Published in final edited form as:

Int J Parasitol. 2009 January ; 39(2): 145–151. doi:10.1016/j.ijpara.2008.10.002.

New Eukaryotic Systematics: A Phylogenetic Perspective of Developmental Gene Expression in the Apicomplexa

Mathieu Gissot¹, Kami Kim¹, Dick Schaap², and James W. Ajioka^{3,*}

¹Departments of Medicine (Infectious Diseases) and Microbiology & Immunology, Albert Einstein College of Medicine, Bronx, NY 10461, USA Tel: +01 718 430 2611 Fax: +01 718 430 8968
Mathieu Gissot: mgissot@aecom.yu.edu Kami Kim: kkim@aecom.yu.edu

²Parasitology R&D, Intervet International bv. P.O. Box 31, 5830AA Boxmeer, The Netherlands
Tel: +31 485 587 373 Fax: +31 485 585 313 dick.schaap@sp.intervet.com

³Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK

Abstract

The phylum Apicomplexa consists of obligate intracellular protistan parasites, some of which are responsible for global disease causing serious morbidity and mortality in humans and animals. Understanding the mechanisms of gene expression that drive the cellular changes required to complete their life cycles will be critical in combating infection and disease. *Plasmodium* spp. and *Toxoplasma gondii* have served as good models for growth and development in the Apicomplexa. Elucidating developmental gene expression relies on comparisons to known mechanisms and their DNA, RNA and protein components. Transcriptional profiling across asexual development suggests a model where a cascade of gene expression results in a “just in time” production process that makes products only when needed. Some mechanisms that control transcription such as chromatin/histone modification are highly conserved in the phylum compared to the traditional model organisms, yeast, worms, flies and mammals. Studies exploiting this phenomenon show great potential for both investigating the effects of chromatin structure on developmental gene expression, and helping to identify genes that are expressed in stage-specific manner. Transcription factors and their cognate cis-acting binding sites have been difficult to identify. This may be because the DNA binding motifs that have evolved to act as transcription factors in the Apicomplexa, e.g. the AP2 family, may be more like plants than the traditional model organisms. A new eukaryotic phylogenetic model comprised of six supergroups divides the traditional model organisms, plants and the Apicomplexa into separate supergroups. This phylogenetic model helps explain why basic functions such as transcriptional regulation appear to be a composite of mechanisms in the Apicomplexa compared to what is known from other eukaryotes.

Keywords

Chromaveolata; Opisthokonta; Archaeplastida; chromatin; histone; transcription factor; cis-acting element; AP2; siRNA; plasmodium; toxoplasma

*Corresponding author Tel: +44 1223 333 923 Fax: +44 1223 333 346 ja131@cam.ac.uk.

1. Introduction

The phylum Apicomplexa is composed exclusively of parasitic species with complex life cycles that have evolved highly specialized infection, replication and transmission cycles within and between a wide variety of hosts (see for example, Levine et al., 1980). Nevertheless, as obligate intracellular pathogens, apicomplexan growth and development are intimately associated with this specific environmental niche, so the underlying molecular mechanisms required for these processes are likely to be shared across species. This review will focus primarily on *Toxoplasma gondii* and *Plasmodium* spp. as they represent the most well studied species within the phylum. Sexual development in both *Plasmodium* spp. and *T. gondii* have been characterized at a cellular level but the gene expression studies in this stage have been limited (see below; Alano, 2007; Ajioka & Sibley, 2007). Asexual growth and development are the most well characterized aspects of their life cycles, in part because good *in vitro* models exist for these processes. In *Plasmodium* spp. the asexual intraerythrocytic cycle, starting with the ring-form to growing trophozoites through schizogony has been extensively characterized and is the main developmental model. For *T. gondii*, the transition from the rapidly dividing tachyzoite seen in acute infection to the nearly dormant bradyzoite found in tissue cysts associated with the chronic infection, is the principle developmental model.

For elucidating the genetic mechanisms that drive development, comparative analysis at virtually every taxonomic level is crucial, requiring an understanding how apicomplexans are related to one another and how the phylum is related to other well-studied model organisms. For many reasons, there hasn't been a serious attempt to reconstruct a phylogeny for species within the Apicomplexa since the first designation and descriptions of the phylum. However, at a higher level, recent studies using a combination of morphological and molecular data have re-organized eukaryotes into six “super-groups” where the hierarchical system is without formal ranking and can more easily accommodate secondary losses, gains and lateral transfers (see Fig. 1; see for example, Simpson & Roger, 2004; Adl et al., 2005). The most significant change in this reorganization is based on the identification of several monophyletic lineages within the protists and that traditional Kingdoms such as Metazoa, Fungi and Plantae are derived from these lineages. This new eukaryotic systematics differs from the old “ladder-like” model where most protists were placed as early diverging organisms with a progression toward a “crown radiation” of plants, animals and fungi (Dacks et al., 2008). This new systematic view, where the supergroups are equally modern/ancient, should guide computational and experimental work and therefore greatly assist the formation of hypotheses about how the scope and use of particular mechanisms of gene expression drive growth and development in the Apicomplexa. Although this re-organization is the appropriate theoretical framework for comparative studies, this review will continue to use traditional taxonomic ranks to simplify discussions.

