A Hybrid Whole Genome Sequencing Approach to Studying the Population Structure of Eimeria Parasites

Conor Noonan[1], *Justin Pachebat*[2], *Melanie Hay*[1], *Sarah Hill*[1], *Damer Blake*[1], *Dong Xia*[1]

[1]Department of Pathobiology and Population Sciences, Royal Veterinary College, Hawkshead Lane, North Mymms AL9 7TA, UK
[2]Department of Life Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA, UK

Background

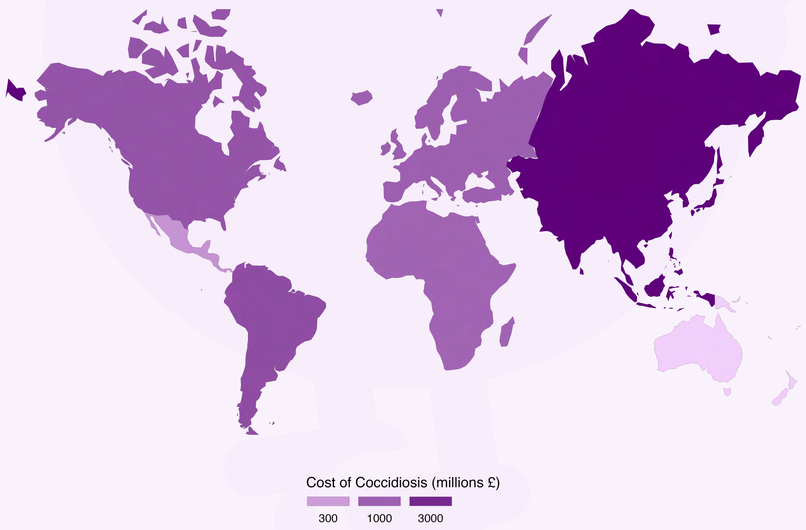


Figure 1. Extrapolated region-specific annual cost of coccidiosis in millions (£).

Eimeria are a genus of Apicomplexan parasites of veterinary clinical importance. Ingestion of this parasite leads to coccidiosis, an enteric disease whose clinical signs include haemorrhagic diarrhoea, diminished weight gain, and mortality in severe infection. Species which infect chickens are of economic importance as coccidiosis incurs costs upwards of £10.4 billion to annual global poultry production [1]. Despite their impact in the agriculture sector, little is known about the genetic diversity of these parasites, and how this variation contributes to the rising level of resistance to current treatment strategies. These knowledge gaps have arisen due to the repetitive structure of *Eimeria* genomes, which pose a computational challenge in *de novo* genome assembly, as well as the use of limited marker sets.

Objectives

1. Develop a Nanopore long read sequencing workflow for *Eimeria* parasites.
2. Improve current *Eimeria* reference genome assemblies.
3. Use newly-complete reference sequences for whole-genome analyses.

