing of just a few genes is now feasible. Therefore, DNA barcoding has been developed for *Blastocystis* and is frequently used by the scientific community to identify STs with the *ssu rRNA* gene [21]. With this unique marker, the genus had been classified into 17 STs [13, 22], but currently 28 STs have been reported [23]. The use of this marker to determine STs has revealed a high diversity among them [5, 23], so it has been proposed that each ST could correspond to a different species [24]. However, due to the increasing number of STs found, it would be important to add other markers with different resolution power to explore the diversity in other genetic targets and conjunction with *ssu rRNA*, solving the variation both inter and intraspecific, even more considering the heterogeneity of *ssu rRNA* which was reported in the ST7 strain B isolate [25], with 17 non-identical copies [20, 25]. A few additional markers have also been evaluated, such as the pyruvate ferredoxin oxidoreductase (*PFOR*) gene, which allowed finding three clades with different ST samples, a lower nucleotide diversity and haplotype polymorphism for clade III [26]. Other study used the internal transcribed spacer *ITS* which revealed novel variants intra ST1 and a high gene flow among different countries of Europe and the Americas [27]. However, many aspects of *Blastocystis* remain unknown, so it is necessary to develop additional markers that allow for typing of the genus and determining some aspects about their biology, evolutionary history, reproduction mechanisms, possible recombination, and population genetic events.

Recently, studies of the genome for *Blastocystis* demonstrated the existence of synteny between ST1, ST4, and ST7 [28,29], with the presence of genes corresponding to complex I and II of the electron transport chain and an absence of III and IV [30], where they also determined four subunits, encoded by nuclear DNA, of the mitochondrial respiratory chain complex II that could work via succinate dehydrogenase (*SDH*) or fumarate reductase [20]. These pathways reportedly reverse reactions in other protozoa, such as *Trypanosoma cruzi* [31]. In eukaryotes, such as yeasts, *SDH* is a key enzyme that catalyzes the passage of succinate to fumarate during the tricarboxylic acid (TCA) cycle [32], it basically has four subunits called A, B, C and D. Subunit A is strongly associated with the mitochondrial inner membrane and covalently bound to flavin adenine dinucleotide (FAD), B contains three Fe–S groups and C and D which are integrals membrane proteins [33]. Although *SDH* is a mitochondrial protein, it is encoded by a nuclear gene and part of its structure is conserved among eukaryotes. Because nuclear genes are less affected by deleterious mutations associated with asexual reproduction [34] and due to the importance of this enzyme, we wanted to establish if this gene could be informative and could be used as an additional marker in the subtyping and/or discrimination of *Blastocystis* subtypes creating a better understanding of the genetic diversity and population structure of this stramenopile. Therefore, the main objective of this study was to determine the usefulness of the succinate dehydrogenase gene of the subunit A (*SDHA*) as a possible additional marker in the subtyping, study of the genetic diversity and population genetic structure of *Blastocystis*.

2. Materials and methods

2.1. Ethics approval and consent to participate

This study was approved by the ethics committee of the National University of Colombia (002-012-15 February 12, 2015) and the ethics committee of the Universidad del Rosario (registered in Act No. 394 of the CEI-UR). The patients approved and signed the written informed consent. This project was conducted under the contract number RGE131 of access to genetic resources granted by the "Ministerio de Medio ambiente y Desarrollo sostenible" from Colombia.

2.2. Selection of new genetic markers

A search was made of the genes corresponding to the constitutive enzymes, primarily those involved with the glycolysis cycle and a few from the Krebs cycle, using the genome available for ST7 in the *Blastocystis* Genome Browser (http://www.genoscope.cns.fr/). Subsequently, the larger sequences were downloaded into a FASTA format. The genes included in the search were *Triosephosphate isomerase (TPI), biphosphate aldolase (BPA), glucokinase (GK), Glucose 6 phosphate isomerase (GPI), Hexokinase (HK), Phosphofructokinase (PFK), Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Phosphoglycerate kinase (PGK), Phosphoglycerate mutase (PGM), Enolase (enol-1), Pyruvate kinase (PK)* and the subunit A *Succinate dehydrogenase (SDHA*). Initially, these sequences obtained from the ST7 genes were used as input in the Primer-BLAST tool. Once the pairs of primers were obtained, the following parameters were considered: the foll absence of dimer formation, absence of fork formation, melting temperature (Tm) in the range of 55 °C to 65 °C, percentage of 40–60% of GC, size of each primer from 18 to 30 bases in length, size of the amplified region from 300 bp to 700 bp, and specificity [35]. Subsequently, each gene was aligned with the primers using MUSCLE [36] implemented in MEGA 7.0 [37] and the sequence obtained after aligning these was verified through BLASTn, which ensured that a specific hit with 100% identity for *Blastocystis* was obtained.

Then, thanks to the collaboration of Doctors Lee O'Brien Andersen and Christen Rune Stensvold from the Department of Microbiology and Infection Control, Statens Serum Institut, Bruce Curtis and Andrew Rogers from the Center for Comparative Genomics & Evolutionary Bioinformatics, Biochemistry and Molecular Biology, Dalhousie University and Ivan Wawrzyniak from Université Clermont Auvergne, 3iHP, CNRS, Laboratoire Microorganismes: Génome et Environnement, we obtained the raw reads of the genomes available in Genbank for ST1-ST4, ST6, ST8 and ST9. We used Short Read Sequencing Typing 2 (SRST2) tool [38] to extract the genes of interest in ST1-ST4, ST6, ST8 and ST9. This tool maps the reads (in fastq format) of each genome on a database of reference alleles in fasta format (the 12 selected genes), in such a way that it allows identifying the presence of the locus and the allele that best matches said locus between all allelic sequences used as reference. In our case, the database used as a reference consisted of the sequences of each of the genes obtained from ST7. Consensus sequences were obtained for the *SDHA* gene of all STs, except for ST8, 1 consensus sequence for the *BPA* gene of ST2, 1 for the *PFK* gene of ST3 and 3 consensus sequences for the *GAPDH* gene of ST2, ST3 and ST6. The program was run using the default depth and coverage parameters.

The consensus sequences obtained in fasta format for the STs evaluated for the *SDHA* gene were aligned using MAFFT [39] together with the primers initially designed on the sequence of the *SDHA* gene of ST7, in order to verify the position of the primers and their alignment with the target sequences.

Additionally, the Ortholog Groups of Protein Sequences (OrthoMCL DB) database available at https://orthomcl.org/ was used to explore the copy number of each of the genes, where a single copy was found for the *SDHA*, *GPI*, *PFK* and *PGM* genes. In addition, for the particular case of the *SDHA* gene, we searched for this sequence in the genome of each of the STs available in Genbank, using BLAST Genomes. For the cases in which more than one sequence was obtained, an alignment was performed between these using MUSCLE [36] implemented in MEGA 7.0 [37], to identify how different or similar they were. Furthermore, previously designed primers were also aligned to verify location on these found sequences. All pairs of primers designed were evaluated on control DNA. The amplification conditions for each pair of primers were standardized using the *Blastocystis* xenic culture DNA ST3.

2.3. PCR of control DNA with newly designed primers

Once the primers designed for each marker were obtained, we wanted to evaluate them experimentally, for this a PCR with control DNA from the *Blastocystis* xenic culture ST3 was performed to verify the size of the amplified products. The reactions were performed in a final volume of 12.5 μL, 2 μL of DNA, 6.25 μl of Go Taq Master Mix Green (Promega) (cat. No. M7122) at a final concentration of 1X and at a final concentration for each primer of 1 μM. The thermal cycling parameters were as follows 95 °C for 5 min, 35 cycles of 95 °C for 1 min, 59 °C for 1 min, 72 °C for 1 min, and 72 °C for 10 min. Each amplification reaction was observed in 2% agarose gels and stained with SYBR Safe, Thermo Fisher Scientific (cat No. S33102) verifying the presence of a single band and the expected size.

Subsequently, 4 of the 12 markers evaluated were amplified with the control DNA, creating unique bands with an expected size, including *TPI* (550 bp), *PFK* (437 bp), *PK* (479 bp), and *SDHA* (514 bp). The sequences of the primers that worked for each locus, with the control DNA, were the following: *TPI* (Fw 5′ GCGTTCACAGAACCTCCGTA 3; Rv 5′ CCTCCAACCTGAACAGCGAT3′), *PFK* (Fw 5′ TACCACTTCGTGC GCTTGAT3; 5′ACGCAGGACACGATGAACTT3′), *PK* (Fw 5′ CGTCAGATCACCGTCGGAAA3; Rv 5′ ACCAAGATCGTATGCACGCT3′) and *SDHA* (Fw5′ GTCGATCCATCGCTTCCACT3; Rv 5′ CAGTCC GCCCATGTTGTAGT3′).

These four genes were chosen to evaluate them in the *Blastocystis* positive DNA fecal samples.

2.4. Study population

A convenience sampling was performed in the departments of Amazonas, Bolívar, Casanare, and Córdoba in Colombia (Figure 1). In total, 375 samples of human feces were collected. From the municipality of Puerto Nariño in the department of Amazonas, samples (75.7%, n = 284) were collected from 3 indigenous rural settlements, as described by Sánchez and collaborators [40]. Samples were also collected from the urban area of Mompós in the department of Bolívar (8%, n = 30), the municipalities of Poré, Tamara and Yopal in the department of Casanare (14.1%, n = 53), and samples from Montería in the department of Córdoba (2.1%, n = 8) as depicted in Figure 1. The samples were collected in plastic containers with 70% ethanol (ratio 1: 4 feces: alcohol) and kept refrigerated for subsequent DNA extraction.