2. Control of gene expression: a phylogenetic perspective

The paradigms for the control of gene expression have been largely established from studies of traditional model organisms that represent a fairly narrow slice of life. Although protists,

including the phylum Apicomplexa, probably account for the majority of eukaryotic life (Moon-van der Staay et al., 2001), our biological knowledge of eukaryotes is disproportionately skewed towards a small number of model organisms, mostly multicellular (Anantharaman et al., 2007). Given the new systematic view (see Fig. 1), the bias is compounded by the fact that the traditional model organisms of yeast and metazoans (e.g. *Saccharomyces cerevisiae*, *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*) are placed together in the supergroup Opisthokonta. Moreover, Plants and algae (e.g. *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*) belong to the Archaeplastida supergroup and the Apicomplexa in yet another supergroup, the Chromalveolata. Hence, the comparison of basic functional processes like gene regulation between apicomplexans and the major eukaryotic model organisms is likely to be complex due to lineage-specific adaptations and secondary gains/losses (Anantharaman et al., 2007). For example, in the Apicomplexa, the secondary endosymbiosis with a photosynthetic algae (Waller & McFadden, 2005) raises the possibility that our lack of finding “conventional” transcription factors and cognate regulatory motifs (i.e. those found in yeast or metazoans) may be because the majority of these elements in the Apicomplexa have a different evolutionary ancestry and have a greater relationship with photosynthetic organisms in this regard. Some epigenetic modifications such as cytosine methylation also appear to be missing in the phylum (Gissot et al., 2008). Conversely, histone modifications that specify chromatin structure and mark active vs. repressed transcriptional states appear to be much more “conventional”, in that antibodies raised against mammalian methylation and acetylation modifications cross react with apicomplexan chromatin (see for example Gissot et al., 2007; see below). Together, the computational and experimental data suggest that elucidation of gene expression in apicomplexans will require an evolutionary framework that allows for the mosaic nature and composition of these species.

3. Developmental transcriptional profiling

In *Plasmodium* spp. and *T. gondii*, developmental transcriptional profiling by microarray analysis and serial analysis of gene expression (SAGE) show that specific sets of genes are coordinately transcribed where particular sets of expressed genes define developmental stages (Bozdech et al. 2003, Le Roch et al., 2003; Radke et al., 2005; Behnke et al. 2008). In a comparison of major *P. falciparum* life stages, cluster analysis of transcriptional patterns showed that each life stage is characterized by 100–200 genes that are highly expressed, but vary in their functional classification (Le Roch et al., 2003). The sporozoite cluster is dominated by characterized sporozoite-specific genes whereas the gametocyte cluster is composed of 88% hypotheticals. This difference highlights both the intensity of study of the respective life stages and the potential complexity of gametogenesis. A statistical analysis of the chromosomal organization of stage-specific genes identified a greater than expected number of genes to be adjacent to one another, suggesting some level of chromatin-dependent regulation. In contrast to these snapshots of developmental stages, the developmental process itself was interrogated by a highly detailed time series analysis of transcription through the *P. falciparum* (3D7 strain) intraerythrocytic developmental cycle (Bozdech et al., 2003). This study revealed a remarkably tight regulatory process that generates a “continuous cascade” of gene expression. At least 60% of the genome is

transcribed during this asexual cycle beginning with the ring-form and culminating with mature merozoites. In general, the induction of any given gene happens only once per cycle, so the proposed transcriptional model from these data is akin to a “just-in-time” production process where products are made only when required. In this study, co-regulated genes were dispersed across the genome and adjacent genes were only rarely seen to follow the same pattern, arguing for a predominately trans-acting mechanism of co-regulation. Transcriptional analysis of the *T. gondii* asexual tachyzoite-to-bradyzoite transition presents a generally similar picture to *P. falciparum* despite a less synchronous population of parasites. A screen for *T. gondii* tachyzoites that fail to become bradyzoites under bradyzoite culture conditions produced several mutants (Singh et al., 2002). In a comparison of the microarray transcriptional profiles for these mutants under bradyzoite culture conditions, each mutant expressed a different subset of genes to form a nested hierarchy, raising the possibility that the tachyzoite-to-bradyzoite transition is controlled by a cascade of gene expression. Complementing this genetic approach, transcriptional profiling of a sporozoite initiated asexual cycle with Serial Analysis of Gene Expression (SAGE) showed that *T. gondii* stage-specific mRNAs accounted for 18% of the SAGE tags (Radke et al., 2005). The comparison of strain types with different capacities to develop into bradyzoites showed that the virulent strain (RH) that does not readily form bradyzoites, has a unique expression profile compared to the two less virulent strains (ME49 and VEG) that easily form bradyzoites. The main difference occurred just prior to the onset of bradyzoite development where ME49 and VEG commenced expression of bradyzoite genes but RH appeared to be arrested in a tachyzoite-like expression pattern. This “priming” of developmental expression is reminiscent of the “just-in-time” transcriptional model and may share underlying mechanisms. Like the *P. falciparum* study, the developmentally co-regulated genes are dispersed across the genome, again supporting a trans-acting regulatory model. A recent follow-on study of the *T. gondii* tachyzoite-to-bradyzoite developmental transition provided more direct evidence for a trans-acting model by defining cis-acting elements upstream from bradyzoite-specific genes that are sufficient for determining stage-specific expression (Behnke et al., 2008).

