De Novo Assembly and Comparative Analysis of Poultry-Infecting *Eimeria* Genomes Support Insights into Drug Resistance and Precocious Development

Conor Noonan[1], Justin Pachebat[2], Melanie Hay[1], Sarah Hill[1], Damer P. 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, Ceredigion, SY23 3DA, UK*

Eimeria are a genus of parasite that can infect all livestock, the ingestion of which can lead to the enteric disease coccidiosis. Of particular economic importance are species which infect chickens, incurring costs upwards of £10.4 billion annually to global production. Despite their impact on animal health and production costs, little is known about the genetic diversity of these parasites and how this variation contributes to the rising level of resistance to existing control strategies. These knowledge gaps exist primarily due to the incomplete nature of the reference genomes, the highly fragmented nature of which is a direct consequence of the technical limitations of short read sequencing technologies and the inherent repeat-rich content of *Eimeria* genomes.

To address these shortcomings, we developed an experimental workflow using a hybrid Illumina and Nanopore sequencing pipeline to improve upon the quality of *Eimeria* reference genomes by generating long reads capable of spanning repeat-rich genomic regions. Acting as scaffolds, these large sequences improve the *de novo* assembly of *Eimeria* genomes by properly orienting shorter sequences and bridging gaps between them. In doing so, more information can be retained, and more accurate genome models can be generated.

We improved the completeness and compositional quality of six existing reference genomes for poultry-infecting *Eimeria* species, reducing them from thousands of contigs to near-chromosomal scale, and producing novel high-quality reference sequences for the newly characterised species *E. zaria* and *E. lata*. A subset of contigs within each assembly exhibit telomeric repeats, with those at both ends forming pseudochromosomal sequences. Comparative genomic analysis of these assemblies revealed evidence of extensive genomic rearrangements within the genus, including a reciprocal translocation event leading to the divergence between the closely related *E. lata* and *E. maxima* species. Analysis of orthogroups across the genus highlighted the lineage-specific expansion of functional gene families involved in processes such as invasion, as well as novel families of uncharacterised function. Furthermore, lineage-specific expansions of transposon families were identified, reflecting patterns of genome expansion within *E. mitis*, the largest known *Eimeria* genome to date.

The availability of improved genome models has enabled us to re-examine historical whole genome sequencing data to derive genetic loci for traits relevant to parasitism and infection severity. Candidate genes underpinning key clinical phenotypes like attenuated development and arprinocid resistance have been highlighted using linkage group selection and low coverage whole genome sequencing.

Ultimately, the generation of high-quality, complete reference sequences has proven integral to resolving the genomic diversity and regional population structures of these veterinary pathogens, and the genetic determinants of clinically significant traits.