Update TRENDS in Parasitology Vol.20 No.5 May 2004

Genome Analysis

The *Eimeria* genome projects: a sequence of events

Martin W. Shirley[1], Al Ivens[2], Arthur Gruber[3], Alda M.B.N. Madeira[3], Kiew-Lian Wan[4], Paul H. Dear[5] and Fiona M. Tomley[1]

[1]Institute for Animal Health, Compton Laboratory, Compton, Nr Newbury, Berkshire, RG20 7NN, UK
[2]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[3]Universidade de Sao Paulo, Faculdade de Medicina Veterinária e Zootecnia, Departamento de Patologia, Av. Prof. Dr. Orlando Marques de Paiva 87, São Paulo, 05508-000, Brazil
[4]Centre for Gene Analysis and Technology, School of BioSciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor DE, Malaysia
[5]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

An international consortium is the driving force behind several new genome-related projects, mainly focused on *Eimeria tenella*, the cause of avian, caecal coccidiosis. The largest project is a whole genome shotgun project, which is at 8.3-fold coverage, and is complemented by complete sequencing of the two smallest *E. tenella* chromosomes and the provision of a physical framework through HAPPY mapping. The derivation of expressed sequence tags from various developmental stages of *E. tenella* and other avian *Eimeria* species is under way and a comprehensive annotation of the entire genome will begin in 2004.

Coccidial parasites belonging to the genus *Eimeria* are a major cause of acute disease and ill thrift in poultry, cattle and sheep. The intestinal disease, coccidiosis, caused by *Eimeria* is especially significant to the poultry industry, and all of the ~39 billion chickens reared each year worldwide must be protected against *Eimeria* infection.

Why sequence *Eimeria*?

Chickens are host to seven species of *Eimeria* and each species is responsible for a different form of coccidiosis. The three most economically significant species are *Eimeria tenella, Eimeria acervulina* and *Eimeria maxima*. The cost of clinical and sub-clinical disease, together with the cost of drugs and vaccines, is ~£40 million per annum in the UK, equivalent to 4.5% of the revenue from sales of live birds [1]. The annual cost worldwide is probably in excess of £2 billion.

Disease control relies on prophylactic chemotherapy and, to a lesser degree, vaccination. More than 235 tonnes of anticoccidials were used in the UK during 2000 (Veterinary Medicines Directorate; http://www.vmd.gov.uk), yet sub-clinical coccidiosis remains commonplace because drug efficacy is compromised by drug-resistant parasites [2]. The portfolio of effective drugs is sharply declining, particularly within Europe where a recent EU ban has removed five chemical anticoccidials, and the future use of the ionophores looks increasingly uncertain.

Vaccines, especially live-attenuated vaccines, are used increasingly with great success [3,4], but are relatively expensive because their manufacture relies on production of oocysts in chickens. The sustainable future of vaccination will require new vaccines that can be produced at lower cost on a large scale. Recombinant vaccines based on a portfolio of protective antigens would be ideal. However, drugs are still favored in many countries and development of new anticoccidials is a priority for some parts of the animal health sector. Whether drugs, vaccines or a combination prevail, it is clear that long-term control of coccidiosis will require identification of new targets within *Eimeria* parasites. This imperative underpins the need for genome sequencing, and work has begun with *E. tenella* because of its importance to agriculture and the research community. One significant benefit of the work will be the increased opportunities for comparative genomics and the *Eimeria* data will be of use to scientists with an interest in other apicomplexans, such as *Plasmodium* spp., *Neospora* and *Toxoplasma*.

Progress on the *Eimeria* genome projects

With funding from the Biotechnology and Biological Sciences Research Council (BBSRC; http://www.bbsrc.ac.uk/) in the UK, shotgun sequencing of the genome of the *E. tenella* Houghton (H) strain began at the Sanger Institute (http://www.sanger.ac.uk) in April 2002. By October 2002, >835 000 reads, with an average read length of 522 bp, were completed (http://www.sanger.ac.uk/projects/E_tenella). These have been assembled into ~49.1 Mb of unique sequence with an average base coverage of ~8.3-fold. At the outset, the genome of *E. tenella* was estimated at 55–60 Mb, as judged by pulsed-field gel electrophoresis (PFGE) analyses of the molecular karyotype. The data are presently assembled into 12 416 contigs, but this number will fall as assembly improves and the data are tidied up.

The *E. tenella* genome is proving to be highly distinctive with respect to the types and numbers of repetitive sequences it contains. Repeats of the trinucleotide GCA/TGC were first identified several years ago [5,6] and their ubiquity throughout the *E. tenella* genome has been confirmed. Analysis of random contigs invariably reveals the presence of these triplets, within coding and non-coding sequences, found typically in arrays of up to ~20 perfect triplets interspersed or flanked by degenerate

Corresponding author: Martin W. Shirley (shirley@bbsrc.ac.uk).

triplets. The biological significance of the triplet repeats is not currently known. Another common repetitive sequence is the heptamer TTTAGGG, a telomeric-like sequence found most frequently in non-coding regions in small arrays of up to six tandem repeats. The genome sequence also reveals almost perfect, longer stretches (up to 70 tandem repeats) of this heptamer, but there is no evidence that places this sequence within the definitive telomeric regions in *Eimeria* spp. [7].