4. Transcription factors and cis-acting elements

Historically transcription factors and their cognate cis-acting elements/binding sites in apicomplexans have proved to be elusive targets, possibly because they are less abundant and/or they do not share homology with the bulk of characterized proteins and sequences in the database. Analysis of the *P. falciparum* var gene promoters show protein binding but the mechanisms controlling expression appear to involve epigenetic processes that silence transcription by chromatin modification (see below). Several *T. gondii* studies have identified putative promoter regions with constitutive activity by deletion analysis and some identified motifs within these regions, but none defined specific sequence elements that are both necessary and sufficient for transcriptional control (Soldati and Boothroyd, 1995; Mercier et al., 1996; Roos et al., 1997; Nakaar et al., 1998; Matrajt et al., 2004). However a combination of recent computational and experimental evidence investigations in *T. gondii* and *Plasmodium* spp. provides evidence for a sequence specific DNA binding protein and its

binding site that may confer transcriptional regulation, so this aspect of transcriptional control is probably not as barren as previously thought (see below)

For growth, two different approaches have been taken, utilizing the *P. falciparum* and *T. gondii* annotated genome sequences. Transcription factor families are sometimes highly diverse, as exemplified by the Myb family. Although Myb was originally described as a viral proto-oncogene associated with avian myeloblastosis, the taxonomic distribution of this group of transcription factors is wide, notably represented as a large family with a variety of functions in plants (see for example Allan et al., 2008). In *P. falciparum*, a search of the *Plasmodium* database with a consensus Myb protein sequence revealed a putative protein, PfMyb1, with high similarity match (Boschet et al., 2004). Cloning and analysis of *pfmyb1* from the *P. falciparum* strain 3D7 showed peak transcription at the trophozoite stage and a western blot of 3D7 trophozoite nuclear extracts probed with antibodies against PfMyb1 synthetic peptides, produced a band whose molecular mass was slightly larger than the control recombinant PfMyb1, but close enough to assume they represent the same proteins. Binding specificity was demonstrated by Electrophoretic Mobility Shift Assays (EMSA) with nuclear extracts of 3D7 promoter probes producing a shift relative to the unrelated control probes. A knockdown experiment using *pfmyb1* double stranded RNA tested the function of PfMyb1 and produced a 40% reduction in growth and subsequent mortality during the transition to the schizont stage (Gissot et al., 2005). These observations correlated with the down regulation of cell-cycle genes and a Chromatin Immunoprecipitation (ChIP) experiment demonstrated that PfMyb1 binds to several promoters suggesting regulation of the corresponding genes. Growth/protein synthesis in *T. gondii* are thought to be transcriptionally regulated where an investigation using ribosomal protein (RP) gene upstream sequences provided insight into the coordination of this process (Van Poppel et al., 2006). The 79 RP genes in *T. gondii* are abundantly expressed in tachyzoites and are relatively short as they code for small proteins, so establishing full length coding sequences with 5' UTR sequence can be achieved by aligning representative sequences from the Expressed Tag Sequence data. These assembled RP gene sequences mapped onto the genome sequence allowed a precise definition of the upstream putative promoter sequences associated with them. Motif searching for putative promoter elements uncovered two sequences found in 95% of the RPs: Toxoplasma Ribosomal Protein (TRP)-1 (consensus: TCGGCTTATATTCGG) and TRP-2 ([T/C]GCATGC[G/A]). The distance from the presumptive transcriptional start site ranged from 10 to 330 bp. A motif search of upstream regions defined by histone modification peaks on *T. gondii* chromosome Ib (see below) showed that about 30% of the expressed genes contained TRP-2 (J. Ajioka and K. Kelly, unpublished data). This observation was further validated by a motif search of upstream sequences from a large set of known transcriptional start sites that also showed a similar frequency of TRP-2. This frequency of 30% is similar to the frequency of TATA boxes in other eukaryotes suggesting it may be used for modulating a basal level of transcription. The function of TRP-2 may be elucidated by classifying the genes that are associated with it aside from RPs. Mutagenesis studies of TRP-2 in several putative promoters demonstrate the sequence is essential for expression (K. Kim and M. Gissot, unpublished data). Identification of the protein(s) that bind the sequence is currently underway (K. Kim and M. Gissot, unpublished data).

Although computational approaches to finding transcription factors and their cis-acting/binding sites have identified several putative factors and motifs (see below), experimental studies in the context of development have been limited to date. The *T. gondii* tachyzoite-to-bradyzoite transition is a model that has been exploited in this regard where several genes associated with each of the life stages have been identified. *In vitro* cultivation of *T. gondii* under stress conditions such as alkaline pH, nitric oxide or respiration inhibitors drives tachyzoites to become bradyzoite-like (for review, see Tomavo & Weiss 2007). An early study using deletion analysis of the 1.2 kb putative promoter region upstream from the gene encoding the small bradyzoite-specific heat shock protein BAG1, defined a ~900 bp region that dramatically reduced reporter expression under bradyzoite culture conditions (Bohne et al., 1997). While the specific sequence element(s) were not defined, this study showed that reporter constructs could be successfully used for a phenotypic assay in the tachyzoite-to-bradyzoite conversion. A similar study of *BAG1* and *Tg-hsp70* upstream regions, reduced the size of the putative promoters to 100–200 bp and were able to design ~30 bp double stranded oligonucleotides with GAA repeats (seen in other eukaryotic heat shock elements) for EMSA assays (Ma et al., 2004). These DNAs showed significant band shifts compared to the competition controls with the addition of *T. gondii* protein lysates, demonstrating that these sequence elements could specifically bind protein. Search of the *T. gondii* genome database for heat shock factors did not produce any high similarity matches indicating that these proteins probably do not account for the gel shift. Another tachyzoite-to-bradyzoite differentially expressed gene system involves the two *T. gondii* plant-like enolase isoforms, *ENO2* mRNA found in tachyzoites and *ENO1* mRNA found in bradyzoites (Dzierszinski et al., 2001). Upstream regions from both isoforms were mapped by deletion to identify the putative promoter regions (Kibe et al., 2005). In contrast to transcriptional induction, two regions in the putative *ENO2* promoter region conferred transcriptional repression in bradyzoites, consistent with this being a tachyzoite specific isoform. In addition to identifying the AGGG consensus sequence, tachyzoite and bradyzoite proteins from nuclear extracts bound previously defined stress response elements and heat-shock-like elements. Albeit indirect, these data show that DNA binding proteins may be involved in promoting transcription of *ENO1* and *ENO2*. In order to more systematically investigate the tachyzoite-to-bradyzoite transition, several known bradyzoite putative promoter regions were mapped to identify cis-acting elements (Behnke et al., 2008). Fine mapping of the putative promoter regions for the heat shock protein gene *BAG1* (1195 bp) and a novel bradyzoite-specific *B-NTPase* (1495 bp) defined minimal cis-acting elements of 20 and 30 bp respectively. Analysis of these sequence elements provided evidence that they could convert a constitutive promoter to one that is upregulated by bradyzoite culture conditions. Increased copy number of the 20 or 30 bp elements led to a concomitant increase in expression. EMSA analysis of the *B-NTPase* and *BAG1* minimal elements with nuclear extracts prepared from the type III CTG strain under bradyzoite culture conditions showed dominant shift bands that could be effectively competed with excess unlabeled wild-type DNA, arguing for sequence specific protein binding.

