A novel target enrichment based approach for genomically characterising *Giardia duodenalis* directly from human and animal clinical samples

Rossella Panarese[a], Paul Capewell[a], Claire Alexander[b] and William Weir[a]

[a] School of Biodiversity, One Health and Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Garscube Estate, 464 Bearsden Road, Glasgow, G61 1QH, UK; [b] Scottish Microbiology Reference Laboratories (Glasgow), New Lister Building, 10-16 Alexandria Parade, Glasgow Royal Infirmary, Glasgow, G31 2ER, UK

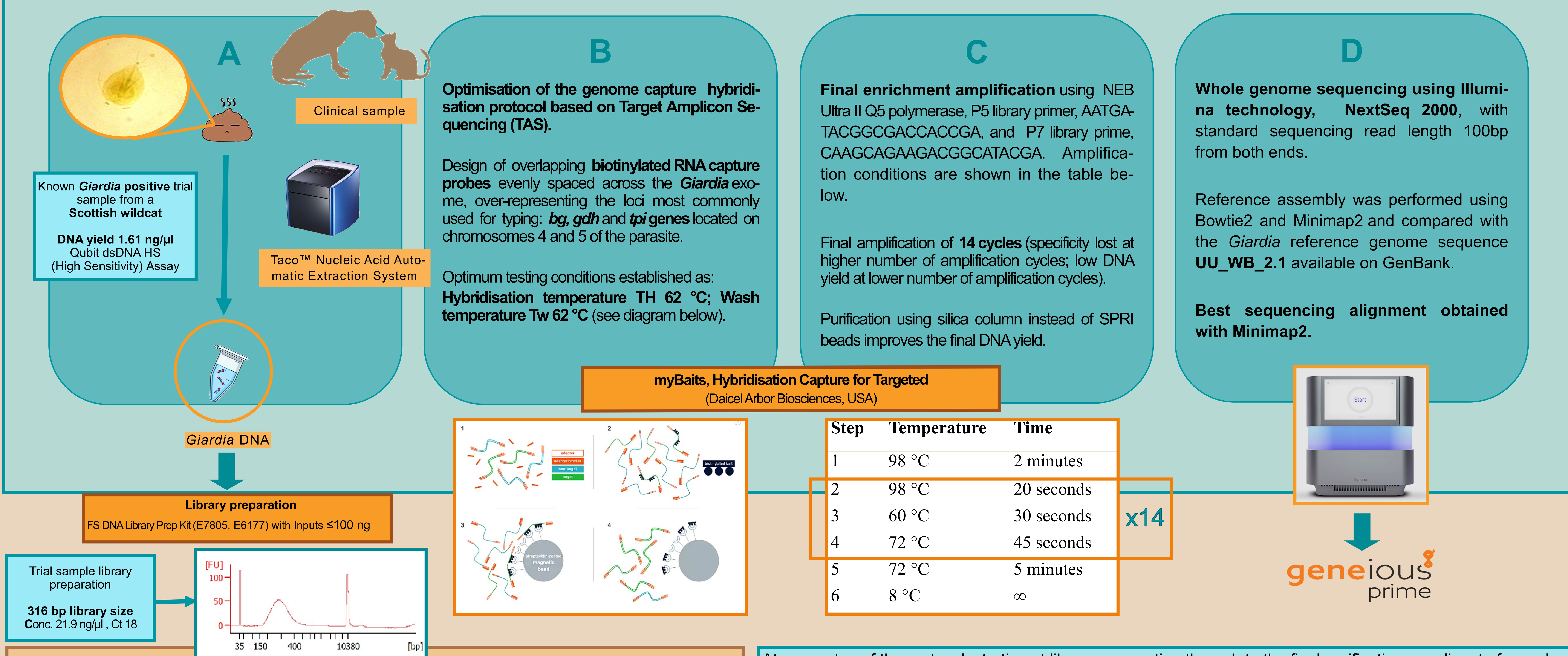
Rossella.Panarese@glasgow.ac.uk

Introduction

Giardia duodenalis (*Giardia*) is an intestinal protozoan parasite which causes disease in both animals and humans across the globe. It is becoming increasingly apparent that in developed countries there exist major knowledge gaps on *Giardia* ecology and epidemiology. For example, in Scotland, despite being considered an "exotic" disease, recent studies have suggested the presence of an endemic cycle of disease within the country which is not associated with travel-related infections[1]. However, endemic transmission routes are poorly understood and, in particular, the level of zoonotic risk is unknown. Current molecular markers for parasite genotyping are low-resolution and are only effective on a minority of field samples[2,3]. Moreover, only a low proportion of *Giardia* isolates can be adapted to culture, which presents a major obstacle to generating good-quality DNA preps for genomic sequencing[2]. Therefore, the present study aims: (i) to establish a laboratory workflow to allow sequencing of *Giardia* directly from faecal samples collected from human and animal cases in Scotland; (ii) to generate high-quality exome sequences to establish a robust minimum set of genotyping markers for analysing relationships between parasite isolates and (iii) to compare the genomes of endemic and foreign travel-associated isolates to begin to investigate transmission routes in Scotland.

