Molecular Diagnosis and Characterization of Clinical Isolates of *Entamoeba histolytica*, *Giardia lamblia* and *Cryptosporidium* Species from the United Arab Emirates and South Africa

By

Ali Mohammed Kamal ElBakri

(Student Number: 11591474)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Phylosophy in Microbiology in the University of Venda

Promoter:

Dr. Samie A. (Department of Microbiology, University of Venda)

Co-Promoters:

Prof. Bessong PO (Microbiology Department, University of Venda)

Prof. Potgieter N (Microbiology Department, University of Venda)

Dr AbuOdeh, RO (Medical Laboratory Sciences Department, University of Sharjah)

ABSTRACT

Diarrhea is an important cause of morbidity and mortality, globally. *Entamoeba histolytica, Giardia lamblia* and *Cryptosporidium* spp., are the most common diarrhea-causing parasitic protozoa. Diagnosis of these parasites is usually performed by microscopy. However, microscopy lacks sensitivity and specificity. Replacing microscopy with more sensitive and specific nucleic acid based methods is hampered by the higher costs, in particular in developing countries. Moreover, data on the molecular epidemiology and zoonotic potential of these organisms are still lacking in the Arabia Gulf countries particularly in the United Arab Emirates (UAE) and in South Africa. Thus, the development of affordable molecular diagnostic methods and an understanding of the molecular epidemiology of the three mentioned parasites will have a profound impact on the public health systems on countries worldwide, particularly in the UAE and South Africa.

Amoebiasis is one of the most important infectious diseases afflicting mainly tropical and subtropical countries. A study was carried out in the Sharjah Emirate, UAE in order to detect and differentiate *Entamoeba histolytica, E. dispar* and *E. moshkovskii* in 120 fecal samples collected from the Sharjah municipality public health clinic by Microscopy, ELISA and a nested PCR (Chapter 2). The PCR was positive for *E. histolytica, E. dispar* and *E. moshkovskii* (collectively referred to as *Entamoeba* complex) in 19.2% of the samples. Of those, 10% were single infections by *E. histolytica*; 2.5% by *E. dispar*; and 2.5% by *E. moshkovskii.* The nested PCR detected mixed infections by both *E. histolytica* and *E. dispar* in 3.3% and *E. dispar* and *E. moshkovskii* in 0.8% of the tested samples. The TechLab ELISA kit failed to detect *E. histolytica* in any of the *E. histolytica* PCR positive samples. Therefore, the overall percentage of *E. histolytica* including those found in mixed infections was 13.3%. Compared to nested PCR, microscopy was found to have an overall sensitivity of 52.2% and a specificity of 75.2% for the detection of *Entamoeba* complex. The present study indicates that *E. histolytica* is present in the UAE with a prevalence of 13.3%. However, studies with larger sample population need to be conducted in order to confirm these findings. The findings of this study recommends the the use of PCR in both the routine diagnosis of amoebiasis and epidemiological survey in the UAE. The

genetic diversity of 20 *E. histolytica* isolates from asymptomatic individuals from the UAE, was investigated by analyzing polymorphism in the serine-rich *E. histolytica* gene by nested PCR amplification followed by restriction fragment length polymorphism analysis (RFLP) on DNA extracted directly from stool samples. The serine rich *Entamoeba histolytica* protein (SREHP) gene was successfully amplified in 15 out of all 20 *E. histolytica*-positive samples. Twelve different profiles were obtained from the 15 isolates successfully amplified and digested. Thus, demonstrating extensive genetic variability and reinforcing the argument that *E. histolytica* has an extremely polymorphic genetic structure. Despite the sample size limitation, a major finding in the study was the occurrence of one profile common to an India isolate whilst another profile common to a Pakistan isolate; indicating the possibility of clonal infection. Furthermore, one isolate from a Bangladeshi expatriate was identical to 2 asymptomatic Bangladesh isolates reported in an earlier study. No clear association between the different genotypes and the study population demographics was noted. The results also indicated the possibility of strains clustering by region (Chapter 3).

The study further evaluated the usefulness of the six pairs of tRNA-linked short tandem repeats (STR) loci in identifying the genetic polymorphism using PCR and sequencing in order to clarify the genotypic differences among *E. histolytica* isolates among expatriates from different geographic regions residing in Sharjah, UAE. The genomic DNA of a total of 20 *E. histolytica* samples (19 asymptomatic and 1 diarrheic/dysenteric cases) was analyzed. For genotype analysis, Loci S^{TGA}–D and R-R showed the most number of amplified isolates whereas NK2 amplified the least number. Loci D-A, A-L, and S-Q were all excluded from the study. The sequences obtained for both R-R and S^{TGA}–D loci were aligned using the Codoncode Aligner. Despite successful amplification of R-R locus, sequencing revealed all isolates to be gut flora bacteria such as *Bacteroides vulgates*. However, for the S^{TGA}-D locus, 10 representative isolates were sequenced. BLAST analysis of the 10 sequences unambiguously identified them as distinct fragments of *E. histolytica*. Five isolates successfully assembled and their sequence homology was comparable to published sequences. Four STRs variations in the S^{TGA}-D locus were detected. Two isolates were exactly identical. Unpredictably though, those two isolates were from a symptomatic Iraqi national and an asymptomatic Afghan national residing in Sharjah,

UAE; respectively. The remaining three isolates exhibited dissimilar nucleotide sequences. No meaningful correlation between infections with *E. histolytica* and age, sex or parasite genotypes was observed. The results demonstrate the extensive genetic variability among *E. histolytica* clinical isolates independent of geographic location. The genetic diversity of the S^{TGA}-D locus shows it to be suitable for epidemiological studies such as the characterization of transmission routes, clinical outcome and the epidemiology of the parasite in the UAE (Chapter 4).

