Tropical Biomedicine 40(3): 351-355 (2023)
https://doi.org/10.47665/tb.40.3.012

RESEARCH ARTICLE

Genome-wide *in silico* analysis for novel regulators of gametocyte development critical for transmission of *Plasmodium falciparum*

Zhang, C.[1*]

[1]Department of Microbiology and Parasitology; Anhui Provincial Laboratory of Microbiology and Parasitology; Anhui Key Laboratory of Zoonoses, School of Basic Medical Sciences, Anhui Medical University, Hefei 230032, People's Republic of China
*Corresponding author: happy2008con@aliyun.com

ARTICLE HISTORY

Received: 12 May 2023
Revised: 14 July 2023
Accepted: 17 July 2023
Published: 30 September 2023

ABSTRACT

Malaria, caused by the unicellular Apicomplexan protozoa of the genus *Plasmodium*, is an infectious disease transmitted via female Anopheles mosquitoes. The sexual stage (gametocytes) of malaria parasites is the key to the transmission of parasites from vertebrate hosts to mosquitoes, representing critical bottleneck of the parasite life cycle. This study has established a systematic computational pipeline to achieve the genome-wide *in silico* analysis and find 708 novels potentially indispensable genes for gametocyte development, consisting of 644 protein coding genes, 56 ncRNA genes and 8 pseudogenes, with a total of 191 genes in the transmembrane, 29 protein coding genes to be exported proteins, and 58 genes in apicoplast regions. Furthermore, Gene Ontology analysis showed that the largest cluster was cellular processes with nucleus and cytosol highest, followed by molecular function with binding and oxidoreductase activities abundant. Meanwhile, when a text searched, using PlasmoDB, there were 300 genes with annotations of "putative", and 196 genes with annotations of "unknown function". These data would be helpful to provide potential targets for effective malaria transmission-blocking strategies.

Keywords: Malaria; *Plasmodium falciparum*; gametocyte; indispensable genes; *In Silico* Analysis.

INTRODUCTION

Malaria, caused by the unicellular Apicomplexan protozoa of the genus *Plasmodium*, remains the main burden on global public health, despite widespread efforts to eradicate malaria (Hema *et al*., 2021; Uwimana *et al*., 2020). The latest report released by the World Health Organization (WHO) shows that there are approximately 247 million malaria cases worldwide, with an estimated 619, 000 deaths (WHO, 2023). Most malaria-related pathologies are caused by the adhesion of parasitized red blood cells (pRBCs) to endothelial cells, platelets, and uninfected red blood cells (rosetting), leading to the sequestration of pRBCs, thereby blocking microvessels and leading to complications such as cerebral malaria, stroke, and death (Miller *et al*., 2002; Wassmer *et al*., 2015). As the burden of parasites increases, malaria parasites face an important decision whether to continue spreading or invest in sexual transmission, a dead end in development if not ingested by susceptible Anopheles mosquitoes. The *Plasmodium falciparum* (*P. falciparum*) is one of the most virulent human malaria parasites. Because only a limited proportion (about 5-10%) of asexually replicating red blood cell parasites differentiate into non replicating gametocytes, and fewer gametocytes are allowed to develop in the Anopheles vector, this sex conversion process is a promising transmission intervention point, taking the regulator of gametocyte development as the direct target of vaccines and drugs to block transmission (Smith *et al*., 2014; Bennink *et al*., 2016).

After several years of international efforts, the genome of *P. falciparum* was first published in 2002 (Gardner *et al*., 2002). With the emerging sequencing technologies, various efforts were facilitated by annotating the *Plasmodium* genome (Kissinger *et al*., 2002; Bahl *et al*., 2003). Multiple studies based on proteome and genome were conducted to develop novel technologies to understand disease-resistance mechanisms in *Plasmodium* (Sardar *et al*., 2021; Sourabh *et al*., 2021). While the advancement of sequencing technologies are highly bene cial to understand speci c pathways and mechanism related to the disease, the most undesirable aspect for any newly sequenced genome is when almost half of the annotated proteins or genes are in the uncharacterized category. These uncharacterized proteins with predicted ORF regions without validated translation evidence can be categorized as "Hypothetical proteins" (Ijaq *et al*., 2015), which are conserved proteins and found across diverse phylogenetic lineages, thus the absence of functional annotations of the proteins is a serious concern. These proteins may be performing crucial functions, which can unravel more details of the molecular basis of the disease infection and pathogenesis (Singh & Singh, 2018).