2.5. DNA extraction, Blastocystis DNA detection, and subtyping

Prior to DNA extraction, each sample was washed with 1x sterile PBS. Genomic DNA from the stool samples was obtained using the commercial kit, Stool DNA Isolation kit from Norgen (cat. No 27600), following the manufacturer's instructions. DNA samples were stored at -30 °C until use. Initially, the samples were subjected to PCR to detect *Blastocystis*. Samples from the department of Amazonas were processed for the detection of *Blastocystis* according to the protocol reported by Sanchez et al. [40]. Conventional PCR was performed for the detection of *Blastocystis* of the samples obtained from the departments of Bolívar, Casanare, and Córdoba. PCR for detecting *Blastocystis*, was performed in a final volume of 9 μL containing 3.5 μL of GoTaq Green Master Mix (Promega) (cat. No M7122), 2 μL of the template DNA, and the primers. Species-specific primers were used at a final concentration of 1 μM. The sequences of the primers used for *Blastocystis* were FWD F5 (5′-GGTCC GGTGAACACTTTGGATTT-3′) and R F2 (5′-CCTACGGAAA CCTTGTTACGACTTCA-3′) [41]. The thermal cycling parameters were as follows 95 °C for 5 min; 35 cycles of 95 °C for 15 s, 58 °C for 1 min, 72 °C for 30 s and then 72 °C for 10 min. The expected size of the amplified fragments was 119 bp [5].

Conventional PCR was performed in order to determine the subtypes and alleles of the samples that were positive for *Blastocystis*. Thus, through the amplification of a 600pb region of the 5′end of the *ssu rRNA*

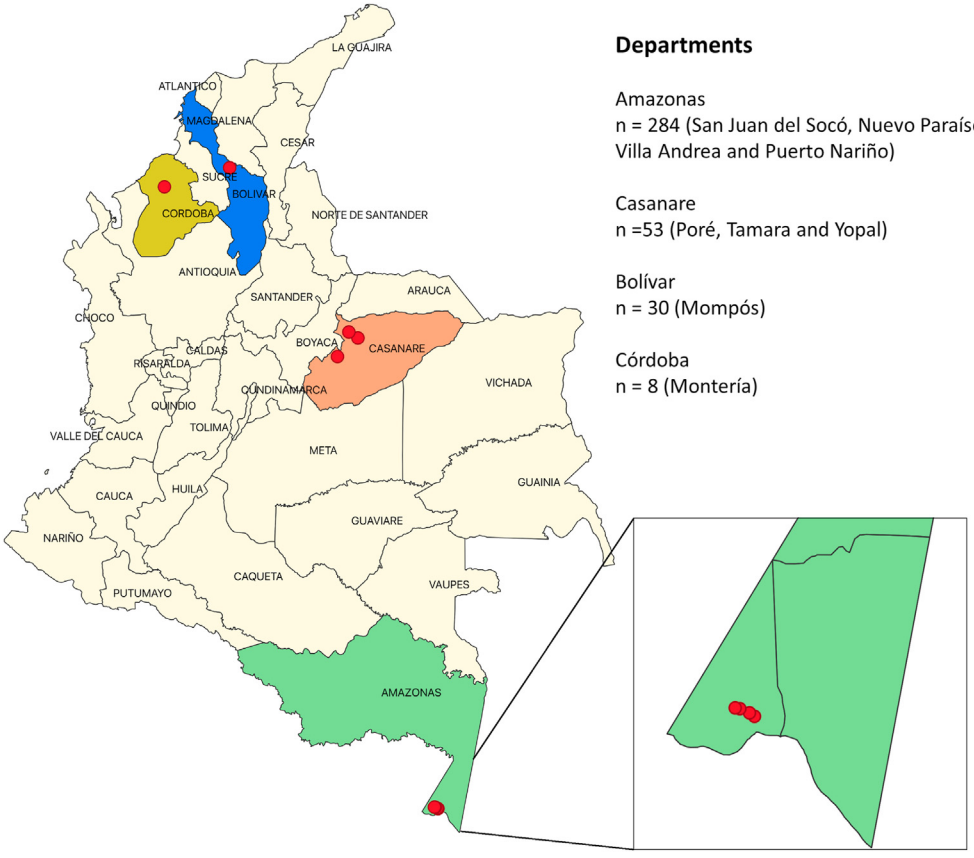

Figure 1. Geographic locations of departments in which samples were collected. The departments where the sample collection was conducted are highlighted in color. The red dots indicate the exact location of the sampled municipalities. In the table a zoom of the department of the Amazon is observed, for a better observation of the sampled sites. In the legend on the right, the municipalities of each department and the number of samples taken are indicated. The map was created using ArcGIS version 10.7.

using primers BhRDr (5′-GAGCTTTTTAACTGCAACAACG-3′) and RD5 (5′-ATCTGGTTGATCCTGCCAGT-3′) as reported previously [21]. Once all of the PCR subtyping were performed, the size of each amplicon was assessed using 2% agarose gel electrophoresis followed by staining with SYBR Safe, Thermo Fisher Scientific (cat No. S33102). Subsequently, each product was purified with ExoSAP-IT®, Affymetrix™ (cat. No 15513687) following the manufacturer's recommendations. Both strands of each product amplified were sequenced using the Sanger method. Sequences were edited in MEGA 7.0 [37] and submitted to a database to identify the subtypes and alleles (https://pubmlst.org/blastocystis/).

2.6. PCR testing of fecal samples with designed primers

Here, we wanted to experimentally evaluate the specificity of the primers on the evaluated markers, using them in different DNA samples obtained directly from human feces. For this, a PCR was performed on DNA from positive and identified human stool samples of ST1, ST2 and ST3.

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the National University of Colombia (002-012-15 February 12, 2015) and the ethics committee of the Universidad del Rosario (registered in Act No. 394 of the CEI-UR). This project was conducted under the contract number RGE131 of access to genetic resources granted by the "Ministerio de Medio ambiente y Desarrollo sostenible".

Only the primers that produced the best PCR results with DNA control (*TPI, PK, PFK* and *SDHA*), were chosen to be evaluated in the DNA testing of the human fecal samples. The conditions for these PCR were the same as those used with the DNA control. Then, each of the products of each marker was purified with ExoSAP-IT®, Affymetrix™ (cat. No 15513687) following the manufacturer's recommendations. Each product amplified were sequenced using the Sanger method. Sequences were edited in the MEGA 7.0 [37]. However, after PCR, sequencing and editing processes of sequences, suitable results were only obtained for the *SDHA* marker. With the other markers, the intensity of the band was very weak, and in some cases, there was no amplification, or the sequencing results were not successful.

2.7. Phylogenetic reconstructions and determination of the haplotype networks

The edited sequences for *ssu rRNA* and the *SDHA* markers in the fasta format were aligned using the multiple sequence alignment program MAFFT v7 [39]. Subsequently, different phylogenetic trees using each marker and the concatenated sequences of both markers were constructed. Also, phylogenetic trees by ST with *SDHA* and *ssu rRNA* were constructed. The trees were run with the maximum likelihood method. The type of analysis carried out included the determination of the evolution model for each data set of each gene using ModelFinder [42] and the tree reconstruction with 1000 bootstrap replicates using IQtree [43]. The visualization and edition of the phylogenetic trees was carried out with the online tool Interactive Tree Of Life V32 [44]. The consensus sequences obtained for the *in silico* analysis of each ST were included in the tree. The ST7 was used as the outgroup. In addition, phylogenetic networks were constructed to detect reticulation signals between samples evaluated by gene and their concatenation. The analysis used the SplitsTree5 program [45] with the Neighbornet algorithm and 1,000 iterations.

Also, we used https://microreact.org/showcase to visualize the relation between geographical regions with the phylogenetic trees by ST constructed previously with IQtree software.

Furthermore, to determine the number of haplotypes in the population for the *SDHA* gene, a fasta sequence matrix was constructed for the haplotype network analysis using prior alignment with MAFFT v7 [39]. This alignment of the fasta format was imported into the Geneious Prime software (available at https://www.geneious.com/) to export a file with.phy extension and subsequently submitted to the Network 5.0 program [46] (available on http://www.fluxus-engineering.com/sharenet.htm) to build a haplotype network by geographical region based on the median-joining model with 1000 iterations. Analysis of the haplotype networks allowed for the determination of the intraspecific relationships between the different haplotypes and the mutational positions generated between them.

2.8. Evaluation of discrimination power and typing efficiency

The sequences of each gene and their concatenation were aligned using the multiple sequence alignment programme MAFFT v7 software [39]. Then, these alignments were included as input data to MLSTest software, to evaluate the discriminatory power (DP) with 95% CI and typing efficiency (TE) of *SDHA* and *ssu rRNA* genes. TE is an indicator of grouping of members with common characteristics and DP allows differentiation of individuals belonging to different groups [47].