Giardia lamblia, the intestinal protozoan causing giardiasis, infects about 200 million individuals annually worldwide. The spectrum of clinical manifestations ranges from an asymptomatic infection to a severe diarrhea. There are 7 different assemblages of *G. lamblia*, where two (A and B) assemblages have been associated with humans, while the others mostly infect animals. In the present study, a PCR protocol was used to detect and identify *G. lamblia* using DNA extracted from human stool samples in order to determine its prevalence and genetic diversity in Sharjah, UAE (Chapter 5). One hundred and eleven healthy expatriates residing in Sharjah and attending the Sharjah Municipal Public Clinic (SMPHC) were screened for *G. lamblia* using a nested PCR amplification of the *ssu-rRNA* gene. Samples identified as positive for *G. lamblia* were genotyped using a nested PCR amplifying the triosephosphate isomerase (*tpi*) gene to differentiate between the two human assemblages (A and B). The PCR products of (n= 11 for *ssu-rRNA* and n= 5 for the *tpi* gene) samples were also sequenced in order to confirm the presence of the organism and to determine the potential phylogenetic relationship between the isolates. Of the 111 samples tested for *Giardia* spp., 60.4% were identified as positive for the *ssu-rRNA* gene. When genotyped for the *tpi* gene, 18.9% were assemblage A, 17.1% belonged to assemblage B and 5.6% showed patterns that were compatible with the presence of mixed infections. However, after excluding 3 isolates for lack of demographic details, of the remaining 64 isolates, assemblage A constituted 19.4% (21), assemblage B 17.6% (19) and mixed infection (A+B) constituted 5.8% (6). A strong correlation between the presence of clinical signs/diarrhoea and assemblage B was observed. Moreover, a strong association was also observed between mixed infections (A+B) and clinical signs/diarrhoea as well. Whether this is due to the presence of assemblage B in the mixture could not be established at this stage. No correlation between age, gender, and geographic origin of the infected individual was noted. The remaining 21 (31%)

isolates were neither typed as assemblage A nor assemblage B raising the possibility of animal associated assemblages among the human population. Phylogenetic analysis showed two distinct clusters for *Giardia* among the participant possibly because of the diverse origin of the study samples. To our knowledge, this study is the first to examine the degree of infection with this important parasite in the country and to accurately determine the infection rate and its genotypic composition. It also raises questions about the possible zoonotic potential of the organism as well as the transmission dynamics between expatriates and natives of the UAE. Moreover, there is an urgent need for further studies within the native Emirati population before any conclusion can be made about expatriates transmitting *G. lamblia* to the local community. The study also provided information about *G. lamblia* genotypes in Sharjah, UAE and may therefore contribute to a better understanding of epidemiology of *Giardia* in the UAE.

Cryptosporidium is the third protozoan parasite that was investigated in the present thesis. The rate of infection with *Cryptosporidium* among expatriates residing in Sharjah, United Arab Emirates was determined. Real time PCR (qPCR) was used for detecting positive samples. A total of 134 stool samples were tested among whom 103 (76.8%) were males and 31 (23.1%) females. The age of the subjects varied between 5 and 58 years. Twenty six (19.4%) were found to be positive for *Cryptosporidium* by PCR. Although the infection rate was highest in Afghan nationals (33.3%) compared to the rest of the study population, no significant association was found between nationality and infection rate. Moreover, no association was observed between infection rate and gender ($\chi^2 = 2.439$; p=0.118) nor infection rate and age group ($\chi^2 = 1.219$; p=0.544).Thus, for the first time, the infection rate with *Cryptosporidium* spp. has been shown to be 19.4% in Sharjah, UAE. Further studies should address the genotypic distribution of the different species of *Cryptosporidium* and to screen and determine the species genotype from native Emirati nationals (Chapter 6).

The existence of significant genetic variation among *E. histolytica* isolates collected from a wide geographical range has already been demonstrated in numerous studies. However, the level of intra-species genetic variation in *E. histolytica* populations in South Africa remain unknown (Chapter 7). Therefore, using samples from South Africa, the genetic variation and

phylogenetic relationships among the detected *E. histolytica* isolates were studied by nucleotide sequencing. Sequence analysis revealed 13 unique genotypes from among 61 *E. histolytica* stool isolates. Interestingly, two strains were exactly identical, in that A is substituted by G in the same position, indicating that they are possibly from the same source. The other genotypes showed varying degree of polymorphism in the 439 bp PCR product of the 16S-like rRNA gene. Moreover, molecular phylogenetic results have demonstrated that the 16S-like rRNA gene of all *E. histolytica* isolates tested are closely clustered together and are possibly restricted to the study region and have shown how valuable the 16S-like rRNA gene is in epidemiological studies.

In conclusion, molecular methods enabled us to differentiate between *Entamoeba* species thus provide us with an infection rate and prevalence of *E. histolytica* in the UAE. Furthermore this methodological approach gave us basic knowledge about *Giardia* prevalence and genotypes in humans in the UAE. Moreover, detection of *Cryptosporidium* species infections in the UAE has provided initial data which could be used for further studies. Information generated from this work will be useful for understanding the transmission, source of infection, possible role of animals in spread of the studied parasites and clinical outcome of infection.