EDITORIAL

Plasmodium **comparative genomics**

Lisa Ranford-Cartwright, Elena Gómez-Díaz

Malaria is a serious infectious disease caused by unicellular eukaryotic parasites of the genus *Plasmodium*. Today more than 200 species exist [1], and whole genome sequence data is available for around 22 species [2]. At least five different *Plasmodium* species can infect humans, of which *Plasmodium falciparum* is the major cause of morbidity and mortality worldwide [3]. Other species are specific to vertebrates including rodents, birds and lizards.

The parasite moves between vertebrate hosts via the bite of a mosquito, which serves as the definitive host and vector. A parasite species is usually restricted in the number of different vertebrate host species it is able to infect, as a result of molecular differences, particularly in the complex interactions between the erythrocyte host cell and the parasite [4], as well as environmental factors [5]. While malaria parasites are vertebrate host specialists, they are vector host generalists: human and rodent species can transmit through a variety of *Anopheles* mosquitoes, whereas avian species utilise diverse vector taxa including culicine mosquitoes such as those in the genera *Aedes* and *Culex*. During their life-cycles in the two hosts, malaria parasites must adapt to a variety of environmental changes, for example in temperature, oxygen levels, and nutrient accessibility, inhabiting different tissues and cells within the vertebrate host, and being predominantly extracellular in the invertebrate host. In these transits the parasite requires different sets of proteins to successfully colonize and survive in different host niches. As a result, there are major differences in patterns of gene expression at different lifecycle stages [6], and this is achieved through a very tight control of the cell cycle and differentiation processes, which are sensitive to environmental stimuli.

Malaria parasites share common genomic features such as the division of the nuclear genome into 14 chromosomes, and the presence of two extra-chromosomal elements, the mitochondrial and apicoplast genomes. Extensive genome sequence data exists for >20 *Plasmodium* species, revealing a general similarity in size (18-33 Mb) and gene number (5000-7800) [2]. More closely-related *Plasmodium* species exhibit syntenic blocks of genes, but generally exhibit high divergence in the sub-telomeric regions [7, 8]. The biological and genomic similarities between the *Plasmodium* species have been exploited particularly to study host-parasite interactions using non-human infecting species, as illustrated in this issue by research on simian malaria species described by Galinski [9]. Some simian malaria parasite species are capable of causing zoonotic infections in humans, and thus they are important as causative agents of human disease [10]. However, *P. knowlesi* and *P. cynomolgi* in simian hosts have proven valuable as model organisms for the human malaria species *P. falciparum* and *P. vivax*, for example in following transcriptomic and proteomic changes during longitudinal infections.

Functional genomics studies in *Plasmodium* have been hampered by the unique characteristics of the genomes of malaria parasites compared with those of other eukaryotes. One is the AT-richness of the genome that varies between species: *P. falciparum* has one of the most AT-rich genomes known to date (>80% A or T overall) and ~90% (A+T) in non- coding regions. As a result, many regions of the parasite genome are highly repetitive [11], which together with a biased codon usage, has been suggested to contribute to variation in the kinetics of gene expression across the genome [12, 13]. Importantly, *Plasmodium* species appear to lack the canonical non-homologous end joining repair mechanisms [14], and so genome engineering techniques like CRISPR-Cas9 and zinc-finger nuclease technologies [15] rely on homologous recombination mechanisms to repair double-strand

breaks in these organisms. The application of CRISPR/Cas9 in *Plasmodium* research is described in this issue by Lee and colleagues [16]. They describe the many uses of the CRISPR/Cas9 system for genome editing, since the first reported use in *Plasmodium* in 2014, and provide a thorough analysis of the factors to be considered in designing a CRISPR/Cas9 experiment. Their review also provides a tantalising vision of the future uses of the CRISPR system to alter transcription and to modify the epigenetic landscape of the genome.

