Targeted long-read sequencing for genetic diversity identification of the zoonotic *Cryptosporidium sp.*

Sonia Boughattas[1], Abdelrahman ElGamal[2], Dana Albatesh[1,3] Ismail Alshaikh[3], Nayla Al-Naema[3], Asmaa A. Althani[1], Fatiha M. Benslimane[1*]

[1]Biomedical Research Center, Qatar University; [2]College of Health Science, Qatar University; [3]ExxonMobil, Doha, Qatar; [4]Enivromental Science Center, Qatar University
[*]Correspondence: fatiha@qu.edu.qa

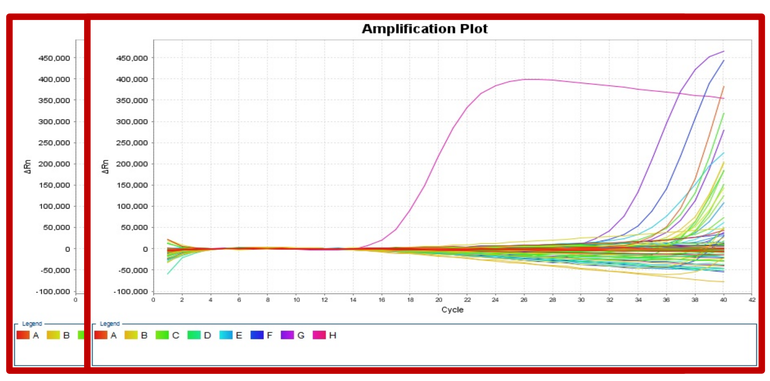
Background

Cryptosporidiosis is ranked sixth in the list of the most important food-borne parasites globally, and it is an important contributor to mortality within infants and the immunosuppressed subjects. The causing agent *Cryptosporidium sp.* is a widespread apicomplexan protozoan exhibiting long-lasting survival rates in water and soil causing high environmental contamination which could become a source of infections for humans and livestock. Work with this parasite is challenging with increasing number of identified species within the *Cryptosporidium* genus. Many of its genotypes are still poorly characterized with scare correlation data either within clinical symptoms or different host manifestations. For accurate genetic analysis we are investigating a method for generating full-length *Cryptosporidium sp.* SSU rRNA gene sequences using the GridION long-read sequencing platform from amplicons obtained using universal eukaryotic SSU rRNA gene primers.