Methods

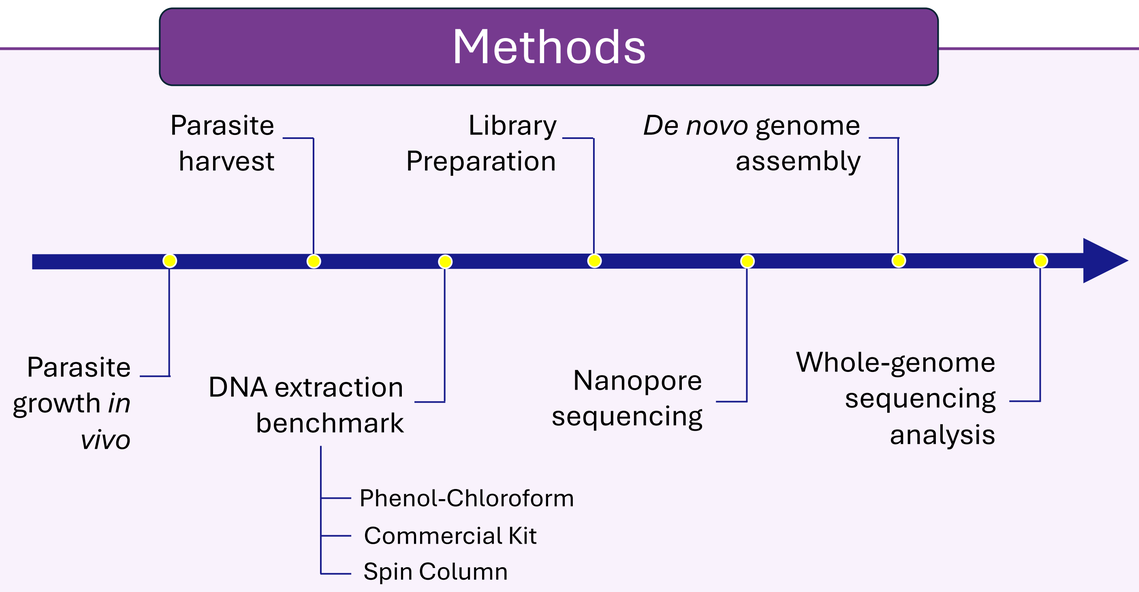


Figure 2. Workflow for long read sequencing of *Eimeria* parasites. Three different DNA extraction methods were compared to assess which was optimal for extracting low-fragmented DNA. Assembly pipeline was derived from Oresegun et al., 2022 [2].

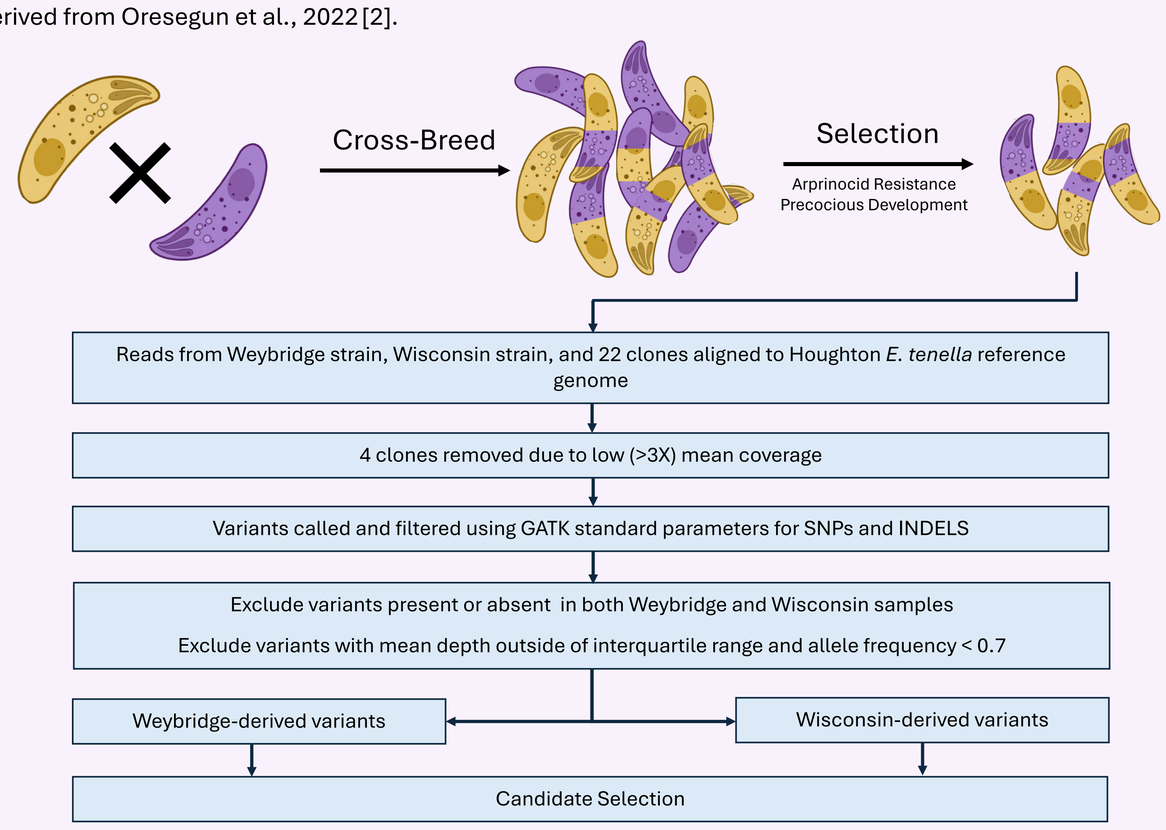


Figure 3. Whole genome analysis workflow of *Eimeria tenella* clonal cross-breeds using Illumina short read data and a complete reference genome to identify causative variants of drug resistance and precocious development [3].

