# ABSTRACT

**Identification and Molecular Characterization of Common Enteric Parasites in Kolkata with Special Reference to *Entamoeba* spp**

Diarrheal diseases are a major health concern, affecting millions of people worldwide every year and causing a high rate of illness and death. The three most common parasites that cause diarrhoea are *Entamoeba* spp, *Giardia* spp, and *Cryptosporidium* spp. In this study, we have identified and characterized the most prevalent enteric parasites responsible for diarrhoea in Kolkata and adjacent areas, with a specific focus on amoebic infection. The primary causative agent for amoebic infection is *Entamoeba histolytica*. Amoebic infection is a complex issue as there are several species that are morphologically indistinguishable from each other, including *E. dispar, E. bangladeshi*, and *E. moshkovskii*. Cysts of the other non-pathogenic amoeba *E. hartmanni* also can be misidentified as the pathogenic *E. histolytica* under a microscope. This makes it challenging to accurately estimate the prevalence of each species and its potential to cause disease in humans.

Over a period of three years, we conducted active surveillance among 6051 patients with diarrhoea in Kolkata, India, to study the prevalence, genetic and phylogenetic structure of the common enteric parasites population. We employed microscopy, PCR, DNA sequencing, and various tools such as MEGA X, DnaSP, PopArt, Clustal W, MultAlin, Seaview, and Tandem Repeats Finder for analyzing the sequencing data. We used GraphPad Prism for statistical analysis.

The prevalence study found *Giardia* in 6.18% and *Cryptosporidium* in 4.36% of cases, indicating a high occurrence of both infections. Furthermore, our study revealed that both cryptosporidiosis and giardiasis were significantly more prevalent among children below 4 years of age. During the study, we also conducted genotyping of *Cryptosporidium* spp using the HSP70 gene. Our results reported the presence of *Cryptosporidium viatorum* in eastern India. This emerging enteropathogen was first described in 2012 and has a considerable prevalence rate. Our DNA sequence analysis revealed that *Cryptosporidium hominis* is the most dominant species in the study regions. This study also reported 4.84% of the examined samples were positive for *Entamoeba* spp. Among these amoebic infections, *E. histolytica* accounted for 0.86% of the cases, *E. moshkovskii* for 3.12%, and *E. dispar* for 0.13%, while the remaining 0.73% could not be identified to a specific species. Statistical analysis confirms a significant association ($P<0.0001$) between *E. moshkovskii* and diarrhoeal incidence. These findings underscore the significant prevalence of *E. moshkovskii* as a major parasitic infection in eastern India. Further analysis found that *E. moshkovskii* infection was significantly associated with the age group of 5-12 years. Surprisingly, this parasitic infection exhibits a distinctive seasonal pattern, not previously seen in other parasitic infections. DNA sequencing revealed that the local *E. moshkovskii* strains were 99.59%-100% identical to the prototype (GenBank: KP722605.1). The study found certain SNPs that showed a correlation with clinical features, but it is not necessarily indicative of direct control over pathogenicity. Neutrality tests of different coinfected subgroups indicated deviations from neutrality and implied population expansion after a bottleneck event or a selective sweep and/or purifying selection in co-infected subgroups. The majority of $F_{ST}$ values of different coinfected subgroups were <0.25, indicating low to moderate genetic differentiation within the subgroups of this geographical area. Additionally, we utilized MLST (Multi-Locus Sequence Typing) to identify genetic markers and potential SNPs associated with clinical features in isolates of *E. moshkovskii*. The markers used in our MLST analysis included KERP1, a protein rich in lysine and glutamic acid, as well as amoebapore C (apc) and Chitinases. By integrating these markers into our genotyping system, we were able to identify one genotype (M1) that was significantly associated ($p=0.0394$) with the sole incidence of diarrhoea. We also identified a few potential SNPs in the three genes that were linked to clinical features. Furthermore, using PCR, RFLP, DNA sequencing, and phylogenetic analysis, the true natural animal reservoir of *E. moshkovskii* has been effectively revealed. The findings indicate that pigs serve as the true animal reservoir for this pathogen, with a prevalence rate of 5.4%. This evidence supports the potential for zoonotic transmission of *E. moshkovskii* from pigs to humans. The implications of my study strongly suggest the possibility of such zoonotic transmission, emphasizing the need for further investigation and increased awareness of this potential health risk.

We also employ a strain typing approach in *E. histolytica* isolates that utilize multiple loci, including SREHP and three polymorphic non-coding loci (tRNA linked array N-K2, loci 1-2, and 5-6), for high-resolution analysis. Distinct clinical phenotype isolates underwent amplification and sequencing of studied loci. The nucleotide sequences were analysed using Tandem Repeats Finder to detect short tandem repeats (STRs). This study found significant polymorphism in the size and number of PCR fragments at SREHP and 5-6 locus, while the 1-2 locus and NK2 locus showed variations in PCR product sizes. Out of 41 genotypes, two (I6 and I41) were significantly associated with their respective disease outcomes and were found in multiple isolates. We observed that I6 was linked with a symptomatic outcome, with a statistically significant p-value of 0.0183. Additionally, we found that I41 was associated with ALA disease outcome, with a p-value of 0.0089. Our study revealed new repeat units not previously reported, unveiling the genetic composition of *E. histolytica* strains in India, associated with distinct disease manifestations. Finally, this study has developed a new multiplex PCR and a PCR-RFLP approach, in addition to qPCR, which enables us to accurately detect and distinguish between different parasite species that appear identical. Selecting the appropriate method for identifying and studying parasites is crucial and depends on the intended purpose, available resources, and facilities.

DR. SANDIPAN GANGULY
Senior Deputy Director (Scientist F) & Head
Division of Parasitology
National Institute of Cholera & Enteric Diseases
Indian Council of Medical Research
Department of Health Research
Ministry of Health and Family Welfare
Govt. of India
P-[illegible] Road, Scheme XM, Beliaghata
Kolkata 700 010, West Bengal, India

11/09/23

Sanjit Kumar Sahu
11/09/23