2.9. Calculation of diversity and the genetic differentiation indices

To determine the number of polymorphisms, present in the sequences of the *ssu rRNA* and the *SDHA* genes, the previous alignment of each locus was used with the MAFFT v7 software [39]. The sequences used in this study were grouped into three populations, determined by the departments of origin of the samples (Amazon, Bolívar, and Casanare). We analyzed 92 sequences in total for the determination of both loci, considering populations by departments and STs. A total number of sites (including gaps) at 376 for the *SDHA* locus and 1314 for the *ssu rRNA*. All sequences, except 1 sequence from Córdoba, were used to calculate the diversity indices such as nucleotide diversity (Pi) and Theta (per site) from the total number of mutations (Eta), number of polymorphic (segregating) sites (S), number of haplotypes (h) and haplotypic diversity (Hd).

In addition, in order to determine if the sequences evaluated presented a neutral evolution or were involved in a selection process, the Tajima D was calculated, which indicate a balancing selection with a positive value and purifying selection if the value is negative [48]. Finally, statistics of the genetic differentiation between populations (departments and STs) for the *SDHA* and *ssu rRNA* loci were calculated. The department of Córdoba was excluded since only one sequence had been obtained for this group. Peer genetic differences were estimated for the populations, calculating the Wright's statistic F (Fst). Then, the average number of nucleotide differences in pairs (Kxy), nucleotide substitutions per site (Dxy), net nucleotide substitutions per site (Da), and gene flow from the haplotypes (Gst). DnaSP v.5 software was used for the analysis (available at http://www.ub.edu/dnasp).

3. Results

3.1. In silico analysis of designed primers

Initially, the 12 gene sequences obtained from the available annotation for ST7 were used for the design of primers. Then, thanks to the collaboration of different researchers who shared with us the sequencing reads (see materials and methods) used for the assembly of the genomes of different STs of *Blastocystis* publicly available in Genbank, we mapped the reads on the sequences of interest using Short Read Sequencing Typing 2 (SRST2) tool [38] and obtained consensus sequences for the *SDHA* gene of all STs, except for ST8, 1 consensus sequence for the *BPA* gene of ST2, 1 for the *PFK* gene of ST3 and 3 consensus sequences for the gene *GAPDH* of ST2, ST3 and ST6. These initial results showed the *SDHA* gene as a potential candidate to evaluate diversity across different STs.

To verify the initially designed primers, the consensus sequences obtained for the *SDHA* gene were aligned using MAFFT [39] together with the primers designed on the sequence of the ST7 *SDHA* gene. Only

two SNPs were found in the position of the first reverse on the ST4 sequence. The first forward was located at the beginning of the alignment in a conserved region for all the sequences evaluated. These first results show these primers as a starting point for the evaluation of the *SDHA* gene in the STs evaluated.

On the other hand, for the particular case of the *SDHA* gene, we wanted to verify the number of copies of this for each STs evaluated and, if there were more than one, determine how similar they were. For this, we did a search for this sequence in the genome of each of the STs available in Genbank, using BLAST Genome. For ST3, ST4 and ST9 we obtained a unique sequence for each one, with minimum coverage and identity percentages of 99 and 80%, respectively. In the case of ST1, two sequences were found, which presented a difference of two SNPs in the region flanked by the designed primers, but which did not affect their alignment, and, in the case of ST7, we found 4 sequences: two were identical to each other with a conserved location that allowed alignment of the primers. The other two sequences had a difference of 1 SNP from each other and were very different from the other two previously detected copies. Furthermore, it should be noted that these last two sequences did not align with the designed primers, indicating that these are specific for the first two identical copies found in ST1. In the case of ST2 and ST6, we were unable to retrieve any sequence after the search performed.

3.2. Newly designed primers with control DNA

Four of twelve new loci evaluated for *Blastocystis* were amplified with a single band of the expected size when we tested them on the control DNA of *Blastocystis* ST3 from a xenic culture. The other 8 loci were discarded because of different reasons, such as, they did not amplify, the size band was incorrect, some faint bands or multiple bands in some cases were showed, preventing their use for the purpose of this study. Due to this, only four loci (*TPI, PFK, PK* and *SDHA*) were evaluated in the DNA of faecal samples.

3.3. Sample description and detection of Blastocystis DNA

The average age of the human population was 8.4 years (SD: 7.5 years, range: 1-70). The largest number of samples collected was in the department of Amazonas (75.7%; n = 261), then 20.3% (n = 76) for San Juan del Socó, 8.3% (n = 31) Nuevo Paraíso, 11.2% (n = 42) Villa Andrea, and 29.9% (n = 112) Puerto Nariño. From Casanare, 14.1% (n = 53) of the total samples were collected, which corresponded to the following municipalities Poré (9.6%; n = 36), Tamara (1.6%; n = 6), and Yopal (2.9%; n = 11). Moreover, for Bolívar (Mompós) and also Córdoba (Montería), 8.0% (n = 30) and 2.1% (n = 8) were collected, respectively (Figure 1).

Regarding the detection of *Blastocystis* DNA by PCR with the *ssu rRNA* marker [5], a total of 86.6% [95% CI: 88.1–94.5, n = 305] of the samples were positive. From the total samples, the percentages of positive samples for each department were Amazonas (66.5%, n = 234), Casanare (9.9%, n = 35), Bolívar, Mompós (8.2%, n = 29) and Córdoba, Montería (1.9%, n = 7). The percentages of positive samples by municipality in the department of Amazonas were San Juan del Socó (18.1%, n = 68), Nuevo Paraíso (7.7%, n = 29), Villa Andrea (10.1%, n = 38), and Puerto Nariño (26.4%, n = 99) and, the percentages of positive samples by municipality in the department of Casanare were Poré (6.9%, n = 26), Yopal (1.6%, n = 6), and Tamara (0.8%, n = 3).

3.4. Blastocystis subtyping (ssu rRNA) and SDHA amplification

A total of 77.4% (n = 236) of the positive samples for *Blastocystis* DNA were subtyped and the following STs identified ST3 (27.5%, n = 84), ST1 (26.2%, n = 80) and ST2 (23.6%, n = 72). The 69 samples that were positive for *Blastocystis* could not be subtyped because the quality of the electropherogram obtained after sequencing was not optimal and the sequence to establish the ST could not be obtained. On the other hand, when we analyzed the electropherograms, we observed mixed nucleotides for the same position in some (17) of the sequences analyzed, indicating possible mixed infections [49].

Initially, we tested DNA primers for four loci in all the positive samples (n = 305). A total of 8.5% (n = 26), 19.01% (n = 58), 12.8% (n = 39) and 41.3% (n = 126) samples were amplified with primers for *TPI, PFK, PK* and *SDHA,* respectively. In the case of *TPI, PFK* and *PK* the amount of sequences obtained was too low and the quality of their electropherograms were not appropriated to get a consensus sequence, so we could not analyze them. By contrast, in the case of *SDHA,* we obtained clean sequences for each sample. The sequences were deposited under the accession numbers MT072325 to MT072444. The quantity of samples amplified for this locus, by department, were Bolívar (15%, n = 18), Casanare (19.6%, n = 23), Córdoba (0.8%, n = 1), and Amazonas (64.1%, n = 75). In order to compare *ssu rRNA* and *SDHA* genes, only those samples where the amplification of both genes could be obtained were used, so 9 samples were discarded.

3.5. Genetic diversity and differentiation

The genetic diversity indexes were calculated using all the sequences obtained for each locus by department and by ST, with the exception of Córdoba (only 1 sample). Comparing the results of haplotype diversity (Hd) and nucleotide diversity (Pi) by department (Figure 2a, b) and by ST (Figure 2c, d) with each loci, greater Pi values were found in the *ssu rRNA* locus, principally in the Bolívar department (Figure 2b), in contrast to *SDHA* locus which showed similar diversity indexes in all departments (Figure 2a). Similar results were obtained by ST, where the highest nucleotide diversity was shown in ST1 of *ssu rRNA* gene and the lower haplotype diversity value showed was in ST1 of *SDHA* gene (Figure 2c, d). In general, *ssu rRNA* gene showed higher diversity in comparison with *SDHA* (Supplementary material Table S1). The number of polymorphic sites obtained for the *ssu rRNA* was 193 segregating sites and the number of haplotypes 56, unlike the *SDHA* with 126 sites and 36 haplotypes. It should also be noted that, with respect to the sampled departments, Casanare and Bolívar showed a larger Theta per site in the *ssu rRNA* locus. Also, a greater number of haplotypes in Amazonas was observed with both genes, being higher with *SDHA* locus. Conversely, in the *SDHA* locus, nucleotide diversity indices were very similar between the three Colombian departments and the ST3 showed higher nucleotide and haplotype diversity for this locus (Figure. 2a, c; Supplementary material Table S1). Minimum number of recombination event was calculated for both genes by geographical region and ST. In all cases *ssu rRNA* showed to be more diverse (Supplementary material Table S1).

Then, we applied the D Tajima test [50] to determine if there were any selective processes occurring in the sequences. We observed that a statistically significant negative value when evaluating the sequences of the three departments and ST for each of the loci. This indicated a process of population expansion with a high frequency circulating among the rare alleles (Supplementary material Table S1).