Material and Methods

The laboratory workflow consists of **four stages**:



Step	Temperature	Time
1	98 °C	2 minutes
2	98 °C	20 seconds
3	60 °C	30 seconds
4	72 °C	45 seconds
5	72 °C	5 minutes
6	8 °C	∞

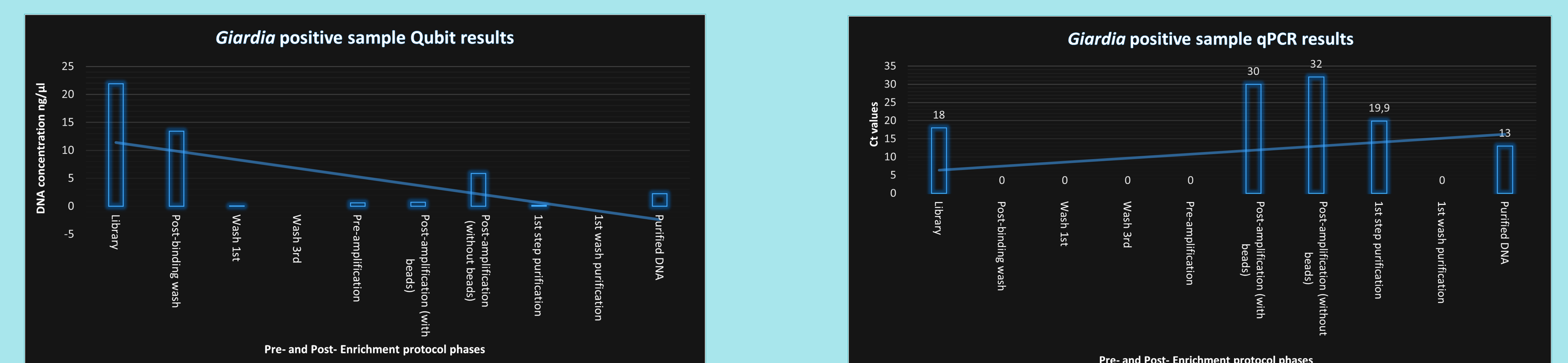
Results

A total of 61,740,610 paired-end reads, averaging 101 bp in length, were generated in the course of the sequencing run. 92,698 successfully mapped to the *Giardia* reference genome, clustering around the target 'bait' loci which were distributed throughout the parasite exome. This resulted in the creation of a 4.7% and 3.6% ref-seq coverage for chromosome 4 and 5, with a mean contig length of 200 and 265 bases for each chromosome, respectively. Those loci that were over-represented on the probe design had a statistically higher depth of coverage than the other regions of the genome and included the three key genes used for routine genotyping. All three loci could be successfully assembled, with read abundances expressed in terms of 'transcripts' per kilobase million (TPM) of 777,977 for *gdh*, 222,022 for *bg* and 1,000,000 for *tpi*. These gene sequences were compared to reference assemblage sequences and at each locus, the top-scoring match was found be 'assemblage F', the classical feline *Giardia* assemblage type.