The identification of genes with an assessment of their function is a crucial activity and a start to this work is under way. Analyses of the *E. tenella* genome are being integrated in Artemis [8], a DNA sequence viewer and annotation tool that allows visualization of sequence features, and the results of various sequence analyses, within the context of the sequence and its six-frame translation. For some genomes, for example, *Leishmania major*, relatively simple tools such as codon usage have proved to be extremely useful for gene prediction [9]. However, preliminary analyses of *E. tenella* contigs show that codon usage and other simple methods have only limited utility for this parasite. Gene predictions are therefore being undertaken using more complex software (e.g. GeneFinder) following the assembly of a comprehensive set of full length *E. tenella* genes on which to 'train' gene-finding algorithms. To facilitate this, ~28 500 publicly available *E. tenella* ESTs have been assembled into ~5100 different clusters (including singlets) to be mapped back onto the genome. Partial gene sequences predicted to contain more than one intron have been chosen for 5′ and 3′ rapid amplification of complementary DNA ends (RACE) extension of transcript sequences. To ensure that genome annotation is accurate and comprehensive, funding has been obtained to support a dedicated curator who will manually inspect the annotation of each protein. All annotated data will be made available to the research community via GeneDB (http://www.genedb.org), as has been the case for *Leishmania*, *Plasmodium* and other organisms. Additional annotation and feedback from the research community is welcomed.

The sequencing focus on *E. tenella* is underpinned by a genetic linkage map and the association of chromosomes 1 (~1 Mb) and 2 (~1.2 Mb) with resistance to an anticoccidial drug and precocious (accelerated) endogenous development, respectively [10]. HAPPY maps [11–13] of these chromosomes are currently being derived and the data will be incorporated with those from the Universiti Kebangsaan Malaysia to determine a complete sequence for the two smallest chromosomes (HAPPY mapping is an *in vitro* approach for defining the order and spacing of DNA markers directly on genomic DNA).

At the Universiti Kebangsaan Malaysia, the sequencing effort on chromosomes 1 and 2 of *E. tenella* H is being done at the Interim Laboratory of the National Institute for Genomics and Molecular Biology, with funding from the Ministry of Science, Technology and the Environment, Malaysia. The chromosomes were chosen because of their linkage with the phenotypic traits of interest and a random shotgun approach is being undertaken using small-insert libraries prepared from chromosome-specific DNAs separated by PFGE. The data comprise ~30 000 reads, and preliminary observations indicate that the two chromosomes have markedly different organization, largely due to the disparity in the distribution of novel repetitive elements.

At the Universidade de Sao Paulo, Brazil, 10 000 ESTs from unsporulated oocysts, sporozoites and second-generation merozoites of *E. tenella* H have been derived using the open reading frame expressed sequences tags (ORESTES) approach [14,15]. ORESTES are produced by low-stringency PCR amplification of complementary DNA (cDNA) transcripts before cloning. One outcome of the ORESTES approach is a bias towards regions in the middle of genes and the discovery of some transcripts that would be missed by traditional EST approaches because of low abundance in the original cells. The group has also derived 10 000 ESTs from sporozoites of *E. acervulina* H and is currently working towards the derivation of 10 000 ESTs from sporozoites of *E. maxima* H. This cDNA sequencing effort has generated >2000 EST clusters, each for *E. tenella* and *E. acervulina* (http://www.lbm.fmvz.usp.br/eimeria/). The sequences of the mitochondrial and plastid genomes have also been obtained through a combination of *de novo* sequencing and bioinformatics (A. Gruber, unpublished).

Where next for the *Eimeria* genome consortium?

An *ad hoc Eimeria* genome consortium first convened in Dublin in 2000, then met again during the VIIIth International Coccidiosis Conference in Australia in 2001*. All participants agreed to share knowledge and to support independent actions to promote *Eimeria* sequencing, and all of the projects on *E. tenella* established during the past two years make use of template DNA from the H strain.

With joint BBSRC/Wellcome Trust funding, a more formal meeting of the *Eimeria* genome consortium was held in May 2003, when the consortium met with other interested parties (e.g. apicomplexan biologists, funders, commercial and academic sectors of the veterinary industry)†. The consortium now has a series of short-, medium- and long-term objectives, and intends to meet formally at least once a year. The immediate challenge is to overcome the difficulties that the repetitive nature of the genome poses for contig assembly and to generate a large set of full length, multiple-intron-containing genes for accurate gene prediction. After gene prediction, it will be vital that genome annotation is carried out as fully and accurately as possible. Another important resource is the large-scale public domain EST sequencing project on *E. tenella*, undertaken jointly by Merck Research laboratories, USA (http://www.merck.com) and Washington University, USA (http://genome.wustl.edu/est/index.php?eimeria=1).

Perspective

In the longer term, the *Eimeria* genome consortium aims to seek funding for sequencing of *E. acervulina* and

* The COST820 Vaccines Against Animal Coccidioses meeting was held 15–18th June 2000, in Dublin, UK, and the VIIIth International Coccidiosis Conference was held 7–13th July 2001, in Cairns, Australia.