3.6. Phylogenetic reconstructions

Figure 3a-c depicts the phylogenetic trees constructed for the concatenated (*ssu rRNA* and *SDHA* sequences), *ssu rRNA* and *SDHA* genes, respectively, where the colors indicate the ST1, ST2 and ST3 that were determined for the samples, together with reference sequences, obtained by the SRST2 tool (see materials and methods), for the ST1 - ST4, ST6, ST9 and the ST7 used as outgroup. In comparison, the concatenated and *ssu rRNA* trees showed similar topology between them with little changes and incongruences among some clusters, for instance, inside of the ST1 cluster we could identify changes in the clustering of samples as Bol_26, Bol_23 and Cor_01 from ST2 and Cas_03 from ST3. Besides, we observed some subgroups clearly established inside of each ST. The *SDHA* tree, exhibited 4 clusters with bootstrap supported greater than 80%, with a

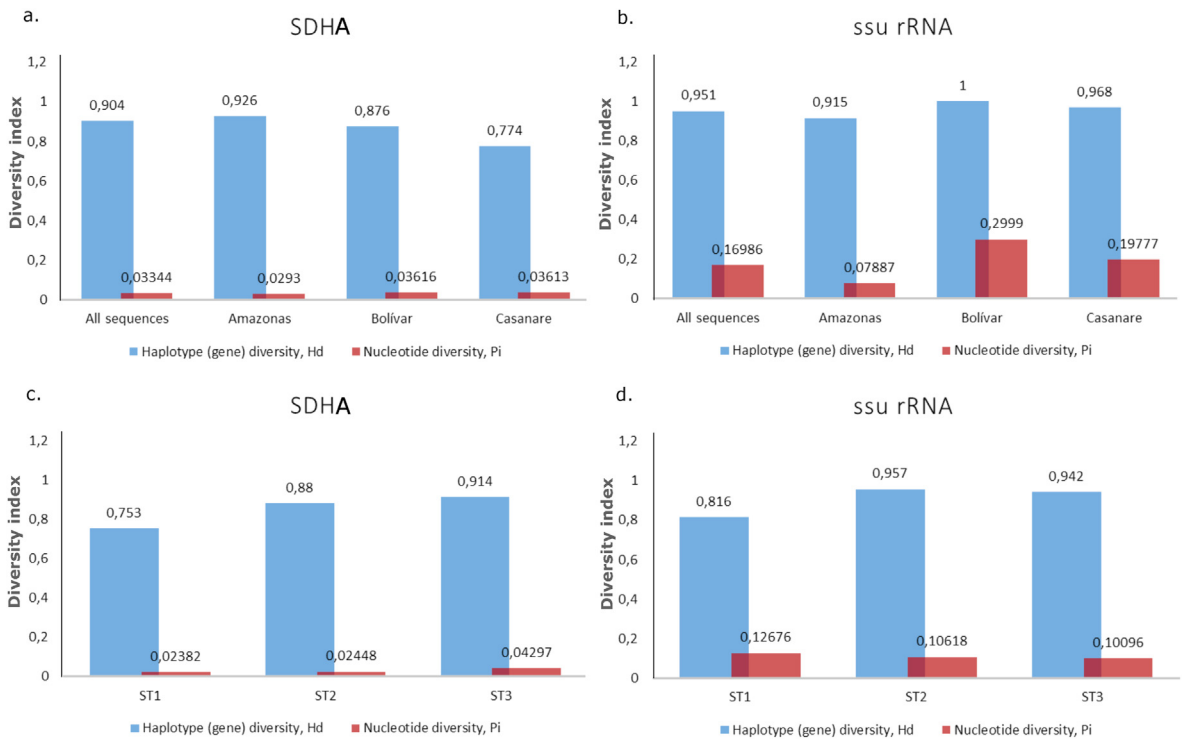

Figure 2. Nucleotide and Haplotype diversity in *ssu rRNA* and *SDHA* markers. Nucleotide and haplotypic diversity values for both loci are shown by geographic region and by ST. a.) *SDHA* by geographical region, b.) *ssu rRNA* by geographical region, c.) *SDHA* by ST, b.) *ssu rRNA* by ST.

relative evidence of clustering associated with the STs where it is possible to detect the clusters of ST1, ST2 and ST3. The concatenated topology obtained for *ssu rRNA* and *SDHA* genes are similar to those reported in *Cryptococcus* [51]. Concatenated and *ssu rRNA* presents signatures of evolution by gradualism where big changes result from many cumulative small changes [52]. Figures 3a and b show this profile. In the case of *SDHA* gene (Figure 3c), phylogenetic reconstruction coincides with the principle of punctuated equilibrium (abrupt and rapid changes that give rise to well-differentiated clusters) [52].

SDHA tree had some incongruences of clustering in comparison with the other trees (Figure 3). For these inconsistences, we wanted to determine some signals of reticulation for the concatenated and each individual gene (Figures 3d – f). Higher signals of reticulation were evidenced in the case of the concatenated and the *ssu rRNA* gene and lower in the *SDHA* gene. On the other hand, we wanted to verify if the *SDHA* marker could resolve differentiation intra ST. For this purpose, phylogenetic trees for each ST with the sequences of *SDHA* were constructed. In comparison with *ssu rRNA* trees by ST, at least three subgroups were detected inside each *SDHA* phylogenetic tree with enough support, indicating diversity intra ST with this genetic target and highlighting the resolution power of this marker (Figure 4). Then, using Micro React tool (https://microreact.org/showcase) some relation between each marker and geographical location was checked (Figures 5a-c). In the ST1, it was clear the cluster of Amazonas department and another cluster conformed by Bolívar and Casanare sequences. In the case of ST3 the sequences from Amazonas could be splitted in two clusters and an additional cluster was conformed for Bolívar and Casanare. In the case of ST2, a cluster is observed for Amazonas sequences and another mixed.

In addition, we use the MLSTest to evaluate the discrimination power (DP) and typing efficiency (TE) [47] of *SDHA* and *ssu rRNA* genes. TE is an indicator of grouping of members with common characteristics. In our case, the value of TE for *SDHA* was higher (TE = 0.257) than *ssu rRNA* (TE = 0.052) and concatenated (TE = 0.029), showing that *SDHA* gene has better grouping capacity which was observed intra STs. DP, which allows differentiation of individuals belonging to different groups showed values between 0.904 – 1. So, both markers can discriminate individuals from different STs with a confidence interval range 0.865–1.

3.7. Haplotypes network and genetic differentiation

Through the sequences obtained for the *SDHA* gene, haplotype networks were constructed under the medium-joining model [46], in order to understand the genetic variability of the circulating strains of *Blastocystis* in the areas of the sampled departments. A total of 36 haplotypes were found, of which a group of them were observed with a small number of mutations among them, which were all from the department of Amazonas. In contrast, it was possible to observe some samples with a very high number of mutations compared to other haplotypes. For example, those samples from Casanare, Bolívar, and Amazonas, that are seen further away (Figure 5d), according to the great diversity present in the genus *Blastocystis*. In addition, some shared haplotypes were observed between the collection regions (Figure 5d). The distribution of haplotypes using the ST of each of the samples was observed. In this case, it was possible to observe that the samples that corresponded to ST1, presented a greater number of haplotypes, which had diverged by a varied number of mutational changes, despite coming from different geographical regions. With respect to the samples classified as ST2 and ST3, no groups of clearly associated haplotypes were observed (Supplementary material Figure S1).

Finally, the genetic differentiation was determined considering each department and STs like different populations. Fst and Gst statistics, gave values between -0.00818 to 0.10163 and -0.00501 to 0.03104, respectively, showing no genetic differentiation among Bolívar and Casanare but in case of Amazonas we found moderated differentiation between Amazonas and Casanare and Amazonas and Bolívar with *SDHA* marker [53]. In case of *ssu rRNA* and *SDHA* by ST, the genetic differentiation was evident between the populations evaluated (Table 1). The number of average nucleotide differences between populations (Kxy) was much higher for the *ssu rRNA* marker, the higher value being 49.72857 and the number of average nucleotide substitutions between