In *Plasmodium*, like other eukaryotes, the regulation of gene expression during development occurs via a combination of transcriptional mechanisms, the amount of mRNA that is produced from a particular gene, and post-transcriptional mechanisms that regulate the translation of mRNA into proteins. In addition, translational and post-translational regulation operate at certain points in the parasite life-cycle, i.e. transmissible stages, where a delay between transcript and protein levels of nearly 30% has been documented [17]. Both chromatin-mediated and epigenetic processes, the latter occurring as non-genetic heritable changes in gene expression, play a key role in the dynamics of transcriptional regulation during development, and in parasite-host adaptation. The control mechanisms have been extensively studied only in *Plasmodium falciparum*.

Plasmodium genes display a conventional bipartite promoter architecture, with a transcription initiation site immediately upstream of the coding sequence that is required for baseline transcriptional activity, and a variable number of upstream *cis*-regulatory sequences that are bound by transcription factors (monocistronic model [18]). In *Plasmodium* there is only one family of transcription factors with master roles in the regulation of transcription. The Apicomplexan AP2 (ApiAP2) family includes 27 members [19, 20], an extremely low number of DNA binding proteins compared with the diversity and abundance of transcription factors identified in other eukaryotes. These transcription factors are expressed in a highly stage specific manner and bind to the promoters of different functional sets of genes that are activated at different times through parasite development. Toenhake & Bartfai in this issue [21] review the peculiarities of core promoter recognition in the compact and AT-rich *Plasmodium* genome, and how the transcriptional apparatus and other DNA-associated proteins have adapted to these unique genomic characteristics. Their review stresses the important role of cis-regulatory sequences and DNA binding proteins involved in life cycle progression during blood-stage development. That is, together with ApiAP2 transcription factors, chromatin remodelling proteins are also emerging as important regulators of developmental progression and the interaction of transcription factors with their cognate sequences. The authors also provide an overview of approaches directed towards the characterization of these gene regulatory networks in *Plasmodium,* such as chromatin accessibility profiling, and how these advances are being used to add insights into the stage-specific and clone-specific regulatory elements involved in transcriptional activation.

In *Plasmodium*, like in other eukaryotes, the structure and function of chromatin is dynamic and contributes chiefly to the regulation of gene expression. Among chromatin-associated processes the most studied are histone variants and post-translational modifications (PTMs) of histones. *Plasmodium* genomes encode a canonical form of each core histone, which are common to other eukaryotes, plus four histone variants H2A.Z, H2B.Z, H3.3 and a centromere-specific H3 (CenH3) [22]. Previous studies have revealed that both histone variant replacement and various histone PTMs are dynamically regulated through *Plasmodium* development and are known to alter chromatin structure and affect levels of transcription [23-25]. Most of the *Plasmodium* genome is in an euchromatic active state and is marked with the variant histones H2A.Z and H2B.Z as well as the histone modifications H3K4me3, H3K9ac and H4K8ac. Telomeres and sub-telomeric repeats are regions of constitutive heterochromatin that are marked with the repressive histone modification mark H3K9me3 and enriched in binding of heterochromatin protein 1 (HP1). Between these two are regions of bistable chromatin that switch between active (H3K9ac) and silent (H3K9me3/HP1)

functional states depending on the phase of the cell cycle and in response to environmental conditions [26, 27].

At higher levels of chromatin structure, the position and distribution of nucleosomes and how the chromatin fibre is organized and located within the nucleus impose another layer of transcriptional control.
In this issue Abel and Le Roch review the link between "form and function" of the *Plasmodium* genome at various levels: chromatin structure, nucleosome dynamics and 3D genome organization [28].The authors describe key similarities and differences at the level of the core promoters, including a divergent TFIID complex, histone PTMs and histone variants contributing to chromatin structure, and discuss the current debate surrounding nucleosome landscapes in *Plasmodium*. The review also explores chromosome conformation capture (3C)-based methods to study 3D nuclear architecture, and how chromatin structure and the nucleus itself are heavily remodelled during the asexual, sexual and transmission stages of the life cycle of *P. falciparum*, with a particular focus on virulence genes.