While sequence specific transcription factors driving development remain unknown, other factors that lack sequence specificity appear to be critical for development in *Plasmodium*. Members of the high mobility group box (HMGB) family, including transcription factors

that are known to be non-sequence specific, are highly conserved across *Plasmodium* species (Briquet et al., 2006). The two *P. falciparum* HMGB proteins HMGB1 and HMGB2 appear to be critical in a developmentally specific upregulation of transcription by activities that may include both RNA polymerase II recruitment and chromatin modification (Briquet et al., 2006). HMGB1 is expressed during the erythrocytic asexual cycle and HMGB2 is upregulated in the gametocyte stage compared to the schizont stage. Given the highly conserved nature of these genes, analysis of HMGB2 in the rodent malaria model *P. yoelli* showed that it is a key regulator of sexual stage gene expression (Gissot et al., 2008). Deletion of the gene encoding the *P. yoelli* HMGB2 (*pyhmg2*) downregulated ~30 genes and the vast majority of these genes showed peak transcriptional expression in the gametocyte stage, including genes known to be critical in oocyst formation on the mosquito gut wall. Interestingly, the primary effect of *pyhmg2* is manifested as a significant decrease in oocyst formation. The simplest explanation for this is translational repression of mRNAs produced in gametogenesis until oocyst formation.

Further analyses of these data suggest that HMGB2 is central to a network of putative transcriptional activators, including plant-like AP2 domain proteins (LaCount et al., 2005; Gissot et al. 2008). This observation is a tantalizing connection to the computational analysis of DNA binding motifs in the Apicomplexa that revealed a family of plant-like AP2 domain proteins, proposed as putative transcription factors (ApiAP2; Balaji et al. 2005). In order to assess the DNA binding capacity of these proteins, two *P. falciparum* ApiAP2 members were tested using a protein binding microarray consisting of 44,000 60-mer oligonucleotides representing all possible 10-mer sequences (Da Silva et al. 2008). The proteins PF14_0633 and PFF0200c showed specific binding to the palindromic TGCATGCA and GTGCAC consensus sequences. The binding of PF14_0633 to TGCATGCA corroborates and extends the earlier findings from the analysis of this element in *T. gondii*, *Cryptosporidium parvum* and *Eimeria tenella* upstream sequences and its potential promoter activity (M. Gissot and K. Kim, unpublished observations). This observation also argues for a basic function common across the Apicomplexa for this ApiAP2/cis-acting-element pair, given the range of genes the motif is associated with (Dear et al., 2003; Van Poppel et al., 2006; Ling et al., 2007; J. Ajioka and K. Kelly, unpublished observations).

Major taxonomic groups appear to have lineage specific expansions (LSE) of protein motifs that are the primary component for the essential functions of signal transduction and gene regulation (Anantharaman et al., 2007). The homeodomain family is the major LSE for transcriptional regulators in animals compared to the AP2 family being the major LSE in plants and the Apicomplexa. Further identification and characterization of the proteins that bind these sequences is the next critical step in understanding how they interact with the sequence elements and the basal transcription machinery to promote transcription in a developmentally regulated fashion.

5. Transcription and chromatin modification

Unlike transcription factors, epigenetic control of transcription by chromatin modification appears to be highly conserved between the Apicomplexa and members of the Opisthokonta including yeast, worms, drosophila and mammals (for excellent recent reviews see Sullivan

and Hakimi, 2006, Hakimi & Deistch, 2007). This conservation will greatly assist in deciphering the “histone code” in the Apicomplexa. Inspection of the *T. gondii* and *P. falciparum* genomes shows that most of the enzymes required for chromatin modification via ubiquitination, acetylation, methylation, sumoylation, ADP-ribosylation and phosphorylation appear to have good homologues in these parasites (Sullivan and Hakimi, 2006; Hakimi & Deistch, 2007). The conservation of histone tail sequences and common methylation and acetylation modifications in *T. gondii* and *P. falciparum*, is sufficient for cross-reaction to antibodies raised against these modifications in mammalian chromatin (see below). This fortuitous observation has afforded the opportunity to investigate how certain modifications correlate with transcriptional activation and repression.