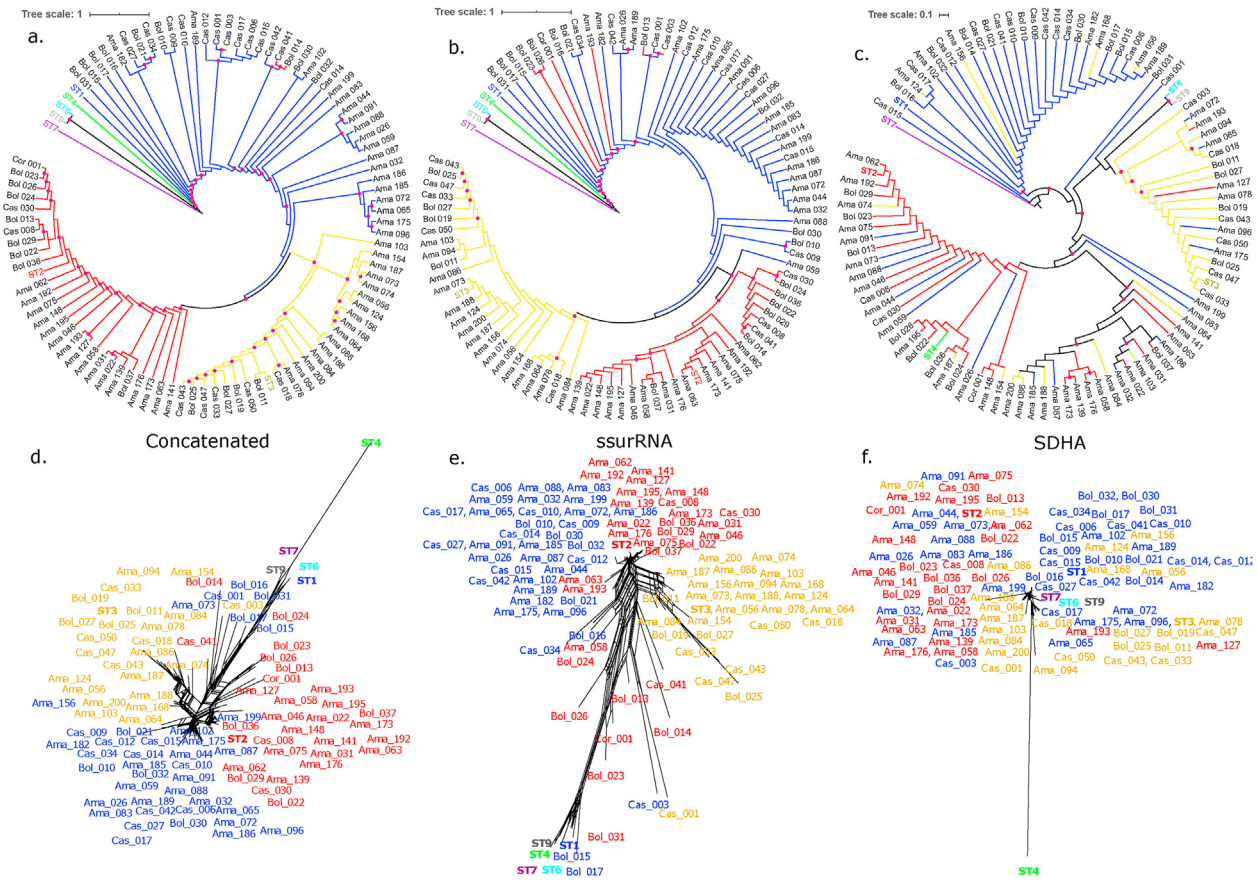

Figure 3. Phylogenetic reconstruction with the *ssu rRNA* and *SDHA* markers. The evolutionary history was inferred using the maximum likelihood (ML) method based on the GTR + F model with 1000 bootstrap replicates. The pink dot on each node represents the bootstrap support >80%. The initials of the departments where the samples come from are indicated at the tips of the branches, followed by the sample code (Cas: Casanare, Ama: Amazonas, Bol: Bolívar, Cor: Córdoba). Phylogenetic trees built from a.) the concatenated sequences obtained with the *ssu rRNA* and *SDHA* marker, b.) *ssu rRNA* marker and c.) *SDHA* marker. The corresponding colors are related with ST, where ST1 is in blue, ST2 in red, ST3 in yellow, ST4 in green, ST6 in turquoise, ST7 in purple and ST9 in gray color. ST7 was used as outgroup. A phylogenetic network, using Splitstree software, was built with the Neighbornet algorithm. d.) the concatenated sequences obtained with the *ssu rRNA* and *SDHA* marker, e.) *ssu rRNA* marker and f.) *SDHA* marker.

populations (Dxy) was higher between Casanare and Bolívar with the *ssu rRNA* gene, 0.24023, followed by ST1 and ST3 with a value of 0.20658 (Table 1).

4. Discussion

Blastocystis globally has been typed into different STs worldwide due to the use of barcoding that amplified 600 bp of the *ssu rRNA* gene, which could easily discriminate between STs [21], showing a high diversity within the genus. This DNA region is interesting because is expected to be highly conserved [25], instead, exhibits relative high diversity within *Blastocystis* STs. It could indicate that the divergence time might not have been sufficient to fix the alleles with their mutations [54] showing an overestimated diversity. In this case, the need to colonize a greater host diversity even with strong competition in the intestinal environment could be the reason to be changing. But, due to the unknown role of this microorganism inside the gut, it has been proposed to explore additional markers with different evolutionary rate in order to elucidate what is happening in these populations. A few authors, have reported the need for use of other genes or non-coding regions of the genome, such as the *internal transcribed spacers* (*ITS*) [27], the use of single copy markers from MROs, which allowed for the detection of coinfections even in the same ST [25] or the use of the *PFOR* gene that generated clades that were different from the STs, because they were subject to different selective pressures that showed an evolutionary history different from that of the *ssu rRNA* gene [26].

Due to these few genes that have been studied in *Blastocystis*, since the *ssu rRNA* gene is usually used for genotyping, it has not really been possible to associate a factor that allows determining why this microorganism looks so diverse between the different STs, if such diversity may be due, for example, to recombination processes that do not allow discrimination between subpopulations and that may be influencing their intestinal role. For this reason, the need arises to search for other genetic regions that allow a better understanding of what is influencing the role of this microorganism and the observed genetic changes that induce such an intriguing variation within the species. In our case, the existing diversity among the STs evaluated was evident, since during the *in silico* analysis carried out, we found that the detection of new genetic markers is quite complex, since genes that are expected to be highly conserved among eukaryotes are not easily redeemable between the different STs analyzed. This added to the scarce information and few existing data on the genome of this microorganism, make it even more difficult to identify molecular genetic targets that allow a more robust study on the genetic diversity of this microorganism. However, our results show the importance of continuing to evaluate other markers, such as the *SDHA* gene together with the *ssu rRNA gene*, where in some cases, it can be very useful not only in genotyping and evaluation of inter and intra ST diversity, but also in the understanding of the evolution that *Blastocystis* has been presenting and that continues to be the focus of different scientific debates.

In the current study, the subtyped samples were amplified with primers designed for the *SDHA* gene. Interestingly, the *SDHA* locus did

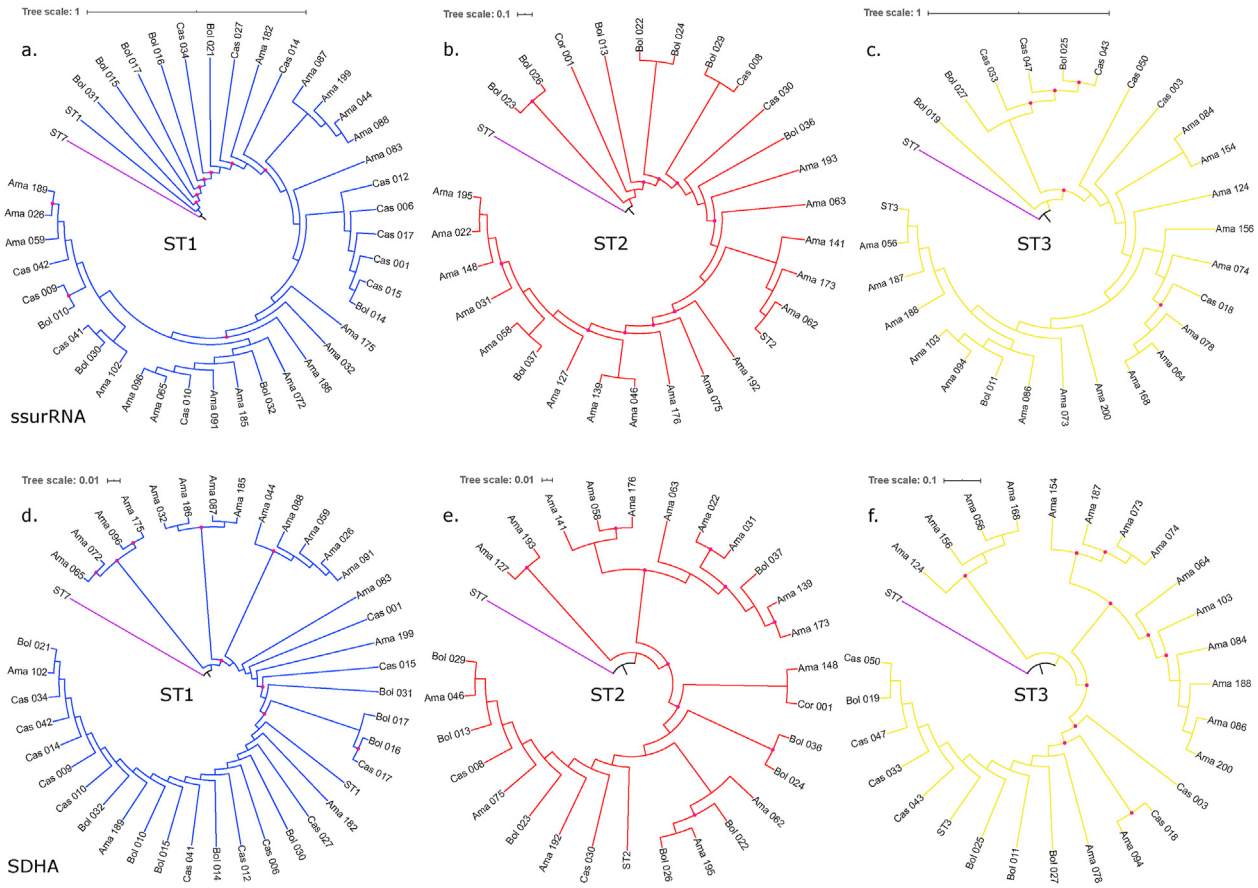

Figure 4. Phylogenetic reconstruction with the *ssu rRNA* and *SDHA* markers by ST. The evolutionary history was inferred using the maximum likelihood (ML) method based on the model established with ModelFinder: F81 + F + G4 and SYM + I to a.) and b.), respectively and GTR + F for c.) – f.). The pink dot on each node represents the bootstrap (1000 replicates) support >80%. The initials of the departments where the samples come from are indicated at the tips of the branches, followed by the sample code (Cas: Casanare, Ama: Amazonas, Bol: Bolívar, Cor: Córdoba). Phylogenetic trees built from the sequences obtained with the *ssu rRNA* by ST a.) ST1, b.) ST2, c.) ST3. Phylogenetic trees built from the sequences obtained with the *SDHA* ST d.) ST1, e.) ST2, f.) ST3. The corresponding colors are related with ST, where ST1 is in blue, ST2 in red, ST3 in yellow and ST7 in purple. ST7 was used as outgroup.