Results

Nanopore Results

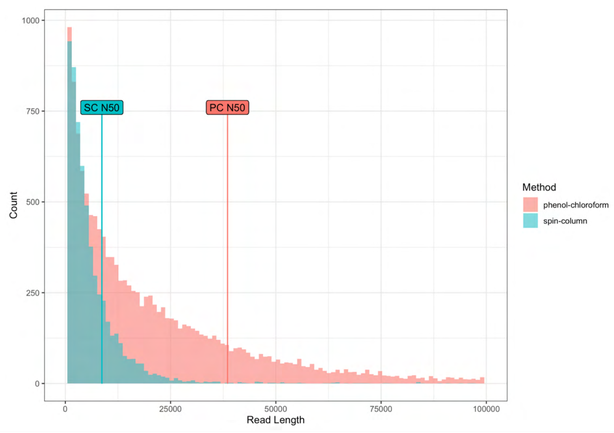
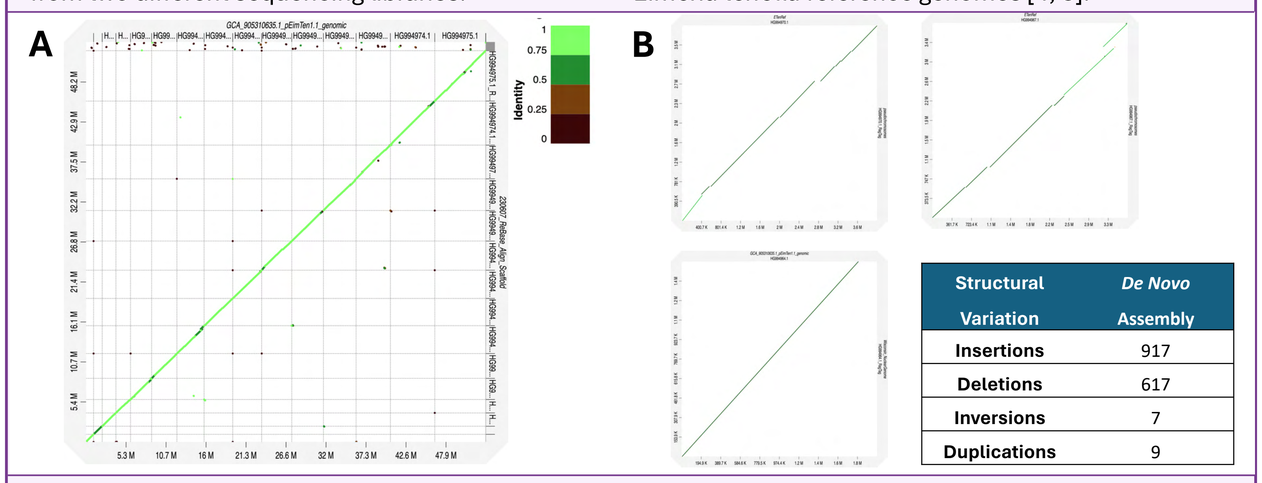


Figure 4. Comparison of read length distributions from two different sequencing libraries.

Statistic	*De Novo* Assembly	2013 Reference	2021 Reference
Assembly Length	53.59 MB	51.9	53.25
Number of Contigs	256	4664	17
N50	440 kb	200 kb	4 MB
Complete BUSCOs	92.8 %	93 %	98.8 %

Table 1. Comparison of *de novo* assembly to two *Eimeria tenella* reference genomes [4, 5].



Structural Variation	*De Novo* Assembly
Insertions	917
Deletions	617
Inversions	7
Duplications	9

Figure 5. **(A)** Structural comparison of *de novo* assembly and 2021 reference. **(B)** Overview of assembly errors and structural variants. **Table 2.** Structural variant summary obtained from Sniffles2 [6].