P. falciparum PfEMP1 proteins encoded by the *var* gene family are highly expressed in erythrocyte stage parasites as means of immune evasion may also have a role in other developmental stages such as gametocytes (Sharp et al., 2006). Changes in gene expression during *in vitro* gametocyte development result in transcription of the type C (chromosome-internal vs. types A and B which are subtelomeric) subset of *var* genes regardless of the *var* gene expression by the progenitor parasite. Although the control for this developmental switch is unclear, analysis of the *var* gene family in the asexual intraerythrocytic cycle show that the mechanism of allelic exclusion and limitation of expression to a single *var* gene, is controlled by chromatin modification and transcriptional repression (for recent review see Kyes et al., 2007). Hypoacetylation by the histone deacetylase PfSIR2, a chromatin state normally associated with transcriptional silencing and the H3K9me3 (histone 3, lysine 9 trimethylation) correlate with transcriptional repression (Chookajorn et al., 2007; Frietas-Junior et al., 2005). Moreover, the upstream type C *var* promoter (*upsC*) region nucleates and propagates stably inherited silencing and is sufficient to “infiltrate” a transgene into the silencing program where activation of the transgene results in the silencing of endogenous *var* gene expression (Voss et al., 2006). These results suggest that as yet unidentified cis-acting elements in *upsC* promoter region are able to propagate an epigenetic signal by the recruitment of silencing machinery.

Using a biochemical approach, analysis of chromatin remodeling complexes involved in chromatin silencing and activation are also proposed to be important in the *T. gondii* tachyzoite-to-bradyzoite transition (Sasouk et al., 2005). In tachyzoites, the TgCRC co-repressor complex containing TgHDAC3 is associated with hypoacetylated chromatin and marks silenced bradyzoite-specific genes such as *BSR4*. In contrast, histone acetyltransferase TgGCN5 and the arginine transmethylase complex TgCARM1, target histone H3 [K18] and H3 [R17] respectively at actively transcribed tachyzoite genes. Further characterization of histone methylation in *Plasmodium spp.* and *T. gondii* demonstrate that the evolutionarily conserved SET8 enzymes target histone H4 [K20] repetitive sequences and telomeric heterochromatin (Sautel et al., 2008). Unlike the human SET8, the apicomplexan SET8 enzymes can di- and tri-methylate their target, showing that heterochromatin is not restricted to metazoans and may have more diverse functions, possibly important in different developmental states.

A ChIP-microarray (ChIP-Chip) based study of a contiguous 1% of the *T. gondii* genome produced a large-scale view of histone modification and transcriptional profiling over part of

chromosome Ib (Gissot et al 2007). In tachyzoites, mapping acetylation (H4ac, H3K9ac) and trimethylation (H3K4me3) modifications produced peaks that marked actively transcribed genes. The H3R17me2 modification marked only a subset of actively transcribed genes suggesting promoter-specific differences in associated chromatin remodeling complexes (Gissot et al., 2007). As predicted, the histone modification peaks associated with gene activation were absent from known oocyst-specific tandemly arrayed kinase genes and bradyzoite-specific *BSR4* genes. The peaks also define a relatively small upstream region on the order of 1000–1500 bp so motif finding and promoter analysis will be statistically tractable. Depending on the probes' location on the genome, a dip in the H3K9ac/H4ac modified histone peak can be identified. This region is presumed to represent a nucleosome free region that provides accessibility for transcription factors and the RNA polymerase II (Kim and Gissot, unpublished). Notably the H3K4me3 peak was closer to the 5' end of the transcribed genes compared to H3K9ac/H4ac, providing polarity to promoter regions defined by these histone modifications. These observations will greatly improve the annotation of genes that are computationally identified as the gene finder programs often miss the 5' exon.

With regard to growth, histone acetylation plays an important role in yeast. In particular ESA1, a specific yeast histone acetylase, is essential and binds to RP-promoters, requiring a specific binding site (Reid et al., 2000). The functional homologue in *Toxoplasma* is Tg-MYST-A, and this histone acetylase also seems essential and cannot be knocked-out or stably overexpressed (Sullivan and Hakimi, 2006). Tg-MYST-A expression is far more abundant in rapidly replicating tachyzoites relative to encysted bradyzoites, and Tg-MYST-A is present in both cytoplasm and nucleus, suggesting a control on its nuclear presence (Smith et al., 2005). Whether and how Tg-MYST-A controls tachyzoite growth is currently under investigation, but may likely involve control on *T. gondii* RP-gene transcription.

Collectively, these data suggest that whole genome coverage of histone modifications over a tachyzoite-to-bradyzoite developmental time series would identify the key genes involved in the transition and provide insight into how the timing and type of chromatin structure influences developmental gene expression and the cellular identity.

6. RNA-based controlling mechanisms controlling gene expression

Over the past decade, several studies have described structural non-coding RNA expression with unknown function in *Plasmodium* spp. and *T. gondii* (see for example Gunasekera et al., 2004; Chakrabarti et al., 2007). Computational analysis of their genomes for various functional proteins that process or utilize these non-coding RNAs for gene expression have generated experimentally testable hypotheses about their function.