not show high genetic diversity compared to the *ssu rRNA* locus (Figure 2, Supplementary material Table S1) and showed a grouping in the phylogenetic reconstruction that was relatively associated with the STs and the collection areas (Figures 3 and 5). This suggested that this gene was probably under different evolutionary forces, which were similar to the *PFOR* gene reported in the Mexican samples [26]. Also, it was important to consider that ribosomal genes were highly utilized in eukaryotes and required a large production of ribosomes in times of massive growth, that could encode hundreds of copies of their transcriptional units [55], leading to recombination events and potentially greater diversity compared to the constitutive metabolic genes, which did not require high variation in order to maintain their function. The latter was characterized by a particular genomic composition with respect to their structure, composition, and conservation of sequences and expression [56]. For example, stabilizing selection for the conservation of metabolic functioning could eventually present recombination and had sufficient discrimination power without being subject to diversified selection [57]. However, significantly conserved genes would not be very informative for identifying variants successfully [57]. In this sense, we wanted to explore reticulation signals between the samples with both markers. The results obtained showed a greater number of reticulation signals in the concatenated and the *ssu rRNA* gene compared to *SDHA* (Figure 3d -f) and agree with the index determined for recombination, which was higher for the *ssu rRNA* gene (Supplementary material Table S1). Future studies should consider including more metabolic genes to understand the true genetic diversity of *Blastocystis*. It is also important to highlight that in some samples mixed infections with the *ssu rRNA* marker were evidenced and that therefore it is possible that STs are present in a lower proportion, which could be amplified with the *SDHA* marker and could influence the phylogenetic inconsistencies shown between both markers (Figure 3). This occurred with three ST2 samples that formed a cluster within ST1, because we obtained an electropherogram with multiple signals in the same position indicating mixed infection, but it was not possible to detect the other sequences present in the samples. In addition, in a certain proportion of the samples analyzed, the presence of multiple peaks was also evidenced in the electropherograms obtained, suggesting the presence of different alleles but at a lower frequency, which could not be detected with the sequencing technique used. The determination of the ST presented in a sample must be resolved using other sequencing techniques as next generation amplicon sequencing [23], that have made it possible to determine different alleles of the *ssu rRNA* gene in the same sample, showing mixed infections with several *Blastocystis* STs [58].

Phylogenetic reconstructions revealed different topology. In comparison of both genes we propose differences in evolutionary patterns (gradualism in the case of *ssu rRNA* and punctuated evolution in the case of *SDHA* gene) as was reported in *Cryptococcus* [51]. Besides, we could establish at least tree subgroups inside each ST using *SDHA* gene (Figure 4). Moreover, we found some clusters related with the geographical distribution of Amazonas and other cluster with Bolívar and Casanare mixed (Figure 5).

Figure 5. Relationships of geographical regions with clusters and haplotypic network. The relationships between geographic location and phylogenetic trees was visualized with https://microreact.org/showcase. These trees correspond with those constructed with IQtree. The location of each department in Colombia is shown. The size of the circles represents the frequency of the samples depending on each ST. Each phylogenetic tree is showing the clusters found inside each ST. a.) ST1, b.) ST2, c.) ST3. The network shows the genetic variability present among the sequences analyzed. d.) The length of the lines is in concordance with the amount of changes to generate a new haplotype. mv: mediate vector. mv, is a hypothesized (often ancestral) sequence which is required to connect existing sequences within the network with maximum parsimony. Without the median vector, there would be no shortest connection between the data set's sequences). The colors represent the regions where the samples were collected, which are indicated in the legend.

Table 1. Genetic differentiation among populations with *SDHA* and *ssu rRNA* locus.

Locus	Population 1	Population 2	Kxy	Gst	Fst	Dxy	Da
SDHA	Amazonas	Bolívar	12.29333	0.01140	0.03612	0.03396	0.00123
	Amazonas	Casanare	13.13600	0.03104	0.09839	0.03629	0.00357
	Bolívar	Casanare	12.97857	-0.00501	-0.00818	0.03585	-0.00029
ssu rRNA	Amazonas	Bolivar	42.44667	0.02318	0.10163	0.20506	0.02084
	Amazonas	Casanare	27.81500	0.01439	0.04528	0.13437	0.00608
	Bolivar	Casanare	49.72857	0.00318	0.03660	0.24023	0.00879
SDHA	ST1	ST2	10.86800	0.07731	0.19569	0.03002	0.00587
	ST1	ST3	14.28797	0.04534	0.15387	0.03947	0.00607
	ST2	ST3	14.20798	0.03763	0.14072	0.03925	0.00552
ssu rRNA	ST1	ST2	35.25926	0.07572	0.34376	0.17033	0.05855
	ST1	ST3	42.76134	0.08018	0.45593	0.20658	0.09418
	ST2	ST3	35.99288	0.05890	0.44264	0.17388	0.07697

Kxy: Average proportion of nucleotide differences between populations.
Gst: Genetic differentiation index based on the frequency of haplotypes.
Fst: estimate gene flow from nucleotide sequences.
Dxy: The average number of nucleotide substitutions per site between populations.
Da: The number of net nucleotide substitutions per site between populations.

Besides, to detect intraspecific variation, it is necessary to increase the phylogenetic resolution using other markers with lower diversity. For this reason, we propose in addition to using *ssu rRNA* to include other markers as *SDHA* which let to obtain clusters with enough support of bootstrap and detect the intra and inter ST variation. These results agree with the TE [47] value obtained to *SDHA*, which showed more capability for this gene to cluster some samples and visualize subgroups inside each ST. However, it is important to evaluate this marker in other STs different from those studied in this study.

On the other hand, several studies have tried to relate this diversity to the different symptoms [59, 60, 61], hosts [7, 12], and even socioeconomic factors [40, 62]. However, these aspects have not yet been fully explicated, and this parasite is associated with many unknown factors regarding the biology and evolution of this microorganism and even the role of this protist in different hosts. In this study, we found a relationship between STs and geographic distribution by department, where the use of the two markers let us to observe a grouping corresponding to the department of Amazonas and another that includes samples from both Bolívar and Casanare. Geographically, Bolívar and Casanare are closer and there is a greater movement of humans between these two regions, so the possibility of transmission between these two areas is higher, instead the samples from the Amazon form a separate cluster, because it corresponds to samples from indigenous communities These communities are located near to forest regions and the possibility of genetic exchange between microorganisms is less. Also, within ST3, we were able to observe two separate groups within the Amazon region, showing different populations of this microorganism in this region (Figures 4 and 5).

Haplotype networks built with *SDHA* sequences in the current study exhibited great diversity of the haplotypes distributed in the different

geographic regions and among the STs. Some haplotypes were found to be associated specifically in the sampled geographic regions, principally in the Amazonas, but some shared haplotypes were observed by region and by ST (Figure 5d and Supplementary material Figure S1). This could be explained by different reasons, such as inadequate sampling (low numbers of samples collected), limited divergence, hybridization, cryptic speciation, and incomplete lineage sorting [63, 64, 65].

Interestingly, when the ST analysis was conducted, groups of haplotypes that were associated with ST1 and ST2 were observed, while the haplotypes associated with ST3 were seen at a greater distribution in the sampled regions (Supplementary material Figure S1). The latter could be in accordance with other reports made, where the ST3 was one of the subtypes that had the greatest diversity [5, 58, 66] and in turn, coincided with the fact that this subtype was the most variable compared to the others since it had more than 50 different alleles reported in the database for ST determination (https://pubmlst.org).

Nevertheless, despite the genetic diversity found and the great geographical distances between some of the sampled departments, no evidence of genetic differentiation for *SDHA* was found between Amazonas - Bolívar and Bolívar – Casanare and for *ssu rRNA* gene, just moderate genetic differentiation between Amazonas – Bolívar was found. High genetic differentiation between STs with either of the two loci evaluated was observed. Furthermore, estimators used to make comparisons between populations, such as the Kxy, Dxy, and Da showed more nucleotide substitutions and differences in the *ssu rRNA* gene when we compared by geographical regions and STs (Table 1). The D Tajima test [50] was based on the comparison of the number of differences using pairs of nucleotide sequences and the number of segregating sites, where the negative values indicated the high frequency of rare alleles and was a signal of population expansion with both markers evaluated.

These results are similar to that observed in protozoa such as *G. duodenalis* [65] and some of the coding genes for the variant surface antigens in *Plasmodium vivax* [67]. This expansion might be due to a possible selection that has occurred in the *Blastocystis* genetic pool, which varied depending on the mutational rate or the recombination rate [68]. Thus, it is necessary to evaluate the reproductive mechanisms of this microorganism, and the possibility of recombination events that tend to maintain the high variation observed, as has happened in protozoa such as *T. cruzi* [69] and *Leishmania* [70] and has been proposed for *G. duodenalis* [71, 72].