Perhaps the most striking and yet most dangerous attribute of *Plasmodium* parasites, is their adaptive potential. Natural genomic and phenotypic variation exist in the parasites found in disease-endemic regions. In addition to standing genetic variation, malaria parasites are masters in their ability to be plastic. To cope with environmental heterogeneity malaria parasites have evolved mechanisms for rapid evolutionary change, including genetic and non-genetic mechanisms. In this issue, Cowell and Winzeler describe one of the important consequences of this evolutionary plasticity: the evolution of resistance, usually through mutation, to almost all antimalarial drugs that have ever been used to treat the disease [29].

Apart from genetic changes that affect transcript levels, epigenetic variation contributes chiefly to phenotypic variation and plasticity in *Plasmodium*. This plasticity is key for adaptation of malaria parasites to changes in their environment. That is, *Plasmodium* possesses the ability to generate and maintain cell-to-cell transcriptional variability for genes that relate to growth, virulence and transmission. This ability is fundamental for parasite transmission and survival in heterogenous environments. Llorà-Batlle and colleagues review the underlying epigenetic mechanisms that allow the parasite to alter its transcriptome and switch between alternative phenotypes [30]. They distinguish between transcriptional changes driven by stochastic events, and those that occur in response to external cues. Transcriptional variation has been studied mainly in the blood stages of the parasite under *in vitro* conditions, while little is known about gene regulation *in vivo* and in other stages of the parasite. In contrast to the other reviews in this issue, Gómez-Díaz & Ruiz focus on transcriptional regulation and plasticity during development and adaptation in the life-cycle of the parasite in the mosquito host [31]. The authors describe the transcriptomes and epigenomes of the parasite in the mosquito stages, the master regulators involved in mosquito-life cycle transitions, and the existing evidence of epigenetic and transcriptional variation. They finally advocate the potential of the mosquito to study *Plasmodium* gene regulation and transcriptional plasticity *in vivo*.

The seven reviews in this special issue thus set out a comparative and integrative view of state-of-the art functional genomics in different *Plasmodium* species. The authors describe how these organisms modify their genomes and epigenomes, and regulate gene expression in order to generate phenotypic plasticity, which ultimately allows the parasites to adapt to changing environments during their two-host lifecycle, and contribute to their virulence and pathogenesis.

This is a fast-moving field, and technological advances will soon allow single-cell transcriptomic and epigenomic analyses, and fine-mapping of variation between parasites within individual hosts. In addition, we anticipate that many of the current limitations faced by functional genomics studies in

malaria will be overcome, including the possibility to study the non-culturable human malaria species, that contribute to human morbidity and mortality. Understanding the mechanisms malaria parasites use to regulate their genomes through the whole parasite life-cycle, how they differ among species, how malaria parasites adapt to the heterogeneous environment of their hosts, are some of the challenges that the field faces over the coming years. This knowledge will be fundamental for the design of more effective malaria control interventions in the context of a globalized and changing world.

Lisa Ranford-Cartwright
Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, UK

Elena Gómez-Díaz
Institute of Parasitology and Biomedicine López-Neyra, Spanish National Research Council (CSIC), Spain.

Address:
Avda. del Conocimiento, 17
18016 Armilla, Granada (Spain)
Tel: +34 958181643
Email: elena.gomez@csic.es