Small RNAs that are derived from double stranded RNA (dsRNA) include ~20–22 nt microRNAs (miRNAs) and ~21–25 nt small interfering RNAs (siRNAs) that function by inhibiting translation, have a variety of functions in eukaryotic organisms including growth and development (for a recent review see Farazi et al., 2008). A search for putative orthologues of key processing enzymes such as Dicer and Argonaute suggests that the usage of these RNAs will vary between species (Ullu et al., 2004). Orthologues for these enzymes

do not appear to be present in the *P. falciparum* genome (Aravind et al., 2003), whereas a good Argonaute orthologue, TgAgo, is present in the *T. gondii* genome (Al Riyahi et al., 2006). Consistent with these observations, at least two experimental searches in *Plasmodium* spp. for miRNAs were negative (Rathgen et al., 2006; Xue et al., 2008), whereas in *T. gondii*, there is some experimental evidence for Argonaute-dependent siRNA inhibition of gene expression of glycolytic enzymes, albeit with variable results (Anavaranich et al., 2006). Curiously, a systematic computational search for conserved structured (double stranded) RNAs across *Plasmodium* spp. uncovered ~600 novel candidates and produced experimental evidence for the expression of 33, some of which showed stage-specific expression (Mourirer et al., 2007). Further genetic and biochemical investigations will be required to resolve this apparent conundrum.

Like metazoans, a subset of *Plasmodium* spp. mRNAs are subject to translational repression (TR) during the development of fertilized female gametes (Mair et al., 2006; Braks et al., 2008). The repression is mediated by DOZI (a DDX6 group helicase) with other protein complexes. In *P. berghei*, the TR is a U-rich element of about 47 nt encoded in the 5' or 3' UTRs of these mRNAs. This mechanism of controlling translation in female gametes closely parallels the cis- and trans-acting factors used metazoan female germline cells, and may represent an ancient post-transcriptional mechanism for controlling gene expression in this cell type.

7. Conclusions

Recent work on how gene expression drives growth and development in the Apicomplexa reveals some important trends. Although the jury is still out as to which aspects of the transcriptional and translational control mechanisms will predominate, there is evidence that most of the major regulatory processes are found at some level. Crucially, the ability to elucidate particular mechanisms has been biased by what is known in the traditional eukaryotic genetic model systems such as yeast, worms, flies and mammals. By chance all of these species are related as members of the Opisthokonta, one of the six supergroups. The concordance within this group will be higher than between groups so the gene expression paradigms may or may not fit very well with species outside the Opisthokonta. This is borne out by the high conservation of chromatin/histone modifying system compared to the paucity of transcription factor homologues. For the latter, identification of the plant-like AP2 family in the Apicomplexa suggests that looking towards the Archaeplastida may be more fruitful. The evolutionary developmental gene expression in the Apicomplexa has resulted in a curious mix of conserved mechanisms when compared to other eukaryotic groups and it will be of great interest to see how they work in concert to generate the diversity of forms seen in the life cycle of these parasites.

Acknowledgments

We would like to thank the Wellcome Trust (JWA), NIAID R01 AI 60496 (KK and MG) and Intervet (DS) for supporting this work.

References

- Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* 2005; 52:399–451. [PubMed: 16248873]
- Ajioka, JW.; Sibley, LD. Development and application of classical genetics in *Toxoplasma gondii*. In: Weiss, LM.; Kim, K., editors. *Toxoplasma gondii*, The model Apicomplexan: perspectives and methods. Academic Press; London, UK: 2007.
- Alano P. Plasmodium falciparum gametocytes: still many secrets of a hidden life. *Mol. Microbiol.* 2007; 66:291–302. [PubMed: 17784927]
- Allan AC, Hellens RP, Laing WA. MYB transcription factors that colour our fruit. *Trends Plant. Sci.* 2008; 13:99–102. [PubMed: 18280199]
- Al Riyahi A, Al-Anouti F, Al-Rayes M, Ananvoranich S. Single argonaute protein from *Toxoplasma gondii* is involved in the double-stranded RNA induced gene silencing. *Int. J. Parasitol.* 2006; 36:1003–1014. [PubMed: 16753169]
- Anantharaman V, Iyer LM, Aravind L. Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. *Annu. Rev. Microbiol.* 2007; 61:453–475. [PubMed: 17506670]
- Ananvoranich S, Al Rayes M, Al Riyahi A, Wang X. RNA silencing of glycolysis pathway in *Toxoplasma gondii*. *J. Eukaryot. Microbiol.* 2006; 53(Suppl 1):S162–163. [PubMed: 17169045]
- Aravind L, Iyer LM, Wellems TE, Miller LH. Plasmodium biology: genomic gleanings. *Cell.* 2003; 115:771–785. [PubMed: 14697197]
- Balaji S, Babu MM, Iyer LM, Aravind L. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2- integrase DNA binding domains. *Nucleic Acids Res.* 2005; 33:3994–4006. [PubMed: 16040597]
- Bankier AT, Spriggs HF, Fartmann B, Konfortov BA, Madera M, Vogel C, Teichmann SA, Ivens A, Dear PH. Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. *Genome Res.* 2003; 13:1787–1799. [PubMed: 12869580]
- Behnke MS, Radke JB, Smith AT, Sullivan WJ Jr, White MW. The transcription of bradyzoite genes in *Toxoplasma gondii* is controlled by autonomous promoter elements. *Mol. Microbiol.* 2008; 68:1502–1518. [PubMed: 18433450]
- Bohne W, Wirsing A, Gross U. Bradyzoite-specific gene expression in *Toxoplasma gondii* requires minimal genomic elements. *Mol. Biochem. Parasitol.* 1997; 85:89–98. [PubMed: 9108551]
- Boschet C, Gissot M, Briquet S, Hamid Z, Claudel-Renard C, Vaquero C. Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 Plasmodium falciparum clones. *Mol. Biochem. Parasitol.* 2004; 138:159–163. [PubMed: 15500927]
- Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol.* Oct.2003 1:E5. [PubMed: 12929205]
- Braks JA, Mair GR, Franke-Fayard B, Janse CJ, Waters AP. A conserved U- rich RNA region implicated in regulation of translation in Plasmodium female gametocytes. *Nucleic Acids Res.* 2008; 36:1176–1186. [PubMed: 18158300]
- Briquet S, Bosch C, Gissot M, Tissandier E, Sevilla E, Franetich JF, Thierry I, Hamid Z, Bourgouin C, Vaquero C. High-mobility-group box nuclear factors of Plasmodium falciparum. *Eukaryot. Cell.* 2006; 5:672–682. [PubMed: 16607015]
- Chakrabarti K, Pearson M, Grate L, Sterne-Weiler T, Deans J, Donohue JP, Ares M Jr. Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA.* 2007; 13:1923–1939. [PubMed: 17901154]
- Chookajorn T, Dzikiowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW. Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci USA.* 2007; 104:899–902. [PubMed: 17209011]