Whole Genome Analysis

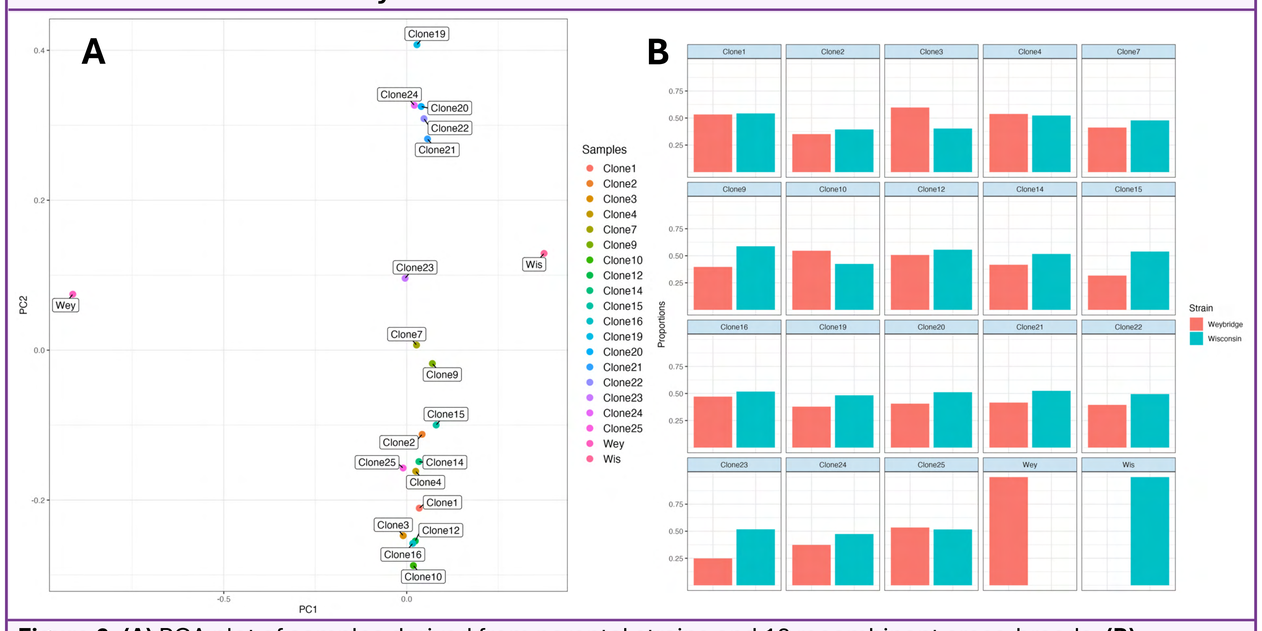


Figure 6. **(A)** PCA plot of samples derived from parental strains and 18 recombinant cross-breeds. **(B)** Proportional composition of parental-derived variants in each cross-breed.

Variant Type	Weybridge	Wisconsin
SNPs	2416	2117
INDELs	415	508
Splice Region Variants	20	35
Missense Variants	150	133
Upstream/Downstream Variants	718	1009
Intergenic Variants	1699	1139

Table 3. Summary of variants followingthe GATK and manual filtering pipeline outlined in Figure 3.

Conclusions

1. Long read sequencing is a cost-effective method of improving the contiguity of existing *Eimeria* genomes.
2. The choice of DNA extraction method has a significant impact on long read sequencing output and fragment length.
3. The availability of complete reference *Eimeria* genomes and the use of whole-genome sequencing revealed loci attributed to drug resistance and precocious development.

Future Directions

- ➤ Optimise sequencing workflow to improve DNA extraction yield, sequencing output, and sequencing quality.
- ➤ Perform sequencing on field samples.
- ➤ Investigate vandidate variants for resistance and precocious development.
- ➤ Integrate sequencing data into ML model for vaccine efficacy.

Acknowledgements

1. Blake DP, Knox J, Dehaeck B, Huntington B, Rathinam T, Ravipati V, et al. Re-calculating the cost of coccidiosis in chickens. Vet Res. 2020;51(1):115.
2. Oresegun DR, Thorpe P, Benavente ED, Campino S, Muh F, Moon RW, et al. De Novo Assembly of Plasmodium knowlesi Genomes From Clinical Samples Explains the Counterintuitive Intrachromosomal Organization of Variant SICAvar and kir Multiple Gene Family Members. Front Genet. 2022;13:855052.
3. Figure 2 Created with BioRender.com.
4. Reid AJ, Blake DP, Ansari HR, Billington K, Browne HP, Bryant J, et al. Genomic analysis of the causative agents of coccidiosis in domestic chickens. Genome Res. 2014;24(10):1676-85.
5. Aunin E, Bohme U, Blake D, Dove A, Smith M, Corton C, et al. The complete genome sequence of Eimeria tenella (Tyzzer 1929), a common gut parasite of chickens. Wellcome Open Res. 2021;6:225.
6. molka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. Detection of mosaic and population-level structural variants with Sniffles2. Nat Biotechnol. 2024.
7. Blake DP, Worthing K, Jenkins MC. Exploring Eimeria Genomes to Understand Population Biology: Recent Progress and Future Opportunities. Genes (Basel). 2020;11(9).

cnoonan22@rvc.ac.uk