Methods and Results

Sampling,

DNA extraction,



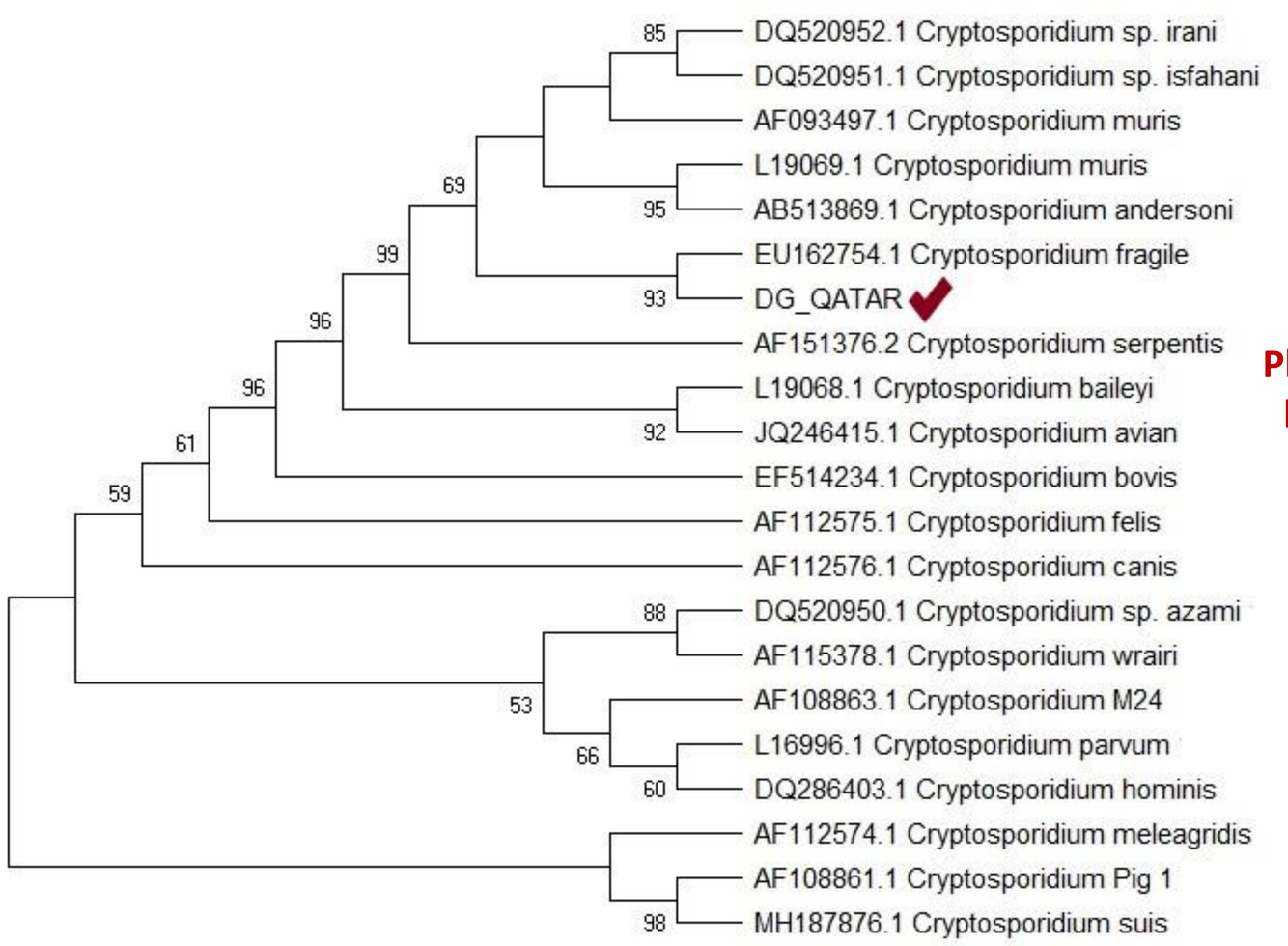
Pathogen screening

Nanopore library

- ✓ PCR amplification of full 18S gene (1800bp) using LongAmp Taq Polymerase (NEB) and the universal primers Af and Br
- ✓ Library preparation by SQK-LSK109 following the manufacturer's protocol for 1D amplicon/cDNA by Ligation (version: ACDE_9064_v109_revG_23May2018).
- ✓ Loading on R9.4 fow cells (FLO-MIN106) on the ONT GridIon platform.
- ✓ Sequencing to approximately 500,000 reads/sample.
- ✓ Basecalling performed using Guppy v4.4.1 with the high accuracy model and a minimum quality score cut off of 7.

Nanopore analysis

- ✓ Passed FASTQ reads were length filtered: 1700 and 2100 nucleotides.
- ✓ Reads were corrected, adapter trimmed using canu v2.1.
- ✓ Reads were clustered using the VSEARCH –cluster_fast command (vsearch v2.15.1) at a 98% identity threshold.
- ✓ Clusters were then polished with racon v1.4.20, clustered again, and polished once more with Medaka v1.4.3 using the model r103_min_high_g360.
- ✓ Consensus sequence was aligned with MAAFT and phylogenetic tree were generated.



Phylogenetic tree with the full-length sequence of 18S gene

Conclusion

With The current work provides the first full gene sequence evidence of the *Cryptosporidium* from Dugongs confirming thus that Dugong populations are exposed to anthropogenic. The ONT enabled the full-length sequence generation of the isolate when conventional subtyping was unsuccessful. This current approach will assist in accurate sequence descriptions in future studies of pathogens epidemiology, transmission routes and genetic diversity.

This work is part of project funded by Qatar petroleum under project number QUEX-BRC-QP-GH-18/19 and The Francophone University Agency under project number QUEX-BRC-AUF-22/23.