- Dacks JB, Walker G, Field MC. Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int.* 2008; 57:97–104. [PubMed: 18180199]
- De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, Bulyk ML, Llinás M. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA.* 2008; 105:8393–8398. [PubMed: 18541913]
- Dzierszinski F, Mortuaire M, Dendouga N, Popescu O, Tomavo S. Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii*. *J. Mol. Biol.* 2001; 309:1017–1027. [PubMed: 11399076]
- Farazi TA, Juranek SA, Tuschl T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development.* 2008; 135:1201–1214. [PubMed: 18287206]
- Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mancio-Silva L, Leal-Silvestre RJ, Gontijo AM, Shorte S, Scherf A. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell.* 2005; 121:25–36. [PubMed: 15820676]
- Gissot M, Ting LM, Daly TM, Bergman LW, Sinnis P, Kim K. High mobility group protein HMGB2 is a critical regulator of plasmodium oocyst development. *J. Biol. Chem.* 2008; 283:17030–17038. [PubMed: 18400754]
- Gissot M, Choi SW, Thompson RF, Grealley JM, Kim K. *Toxoplasma gondii* and *Cryptosporidium parvum* lack detectable DNA cytosine methylation. *Eukaryot Cell.* 2008; 7:537–540. [PubMed: 18178772]
- Gissot M, Kelly KA, Ajioka JW, Grealley JM, Kim K. Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog.* Jun.2007 3:e77. [PubMed: 17559302]
- Gissot M, Briquet S, Refour P, Boschet C, Vaquero C. PfMyb1, a *Plasmodium falciparum* transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J. Mol. Biol.* 2005; 346:29–42. [PubMed: 15663925]
- Gunasekera AM, Patankar S, Schug J, Eisen G, Kissinger J, Roos D, Wirth DF. Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol.* 2004; 136:35–42. [PubMed: 15138065]
- Hakimi MA, Deitsch KW. Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation. *Curr. Opin. Microbiol.* 2007; 10:357–362.
- Kibe MK, Coppin A, Dendouga N, Oria G, Meurice E, Mortuaire M, Madec E, Tomavo S. Transcriptional regulation of two stage-specifically expressed genes in the protozoan parasite *Toxoplasma gondii*. *Nucleic Acids Res.* 2005; 33:1722–1736. [PubMed: 15784612]
- Kyes SA, Kraemer SM, Smith JD. Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. *Eukaryot Cell.* 2007; 6:1511–1520. [PubMed: 17644655]
- LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C, Fields S, Hughes RE. A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature.* 2005; 438:103–107. [PubMed: 16267556]
- Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, Winzeler EA. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science.* 2003; 301:1503–1508. [PubMed: 12893887]
- Levine ND, Corliss JO, Cox FE, Deroux G, Grain J, Honigberg BM, Leedale GF, Loeblich AR 3rd, Lom J, Lynn D, Merinfeld EG, Page FC, Poljansky G, Sprague V, Vavra J, Wallace FG. A newly revised classification of the protozoa. *J. Protozool.* 1980; 27:37–58. [PubMed: 6989987]
- Ling KH, Rajandream MA, Rivailler P, Ivens A, Yap SJ, Madeira AM, Mungall K, Billington K, Yee WY, Bankier AT, Carroll F, Durham AM, Peters N, Loo SS, Isa MN, Novaes J, Quail M, Rosli R, Nor Shamsudin M, Sobreira TJ, Tivey AR, Wai SF, White S, Wu X, Kerhornou A, Blake D, Mohamed R, Shirley M, Gruber A, Berriman M, Tomley F, Dear PH, Wan KL. Sequencing and analysis of chromosome 1 of *Eimeria tenella* reveals a unique segmental organization. *Genome Res.* 2007; 17:311–319. [PubMed: 17284678]