However, 20 years after the first report of stable transfection of *P. falciparum*, only 368 genes have reported knockout attempts (de Koning-Ward *et al*., 2015; Sanderson & Rayner, 2017), and 529 genes are targeted in the rodent *P. berghei* (Janse *et al*., 2011). Until recently, as the first genome-wide genetic screening, the *piggyBac* system identified 2680 genes, which are crucial for the asexual blood-stage growth of *P. falciparum in vitro* (Zhang *et al*., 2018). In addition, based on the large-scale gene targeting vector library in the genome of *P. berghei*, the analysis of 1342 blood stage survival mutants identified 461 genes, which are needed for the optimal transmission

of parasites through mosquitoes (Bushell *et al*., 2017). Similarly, by analyzing a set of 1302 blood stage *P. berghei* mutants, 30 genes are required for the formation of male and female gametocytes, of which 21 genes are required for females, and 14 genes are required for male gametocytogenesis (Russell *et al*., 2023). In *P. falciparum*, except the intra-erythrocytic stage, until now, the genetic screen with full genome coverage has not yet been achieved in other stages of the life cycle. Although the overlap of essential gene orthologs is highly conservative between *P. berghei* and *P. falciparum*, there still are functions of many genes to be experimentally identified in each stage of life cycle, especially gametocyte development stage of *P. falciparum*.

This study, through genome-wide *in silico* analyzed how many genes were still needed be validated to be potentially indispensable genes and characterized functions of these genes to explore their potential as novel regulators of gametocyte development critical for transmission of *P. falciparum*. And, these data could then be used by malaria researchers for future experimental validation. In addition, this research method may also be an effective way to analyze potential new regulatory factors in other stages of the life cycle of *P. falciparum*.

MATERIALS AND METHODS

Data extraction and analysis of potentially indispensable genes for gametocyte development

The gene sequence of *P. falciparum* was searched and analyzed, using PlasmoDB (https://plasmodb.org/plasmo/app), which includes all the genomes of malaria parasites that have been sequenced so far. Phenotypes for disruptions of *Plasmodium* genes were mainly downloaded from PhenoPlasm (http://phenoplasm.org/), a community database of malaria parasite gene phenotypes, which has included 7830 phenotype genes until now, and is always kept up to date, using data from the Rodent Malaria genetically modified parasites database (RMgmDB: https://www.pberghei.eu/index.php), the Adams lab saturation screen in *P. falciparum* (Zhang *et al*., 2018), PlasmoGEM (Bushell *et al*., 2017), and the literature collated by the curator or other users.

Functional characterization of novel potentially indispensable genes for gametocyte development

For functional characterizing protein coding genes, the transmembrane, exported protein, subcellular localization and gene ontology annotations were carried out using self-contained tools from PlasmoDB (https://plasmodb.org/plasmo/app).

RESULTS

The analysis of novel potentially indispensable genes for gametocyte development

To analyze novel potentially indispensable genes for gametocyte development in *P. falciparum*, as shown in Figure 1, a system computing pipeline, based on sequence extensive comparative analysis, was employed, including analysis of *Pf* genes with disruption possible (viable mutant) and without knockout, *Pf* gametocyte transcriptional genes and so on. And the detailed process of analyzing novel potentially indispensable genes for gametocyte development using PlasmoDB, were shown in Figure 2, each calculation step of the pipeline filters out the best available supporting evidence to screen for proteins. Currently, for *P. falciparum* 3D7 [Reference], its nuclear genome as 23.33 Mb in size, including 5720 genes, is defined (Release 59, 2022-08-30, www.plasmodb.org). From PhenoPlasm (http://phenoplasm.org/), *Pf* genes (2591, ~45%), which are refractory to deletion, *Pf* gametocyte transcriptional genes (1592, ~28%) (Young *et al*., 2005), all published genes (70, ~1.2%) that are essential or not essential for *Pf* Gametocyte, *Pf* ortholog genes (330, ~5.8%) of *Pb* Gametocyte, and *Pf* ortholog genes (109, ~1.9%) of *Py* Gametocyte,