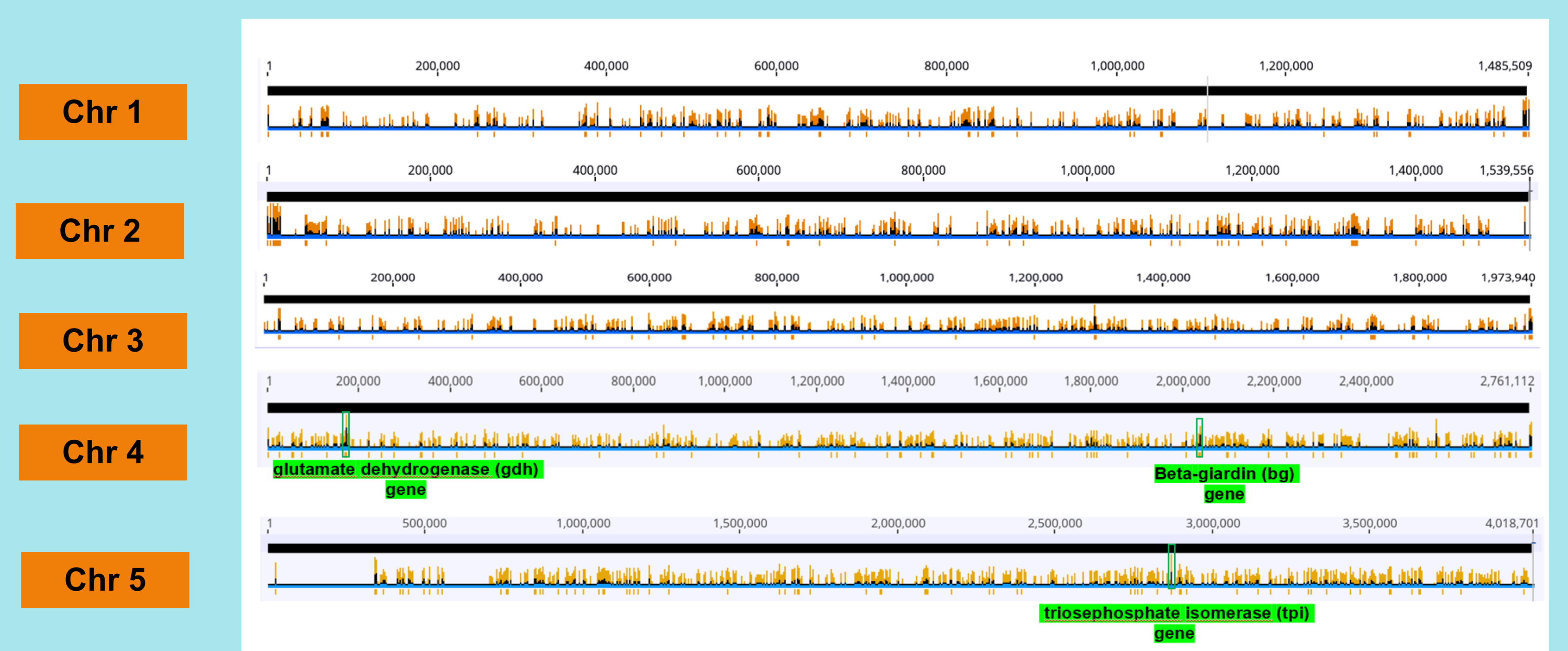
Conclusions

We have successfully applied this methodology to a *Giardia*-positive faecal sample from a Scottish wildcat, achieving a high level of parasite DNA enrichment and generating exome-wide genotyping data. Having demonstrated 'proof of principle' for this approach, we now intend to apply it to a number of human and animal samples. Ultimately, it is envisioned that genome capture methodology may provide a rapid and accurate typing method for this parasite which can be applied directly to clinical samples. This will assist in the typing of samples that do not readily PCR and those for which classical typing loci produce conflicting results. Such ultra-high resolution *Giardia* genotyping methodology will allow a better understanding of the epidemiology of the parasite in Scotland and further afield and has the potential to facilitate the elucidation of transmission routes and the tracing of disease outbreaks.

At every step of the protocol, starting at library preparation through to the final purification, an aliquot of sample material was stored and tested by Qubit and qPCR.



Schematic of the five *Giardia* chromosomes showing the distribution of mapped reads across the genome and location of genes encoding triosephosphate isomerase (*tpi*), beta-giardin (*bg*), glutamate dehydrogenase (*gdh*).



Percentage nucleotide identity between test sequence and reference *Giardia* sequences at the three loci used for assemblage typing: bg similarity (A); gdh similarity (B); tpi similarity (C).

A

	Bg gene_*Giardia* pos sample
Assemblage F (bg)	**99.1%**
Assemblage C (bg)	94.0%
Assemblage D (bg)	93.6%
Assemblage B (bg)	95.7%
Assemblage E (bg)	96.3%
Assemblage A (bg)	96.8%

B

	Gdh gene_*Giardia* pos sample
Assemblage F (gdh)	**100%**
Assemblage A (gdh)	93.8%
Assemblage D (gdh)	86.1%
Assemblage B (gdh)	89.3%
Assemblage C (gdh)	89.7%

C

	Tpi gene_*Giardia* pos sample
Assemblage F (tpi)	**98.3%**
Assemblage B (tpi)	79.0%
Assemblage C (tpi)	80.9%
Assemblage D (tpi)	76.3%
Assemblage A (tpi)	91.9%

References

1. Ferguson LC, Smith-Palmer A, Alexander CL. An update on the incidence of human giardiasis in Scotland, 2011–2018. *Parasites & vectors*, 2020, 13: 1-7;
2. Capewell P. et al. Molecular epidemiology of *Giardia* infections in the genomic era. *Trends in parasitology*, 2021, 37.2: 142-153;
3. Krumrie S., et al. Molecular characterisation of *Giardia duodenalis* from human and companion animal sources in the United Kingdom using an improved triosephosphate isomerase molecular marker. *Current Research in Parasitology & Vector-borne Diseases*, 2022, 2: 100105.