- Ma YF, Zhang Y, Kim K, Weiss LM. Identification and characterisation of a regulatory region in the *Toxoplasma gondii* hsp70 genomic locus. *Int. J. Parasitol.* 2004; 34:333–346. [PubMed: 15003494]
- Mair GR, Braks JA, Garver LS, Wiegant JC, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP. Regulation of sexual development of *Plasmodium* by translational repression. *Science.* 2006; 313:667–669. [PubMed: 16888139]
- Matrajt M, Platt CD, Sagar AD, Lindsay A, Moulton C, Roos DS. Transcript initiation, polyadenylation, and functional promoter mapping for the dihydrofolate reductase-thymidylate synthase gene of *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 2004; 137:229–238. [PubMed: 15383293]
- Mercier C, Lefebvre-Van Hende S, Garber GE, Lecordier L, Capron A, Cesbron- Delauw MF. Common cis-acting elements critical for the expression of several genes of *Toxoplasma gondii*. *Mol Microbiol.* 1996; 21:421–428. [PubMed: 8858595]
- Moon-van der Staay SY, De Wachter R, Vaulot D. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature.* 2001; 409:607–610. [PubMed: 11214317]
- Mourier T, Carret C, Kyes S, Christodoulou Z, Gardner PP, Jeffares DC, Pinches R, Barrell B, Berriman M, Griffiths-Jones S, Ivens A, Newbold C, Pain A. Genome- wide discovery and verification of novel structured RNAs in *Plasmodium falciparum*. *Genome Res.* 2008; 18:281–292. [PubMed: 18096748]
- Nakaar V, Bermudes D, Peck KR, Joiner KA. Upstream elements required for expression of nucleoside triphosphate hydrolase genes of *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 1998; 92:229–239. [PubMed: 9657328]
- Radke JR, Behnke MS, Mackey AJ, Radke JB, Roos DS, White MW. The transcriptome of *Toxoplasma gondii*. *BMC Biol.* 2005; 3:26. [PubMed: 16324218]
- Rathjen T, Nicol C, McConkey G, Dalmay T. Analysis of short RNAs in the malaria parasite and its red blood cell host. *FEBS Lett.* 2006; 580:5185–5188. [PubMed: 16963026]
- Reid JL, Iyer VR, Brown PO, Struhl K. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Mol. Cell.* 2000; 6:1297–1307. [PubMed: 11163204]
- Roos DS, Sullivan WJ, Striepen B, Bohne W, Donald RG. Tagging genes and trapping promoters in *Toxoplasma gondii* by insertional mutagenesis. *Methods.* 1997; 13:112–122. [PubMed: 9405195]
- Saksouk N, Bhatti MM, Kieffer S, Smith AT, Musset K, Garin J, Sullivan WJ Jr, Cesbron-Delauw MF, Hakimi MA. Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma gondii*. *Mol. Cell. Biol.* 2005; 25:10301–10314. [PubMed: 16287846]
- Sautel CF, Cannella D, Bastien O, Kieffer S, Aldebert D, Garin J, Tardieux I, Belrhali H, Hakimi MA. SET8-mediated methylations of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes. *Mol Cell Biol.* 2007; 27:5711–5724. [PubMed: 17562855]
- Sharp S, Lavstsen T, Fivelman QL, Saeed M, McRobert L, Templeton TJ, Jensen AT, Baker DA, Theander TG, Sutherland CJ. Programmed transcription of the var gene family, but not of stevor, in *Plasmodium falciparum* gametocytes. *Eukaryot Cell.* 2006; 5:1206–1214. [PubMed: 16896206]
- Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. *Curr Biol.* 2004; 14:R693–R696. [PubMed: 15341755]
- Singh U, Brewer JL, Boothroyd JC. Genetic analysis of tachyzoite to bradyzoite differentiation mutants in *Toxoplasma gondii* reveals a hierarchy of gene induction. *Mol. Microbiol.* 2002; 44:721–733. [PubMed: 11994153]
- Smith AT, Tucker-Samaras SD, Fairlamb AH, Sullivan WJ Jr. MYST family histone acetyltransferases in the protozoan parasite *Toxoplasma gondii*. *Eukaryot. Cell.* 2005; 4:2057–2065. [PubMed: 16339723]
- Sullivan WJ Jr, Hakimi MA. Histone mediated gene activation in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 2006; 148:109–116. [PubMed: 16644030]
- Soldati D, Boothroyd JC. A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Mol. Cell. Biol.* 1995; 15:87–93. [PubMed: 7799972]

- Tomavo, S.; Weiss, LM. Toxoplasma gene regulation and bradyzoite development. In: Ajioka, JW.; Soldati, D., editors. Toxoplasma, molecular and cellular biology. Horizon Bioscience; Norfolk, UK: 2007. p. 285-301.
- Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. Cell Microbiol. 2004; 6:509–519. [PubMed: 15104593]
- Van Poppel NF, Welagen J, Vermeulen AN, Schaap D. The complete set of *Toxoplasma gondii* ribosomal protein genes contains two conserved promoter elements. Parasitology. 2006; 133:19–31. [PubMed: 16674839]
- Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, Reeder JC, Crabb BS, Cowman AF. A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. Nature. 2006; 439:1004–1008. [PubMed: 16382237]
- Waller RF, McFadden GI. The apicoplast: a review of the derived plastid of apicomplexan parasites. Curr Issues Mol Biol. 2005; 7:57–79. [PubMed: 15580780]
- Xue X, Zhang Q, Huang Y, Feng L, Pan W. No miRNA were found in Plasmodium and the ones identified in erythrocytes could not be correlated with infection. Malar J. 2008; 7:47. [PubMed: 18328111]

Figure 1.

A tree of eukaryotes (with unresolved root) based on Adl et al. () and Dacks et al. (). This “egalitarian” view suggests that all the taxa are equally modern/ancient. The six supergroups and representative taxa are (clockwise): Excavata, *Leishmania major*, *Trypanosoma brucei*, *Trichomonas vaginalis*; Chromalveolata, *Plasmodium falciparum*, *Toxoplasma gondii*, *Tetrahymena thermophila*; Rhizaria; Archaeplastida, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Porphyra purpurea*; Amoebozoa, *Entamoeba histolytica*; Opisthokonta, *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*.