† The *Eimeria* Genome Consortium meeting was held 15–16th May 2003, in Wantage, UK.

Table 1. Current interests and activities in *Eimeria* genomics worldwide[a]

Country and current interests in *Eimeria* genomics	Associated website
Brazil	
Eimeria acervulina, *Eimeria maxima* and *Eimeria tenella* ESTs; ORESTES; extrachromosomal genomes; bioinformatics	http://www.lbm.fmvz.usp.br/
China	
Whole genome sequencing of *E. maxima*	–[b]
Malaysia	
Sequencing of chromosomes 1 and 2 of *E. tenella*	http://cgat.ukm.my/genomicslab/
UK	
Eimeria tenella whole genome shotgun sequencing and ESTs;	http://www.sanger.ac.uk/projects/e_tenella/ and
Eimeria genome annotation; bioinformatics; proteomics of apical organelles; genetic linkage maps of *E. tenella* and *E. maxima*	http://www.iah.bbsrc.ac.uk/eimeria/index.html
Eimeria HAPPY maps	http://www.mrc-lmb.cam.ac.uk/happy/happy-home-page.html
USA	
Eimeria acervulina whole genome shotgun sequencing and ESTs	http://www.anri.barc.usda.gov/pbel/index.asp

[a]This list is not exhaustive, but provides a snapshot of the major projects on *Eimeria* genomics that are now ongoing. Abbreviations: EST, expressed sequence tag; ORESTES open reading frame expressed sequences tags.
[b]Jianping Cai, Guangdong Academy of Agricultural Sciences, Guangdong, China.

E. maxima, two species that are also of great economic significance but which differ from *E. tenella* in many important aspects of their interactions with the chicken. Interest in these species has led to further expansion of the consortium (see Table 1). Detailed analysis of the genomes of *E. tenella*, perhaps with those of the other species, will be invaluable for developing an understanding of the biology and biochemistry of *Eimeria* parasites, for comparative analysis with other members of the Apicomplexa, and for guiding the selection of novel, effective targets for drug and vaccine design.

Acknowledgements
The British Council is thanked for the help provided through the auspices of the Committee for International Co-operation in Higher Education 444 programme between the Universiti Kebangsaan Malaysia and the Institute for Animal Health. The BBSRC is thanked for the support and funding to derive a genome sequence for the H strain of *Eimeria tenella*.

References

1 Williams, R.B. (1999) A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. *Int. J. Parasitol.* 29, 1209–1229
2 Chapman, H.D. (1997) Biochemical, genetic and applied aspects of drug resistance in *Eimeria* parasites of the fowl. *Avian Pathol.* 26, 221–244
3 Chapman, H.D. (2000) Practical use of vaccines for the control of coccidiosis in chickens. *Worlds Poult. Sci. J.* 56, 7–20
4 Shirley, M.W. *et al.* (1995) A live attenuated vaccine for the control of avian coccidiosis: trials in broiler breeders and replacement layer flocks in the United Kingdom. *Vet. Rec.* 137, 453–457
5 Jenkins, M.C. (1988) A cDNA encoding a merozoite surface protein of the protozoan *Eimeria acervulina* contains tandem-repeated sequences. *Nucleic Acids Res.* 16, 9863
6 Shirley, M.W. (1994) The genome of *Eimeria tenella*: further studies on its molecular organisation. *Parasitol. Res.* 80, 366–373
7 Shirley, M.W. (2000) The genome of *Eimeria* spp., with special reference to *Eimeria tenella*-a coccidium from the chicken. *Int. J. Parasitol.* 30, 485–493
8 Rutherford, K. *et al.* (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945
9 Aggarwal, G. *et al.* (2003) Importing statistical measures into Artemis enhances gene identification in the *Leishmania* genome project. *Bioinformatics* 4, 23
10 Shirley, M.W. and Harvey, D.A. (2000) A genetic linkage map of the apicomplexan protozoan parasite, *Eimeria tenella*. *Genome Res.* 10, 1587–1593
11 Dear, P.H. and Cook, P.R. (1989) Happy mapping: a proposal for linkage mapping the human genome. *Nucleic Acids Res.* 17, 6795–6807
12 Dear, P.H. and Cook, P.R. (1993) Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Res.* 21, 13–20
13 Piper, M.B. *et al.* (1998) A HAPPY map of *Cryptosporidium parvum*. *Genome Res.* 8, 1299–1307
14 de Souza, S.J. *et al.* (2000) Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12690–12693
15 Camargo, A.A. *et al.* (2001) The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12103–12108

1471-4922/$ - see front matter
doi:10.1016/j.pt.2004.02.005

Research Focus

Pregnancy-associated malaria – on the brink?

Alister G. Craig

Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

Regarded by many as the best example in falciparum malaria of an association between a specific parasite-adhesive phenotype and disease, pregnancy-associated malaria represents one of the more immediate hopes for control of malaria disease via vaccination. But, are we sure that we have identified the right candidate antigens and do we know how to measure the impact of such an intervention?

Corresponding author: Alister G. Craig (agcraig@liv.ac.uk).