In the case of *Blastocystis*, the evidence of expansion in the samples analyzed might be related to the wide range of hosts presented by this protist, where the capacity for infection in mammals and other animals might have influenced the great diversity, and effectively increased the population size [73]. This population expansion likely occurred due to the ease of fecal transmission through contaminated food and water where the cysts were dispersed to new hosts moving easily throughout the Colombian territory, as has been reported for *G. duodenalis* [65] and for some nematodes, where it has been proposed that host movement is key to gene flow and scattering of these rare alleles [74]. Future studies are necessary to verify these dynamics of transmission and population genetics. Moreover, studies using whole genomes of *Blastocystis* will allow us to understand if recombination is occurring.

The *SDHA* marker could be considered as a possible candidate for the discrimination of groups presented within a population of this microorganism. However, this marker should be evaluated with a larger number of samples and hosts (including animals) as well as more geographic regions in Colombia and South America. Similarly, it could be very useful, along with other markers such as *ssu rRNA*, in the analysis of the genetic diversity and population structure of *Blastocystis*, since in our case it allowed us to show selective pressure forces leading to the expansion process that could be explained by the high number of hosts capable of becoming infected with this microorganism. However, more markers are required to obtain a more robust analysis of what has happened within the populations of this protozoan.

Some limitations of our study include the use of only one additional genetic marker, which could bias these assumptions, the low number of samples that we could amplify because we did not culture the samples and amplify directly from DNA extracted from stool samples, so the amount of DNA available from genes with just one copy or lower number of copies make difficult the amplification. Besides the concentration of this microorganism could be low. It is necessary to increase the number of samples evaluated from different regions to verify our assumptions. It is necessary to extend this study to other STs, since we just could test the ST1, ST2 and ST3.

5. Conclusions

The *SDHA* marker could eventually be used together with *ssu rRNA* marker, for typing and to evaluate the evolution, diversity, and population structure of *Blastocystis*. However, it was necessary to explore more regions of the genome that allowed for the development of new markers to study this microorganism and to elucidate the incongruity still present around the biology of this protist.

Declarations

Author contribution statement

Adriana Higuera: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Marina Muñoz, Juan David Ramírez: Conceived and designed the experiments; Wrote the paper.

Myriam Consuelo López, Patricia Reyes, Plutarco Urbano, Oswaldo Villalobos: Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS) (Colciencias) within the framework of the National Program for Promoting Research Training (sponsorship call 757). This work was also funded by the Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias) through the project “Desarrollo de una estrategia y plataforma de Multilocus Sequence Typing (MLST) para la caracterización molecular de siete parásitos intestinales”,code 122271250.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2020.e05387.

Acknowledgements

We thank to Lee O'Brien Andersen and Christen Rune Stensvold from the Department of Microbiology and Infection Control, Statens Serum Institut, to Bruce Curtis and Andrew Rogers from the Centre for Comparative Genomics & Evolutionary Bioinformatics, Biochemistry and Molecular Biology, Dalhousie University and to Ivan Wawrzyniak from Université Clermont Auvergne, 3iHP, CNRS, Laboratoire Microorganismes: Génome et Environnement for sharing the raw reads of the publicly available genomic data and to Angie Johana Sanchez for her help in the molecular detection of *Blastocystis* in the fecal samples from Amazonas. We thank the Dirección.

References

[1] J.D. Silberman, M.L. Sogin, D.D. Leipe, C.G. Clark, Human parasite finds taxonomic home, Nature 380 (6573) (1996) 398.

[2] N. Arisue, T. Hashimoto, H. Yoshikawa, Y. Nakamura, G. Nakamura, F. Nakamura, et al., Phylogenetic position of Blastocystis hominis and of stramenopiles inferred from multiple molecular sequence data, J. Eukaryot. Microbiol. 49 (1) (2002) 42–53.

[3] A.L. Londono-Franco, J. Loaiza-Herrera, F.M. Lora-Suarez, J.E. Gomez-Marin, [Blastocystis sp. frequency and sources among children from 0 to 5 years of age attending public day care centers in Calarca, Colombia], Biomedica 34 (2) (2014) 218–227.

[4] J.D. Ramírez, A. Sánchez, C. Hernández, C. Flórez, M.C. Bernal, J.C. Giraldo, et al., Geographic distribution of human Blastocystis subtypes in South America, Infect. Genet. Evol. 41 (2016) 32–35.

[5] C.R. Stensvold, M. Alfellani, C.G. Clark, Levels of genetic diversity vary dramatically between Blastocystis subtypes, Infect. Genet. Evol. 12 (2) (2012) 263–273.

[6] D. El Safadi, L. Gaayeb, D. Meloni, A. Cian, P. Poirier, I. Wawrzyniak, et al., Children of Senegal River Basin show the highest prevalence of Blastocystis sp. ever observed worldwide, BMC Infect. Dis. 14 (2014) 164.

[7] I. Wawrzyniak, P. Poirier, E. Viscogliosi, M. Dionigia, C. Texier, F. Delbac, et al., Blastocystis, an unrecognized parasite: an overview of pathogenesis and diagnosis, Ther. Adv. Infect. Dis. 1 (5) (2013) 167–178.

[8] K.S. Tan, H. Mirza, J.D. Teo, B. Wu, P.A. Macary, Current views on the clinical relevance of blastocystis spp, Curr. Infect. Dis. Rep. 12 (1) (2010) 28–35.

[9] R. Verma, K. Delfanian, Blastocystis hominis associated acute urticaria, Am. J. Med. Sci. 346 (1) (2013) 80–81.

[10] T. Roberts, D. Stark, J. Harkness, J. Ellis, Subtype distribution of Blastocystis isolates identified in a Sydney population and pathogenic potential of Blastocystis, Eur. J. Clin. Microbiol. Infect. Dis. 32 (3) (2013) 335–343.

[11] C.R. Stensvold, D.B. Christiansen, K.E.P. Olsen, H.V. Nielsen, Blastocystis sp. subtype 4 is common in Danish blastocystis-positive patients presenting with acute diarrhea, Am. J. Trop. Med. Hyg. 84 (6) (2011) 883–885.

[12] J.D. Ramirez, L.V. Sanchez, D.C. Bautista, A.F. Corredor, A.C. Florez, C.R. Stensvold, Blastocystis subtypes detected in humans and animals from Colombia, Infect. Genet. Evol. 22 (2014) 223–228.

[13] K.S.W. Tan, New insights on classification, identification, and clinical relevance of blastocystis spp, Clin. Microbiol. Rev. 21 (4) (2008) 639–665.

[14] N. Abe, Z. Wu, H. Yoshikawa, Molecular characterization of Blastocystis isolates from primates, Vet. Parasitol. 113 (3-4) (2003) 321–325.

[15] C.G. Clark, Extensive genetic diversity in Blastocystis hominis, Mol. Biochem. Parasitol. 87 (1) (1997) 79–83.

[16] H. Yoshikawa, I. Nagono, E.H. Yap, M. Singh, Y. Takahashi, DNA polymorphism revealed by arbitrary primers polymerase chain reaction among Blastocystis strains isolated from humans, a chicken, and a reptile, J. Eukaryot. Microbiol. 43 (2) (1996) 127–130.

[17] N. Arisue, T. Hashimoto, H. Yoshikawa, Sequence heterogeneity of the small subunit ribosomal RNA genes among blastocystis isolates, Parasitology 126 (Pt 1) (2003) 1–9.

[18] E. Gentekaki, B.A. Curtis, C.W. Stairs, V. Klimes, M. Elias, D.E. Salas-Leiva, et al., Extreme genome diversity in the hyper-prevalent parasitic eukaryote Blastocystis, PLoS Biol. 15 (9) (2017), e2003769.

[19] I. Wawrzyniak, D. Courtine, M. Osman, C. Hubans-Pierlot, A. Cian, C. Nourrisson, et al., Draft genome sequence of the intestinal parasite Blastocystis subtype 4-isolate WR1, Genomics Data 4 (2015) 22–23.

[20] F. Denoeud, M. Roussel, B. Noel, I. Wawrzyniak, C. Da Silva, M. Diogon, et al., Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite, Genome Biol. 12 (3) (2011) R29.

[21] S.M. Scicluna, B. Tawari, C.G. Clark, DNA barcoding of blastocystis, Protist 157 (1) (2006) 77–85.

[22] C.R. Stensvold, G.K. Suresh, K.S. Tan, R.C. Thompson, R.J. Traub, E. Viscogliosi, et al., Terminology for Blastocystis subtypes–a consensus, Trends Parasitol. 23 (3) (2007) 93–96.

[23] J.G. Maloney, A. Molokin, M.J.R. da Cunha, M.C. Cury, M. Santin, Blastocystis subtype distribution in domestic and captive wild bird species from Brazil using next generation amplicon sequencing, Parasite Epidemiol Control 9 (2020), e00138.

[24] C.G. Clark, M. van der Giezen, M.A. Alfellani, C.R. Stensvold, Recent developments in Blastocystis research, Adv. Parasitol. 82 (2013) 1–32.

[25] P. Poirier, D. Meloni, C. Nourrisson, I. Wawrzyniak, E. Viscogliosi, V. Livrelli, et al., Molecular subtyping of Blastocystis spp. using a new rDNA marker from the mitochondria-like organelle genome, Parasitology 141 (5) (2014) 670–681.

[26] P. Alarcon-Valdes, G. Villalobos, W.A. Martinez-Flores, E. Lopez-Escamilla, N.R. Gonzalez-Arenas, M. Romero-Valdovinos, et al., Can the pyruvate: ferredoxin oxidoreductase (PFOR) gene be used as an additional marker to discriminate among Blastocystis strains or subtypes? Parasit Vectors 11 (1) (2018) 564.