REFERENCES

1. Martinsen ES, Perkins SL. The Diversity of Plasmodium and other Haemosporidians: The Intersection of Taxonomy, Phylogenetics and Genomics, Malaria Parasites: Comparative Genomics , Evolution and Molecular Biology 2013:1-15.
2. Aurrecoechea C, Brestelli J, Brunk BP et al. PlasmoDB: a functional genomic database for malaria parasites, Nucleic Acids Research 2009;37:D539-D543.
3. Organization WH. World Malaria Report 2018. Geneva, Switzerland: World Health Organization, 2018.
4. Scully EJ, Kanjee U, Duraisingh MT. Molecular interactions governing host-specificity of blood stage malaria parasites, Current Opinion in Microbiology 2017;40:21-31.
5. Fecchio A, Wells K, Bell JA et al. Climate variation influences host specificity in avian malaria parasites, Ecology Letters 2019;22:547-557.
6. Le Roch KG, Zhou Y, Blair PL et al. Discovery of gene function by expression profiling of the malaria parasite life cycle, Science 2003;301:1503-1508.
7. Carlton JM, Vinkenoog R, Waters AP et al. Gene synteny in species of Plasmodium, Mol.Biochem.Parasitol. 1998;93:285-294.
8. Kooij TWA, Carlton JM, Bidwell SL et al. A Plasmodium whole-genome synteny map: Indels and synteny breakpoints as foci for species-specific genes, Plos Pathogens 2005;1:349-361.
9. Galinski MR. Functional genomics of simian malaria parasites and host–parasite interactions. *Brief Funct Genom* 2019:xx–xx.
10. Singh B, Kim SL, Matusop A et al. A large focus of naturally acquired Plasmodium knowlesi infections in human beings, Lancet 2004;363:1017-1024.
11. Gardner MJ, Hall N, Fung E et al. Genome sequence of the human malaria parasite Plasmodium falciparum, Nature 2002;419:498-511.
12. Sinha I, Woodrow CJ. Forces acting on codon bias in malaria parasites, Scientific Reports 2018;8.

13. Chan S, Ch'ng JH, Wahlgren M et al. Frequent GU wobble pairings reduce translation efficiency in Plasmodium falciparum, Scientific Reports 2017;7.
14. Lee AH, Symington LS, Fidock DA. DNA Repair Mechanisms and Their Biological Roles in the Malaria Parasite Plasmodium falciparum, Microbiol.Mol.Biol.Rev. 2014;78:469-486.
15. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, Trends in Biotechnology 2013;31:397-405.
16, Lee MCS, Lindner SE, Lopez-Rubio J-J, Llinás M. Cutting back malaria: CRISPR/Cas9 genome editing of Plasmodium. Brief Funct Genom 2019:xx-xx.
17. Caro F, Ahyong V, Betegon M et al. Genome-wide regulatory dynamics of translation in the Plasmodium falciparum asexual blood stages, eLife 2014;3.
18. Lanzer M, de Bruin D, Ravetch JV. Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates, The EMBO journal 1992;11:1949-1955.
19. Painter HJ, Campbell TL, Llinas M. The Apicomplexan AP2 family: integral factors regulating Plasmodium development, Mol Biochem Parasitol 2011;176:1-7.
20. Campbell TL, De Silva EK, Olszewski KL et al. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite, PLoS Pathog 2010;6:e1001165.
21. Toenhake CG, Bártfai R. What functional genomics has taught us about transcriptional regulation in malaria parasites. Brief. Funct. Genomics 2019:xx–xx.
22. Miao J, Fan Q, Cui L et al. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation, Gene 2006;369:53-65.
23. Gupta AP, Chin WH, Zhu L et al. Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum, PLoS Pathog 2013;9:e1003170.
24. Bartfai R, Hoeijmakers WA, Salcedo-Amaya AM et al. H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3, PLoS Pathog 2010;6:e1001223.
25. Gómez-Díaz E, Yerbanga RS, Lefevre T et al. Epigenetic regulation of Plasmodium falciparum clonally variant gene expression during development in Anopheles gambiae, Sci Rep 2017;7:40655.
26. Lopez-Rubio JJ, Mancio-Silva L, Scherf A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites, Cell Host Microbe 2009;5:179-190.
27. Cortes A, Deitsch KW. Malaria Epigenetics, Cold Spring Harb Perspect Med 2017;7.
28. Abel S, Le Roch KG. The role of epigenetics and chromatin structure in transcriptional regulation in malaria parasites. Brief Funct Genom 2019:xx-xx.
29. Cowell AN., Winzeler EA. The genomic architecture of antimalarial drug resistance. Brief Funct Genom 2019:xx-xx.
30. Llorà-Batlle O, Tintó-Font E, Cortés A. Transcriptional variation in malaria parasites: why and how. Brief Funct Genom 2019:xx-xx.
31. Ruiz JL, Gómez-Díaz E. The second life of Plasmodium in the mosquito host: gene regulation on the move. Brief Funct Genom 2019:xx-xx.