[27] G. Villalobos, G.E. Orozco-Mosqueda, M. Lopez-Perez, E. Lopez-Escamilla, A. Cordoba-Aguilar, L. Rangel-Gamboa, et al., Suitability of internal transcribed spacers (ITS) as markers for the population genetic structure of Blastocystis spp, Parasit Vectors 7 (2014) 461.

[28] I. Wawrzyniak, M. Roussel, M. Diogon, A. Couloux, C. Texier, K.S. Tan, et al., Complete circular DNA in the mitochondria-like organelles of Blastocystis hominis, Int. J. Parasitol. 38 (12) (2008) 1377–1382.

[29] V. Perez-Brocal, C.G. Clark, Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite Blastocystis: complete sequences, gene content, and genome organization, Mol. Biol. Evol. 25 (11) (2008) 2475–2482.

[30] A. Stechmann, K. Hamblin, V. Pérez-Brocal, D. Gaston, G.S. Richmond, M. van der Giezen, et al., Organelles in Blastocystis that blur the distinction between mitochondria and hydrogenosomes, Curr. Biol. 18 (8) (2008) 580–585.

[31] P.B. Christmas, J.F. Turrens, Separation of NADH-fumarate reductase and succinate dehydrogenase activities in Trypanosoma cruzi, FEMS Microbiol. Lett. 183 (2) (2000) 225–228.

[32] D. Kregiel, Succinate dehydrogenase of Saccharomyces cerevisiae–the unique enzyme of TCA cycle–current knowledge and new perspectives, in: Dehydrogenases, 2012 [Internet]. [211-34]. Available from: https://www.intechopen.com/books/dehydrogenases/succinate-dehydrogenase-of-saccharomyces-cerevisiae-the-unique-enzyme-of-tca-cycle-current-knowledge.

[33] S. Huang, A.H. Millar, Succinate dehydrogenase: the complex roles of a simple enzyme, Curr. Opin. Plant Biol. 16 (3) (2013) 344–349.

[34] O.G. Berg, C.G. Kurland, Why mitochondrial genes are most often found in nuclei, Mol. Biol. Evol. 17 (6) (2000) 951–961.

[35] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.

[36] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (5) (2004) 1792–1797.

[37] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874.

[38] M. Inouye, H. Dashnow, L.A. Raven, M.B. Schultz, B.J. Pope, T. Tomita, et al., SRST2: rapid genomic surveillance for public health and hospital microbiology labs, Genome Med. 6 (11) (2014) 90.

[39] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780.

[40] A. Sanchez, M. Munoz, N. Gomez, J. Tabares, L. Segura, A. Salazar, et al., Molecular epidemiology of giardia, blastocystis and cryptosporidium among indigenous children from the Colombian Amazon basin, Front. Microbiol. 8 (2017) 248.

[41] C.R. Stensvold, U.N. Ahmed, L.O. Andersen, H.V. Nielsen, Development and evaluation of a genus-specific, probe-based, internal-process-controlled real-time PCR assay for sensitive and specific detection of Blastocystis spp, J. Clin. Microbiol. 50 (6) (2012) 1847–1851.

[42] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (6) (2017) 587–589.

[43] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, et al., IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (5) (2020) 1530–1534.

[44] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (W1) (2016) W242–W245.

[45] D.H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies, Mol. Biol. Evol. 23 (2) (2006) 254–267.

[46] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1) (1999) 37–48.

[47] N. Tomasini, J.J. Lauthier, M.S. Llewellyn, P. Diosque, MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms, Infect. Genet. Evol. 20 (2013) 188–196.

[48] E. Lasek-Nesselquist, D.M. Welch, R.C. Thompson, R.F. Steuart, M.L. Sogin, Genetic exchange within and between assemblages of Giardia duodenalis, J. Eukaryot. Microbiol. 56 (6) (2009) 504–518.

[49] J.G. Maloney, A. Molokin, M. Santin, Next generation amplicon sequencing improves detection of Blastocystis mixed subtype infections, Infect. Genet. Evol. 73 (2019) 119–125.

[50] F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism, Genetics 123 (3) (1989) 585–595.

[51] M. Muñoz, M. Camargo, J.D. Ramírez, Estimating the intra-taxa diversity, population genetic structure, and evolutionary pathways of, Front. Genet. 9 (2018) 148.

[52] B. Singh, Evolutionary biology: concepts of punctuated equilibrium, concerted evolution and coevolution, J. Sci. Res. 58 (2014) 15–26.

[53] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (6) (1984) 1358–1370.

[54] T.J. Anderson, The dangers of using single locus markers in parasite epidemiology: Ascaris as a case study, Trends Parasitol. 17 (4) (2001) 183–188.

[55] T.H. Eickbush, D.G. Eickbush, Finely orchestrated movements: evolution of the ribosomal RNA genes, Genetics 175 (2) (2007) 477–485.

[56] K. Wei, T. Zhang, L. Ma, Divergent and convergent evolution of housekeeping genes in human-pig lineage, PeerJ 6 (2018), e4840.

[57] M. Maiden, Multilocus sequence typing of bacteria, Annu. Rev. Microbiol. 60 (2006) 561–588.

[58] L. Rojas-Velázquez, P. Morán, A. Serrano-Vázquez, L.D. Fernández, H. Pérez-Juárez, A.C. Poot-Hernández, et al., Genetic diversity and distribution of blastocystis subtype 3 in human populations, with special reference to a rural population in Central Mexico, BioMed Res. Int. 2018 (2018).

[59] R.T. Mohamed, M.A. El-Bali, A.A. Mohamed, M.A. Abdel-Fatah, M.A. El-Malky, N.M. Mowafy, et al., Subtyping of Blastocystis sp. isolated from symptomatic and asymptomatic individuals in Makkah, Saudi Arabia, Parasit Vectors 10 (1) (2017) 174.

[60] J.D. Ramirez, C. Florez, M. Olivera, M.C. Bernal, J.C. Giraldo, Blastocystis subtyping and its association with intestinal parasites in children from different geographical regions of Colombia, PloS One 12 (2) (2017), e0172586.

[61] F. Zulfa, I.P. Sari, A. Kurniawan, Association of Blastocystis subtypes with diarrhea in children, IOP Publishing. J. Phys.: Conference Series (2017), 012031.
[62] X. Villamizar, A. Higuera, G. Herrera, L.R. Vasquez-A, L. Buitron, L.M. Muñoz, et al., Molecular and descriptive epidemiology of intestinal protozoan parasites of children and their pets in Cauca, Colombia: a cross-sectional study, BMC Infect. Dis. 19 (1) (2019) 190.
[63] M.W. Hart, J. Sunday, Things fall apart: biological species form unconnected parsimony networks, Biol. Lett. 3 (5) (2007) 509–512.
[64] S. Tarcz, E. Przyboś, M. Surmacz, An assessment of haplotype variation in ribosomal and mitochondrial DNA fragments suggests incomplete lineage sorting in some species of the Paramecium aurelia complex (Ciliophora, Protozoa), Mol. Phylogenet. Evol. 67 (1) (2013) 255–265.
[65] S.H. Choy, M.A.K. Mahdy, H.M. Al-Mekhlafi, V.L. Low, J. Surin, Population expansion and gene flow in Giardia duodenalis as revealed by triosephosphate isomerase gene, Parasites Vectors 8 (2015) 454.
[66] D. Meloni, P. Poirier, C. Mantini, C. Noël, N. Gantois, I. Wawrzyniak, et al., Mixed human intra- and inter-subtype infections with the parasite Blastocystis sp, Parasitol. Int. 61 (4) (2012) 719–722.
[67] U.H. Son, S.D. Dinzouna-Boutamba, S. Lee, H.S. Yun, J.Y. Kim, S.Y. Joo, et al., Diversity of *vir* genes in *Plasmodium vivax* from endemic regions in the Republic of Korea: an initial evaluation, Kor. J. Parasitol. 55 (2) (2017) 149–158.
[68] R.R. Hudson, N.L. Kaplan, Deleterious background selection with recombination, Genetics 141 (4) (1995) 1605–1617.
[69] M. Lewis, M. Llewellyn, M. Yeo, L. Messenger, M. Miles, Experimental and natural recombination in Trypanosoma cruzi, in: American Trypanosomiasis Chagas Disease, Elsevier, 2017, pp. 455–473 [Internet].
[70] E. Inbar, J. Shaik, S.A. Iantorno, A. Romano, C.O. Nzelu, K. Owens, et al., Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in Leishmania, PLoS Genet. 15 (5) (2019), e1008042.
[71] S.M. Cacciò, H. Sprong, Giardia duodenalis: genetic recombination and its implications for taxonomy and molecular epidemiology, Exp. Parasitol. 124 (1) (2010) 107–112.
[72] M.A. Cooper, R.D. Adam, M. Worobey, C.R. Sterling, Population genetics provides evidence for recombination in Giardia, Curr. Biol. 17 (22) (2007) 1984–1988.
[73] L.G. Barrett, P.H. Thrall, J.J. Burdon, C.C. Linde, Life history determines genetic structure and evolutionary potential of host-parasite interactions, Trends Ecol. Evol. 23 (12) (2008) 678–685.
[74] M.S. Blouin, C.A. Yowell, C.H. Courtney, J.B. Dame, Host movement and the genetic structure of populations of parasitic nematodes, Genetics 141 (3) (1995) 1007–1014.