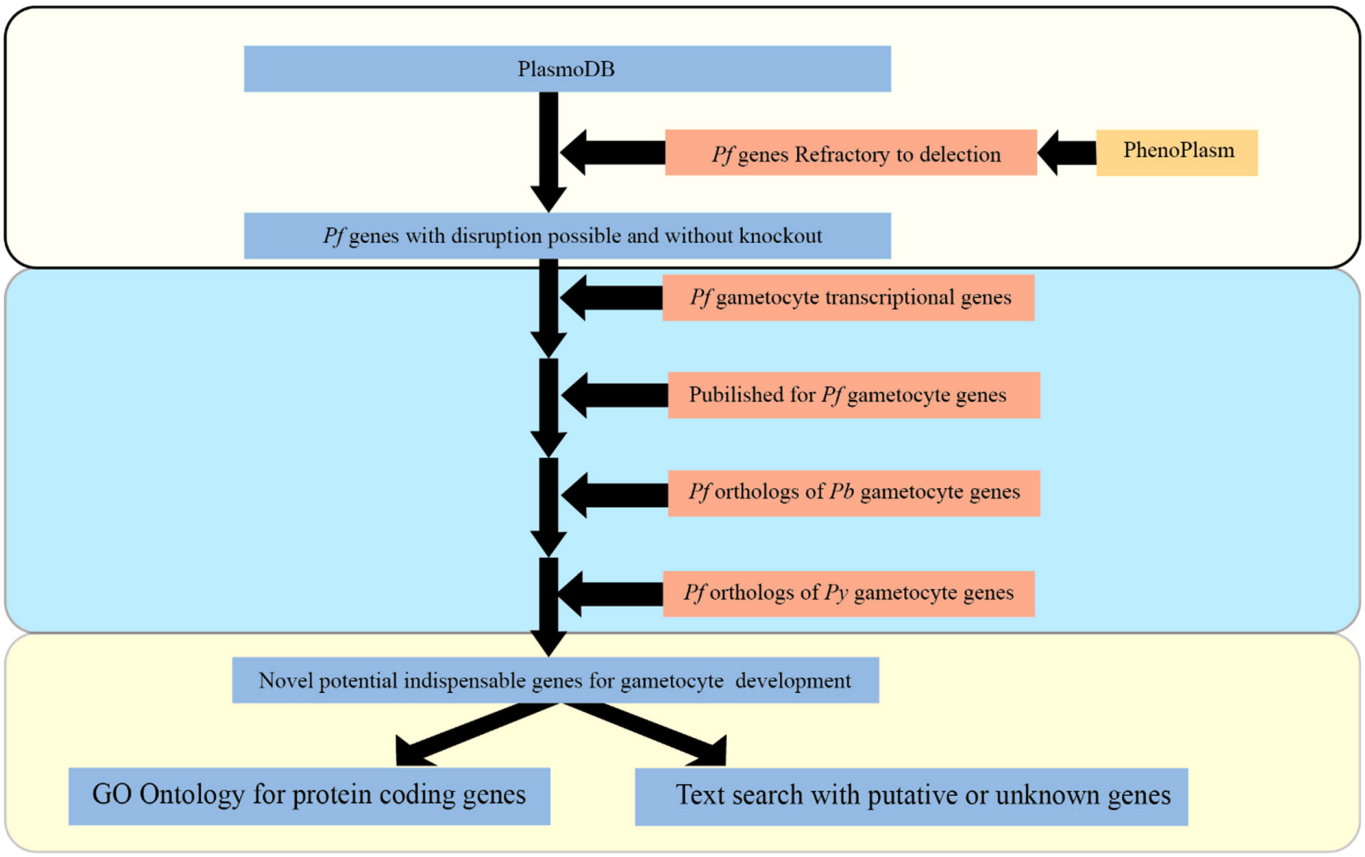

Figure 1. The workflow of the entire research method.

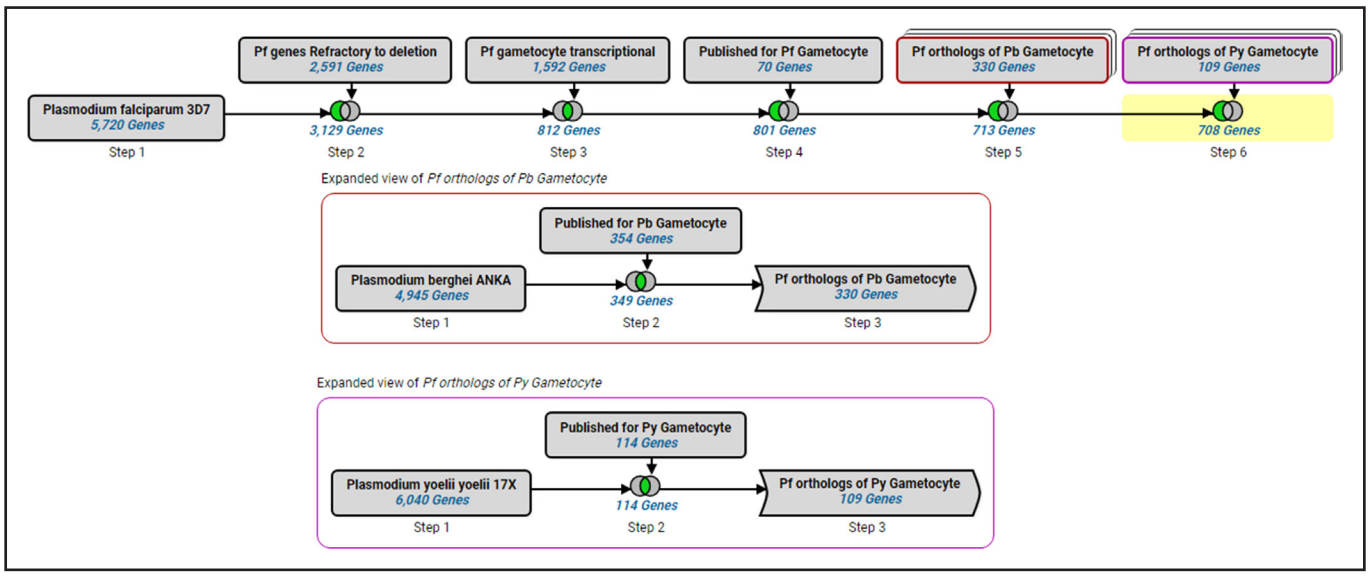

Figure 2. The detailed process of analyzing novel potentially indispensable genes for gametocyte development using PlasmoDB.

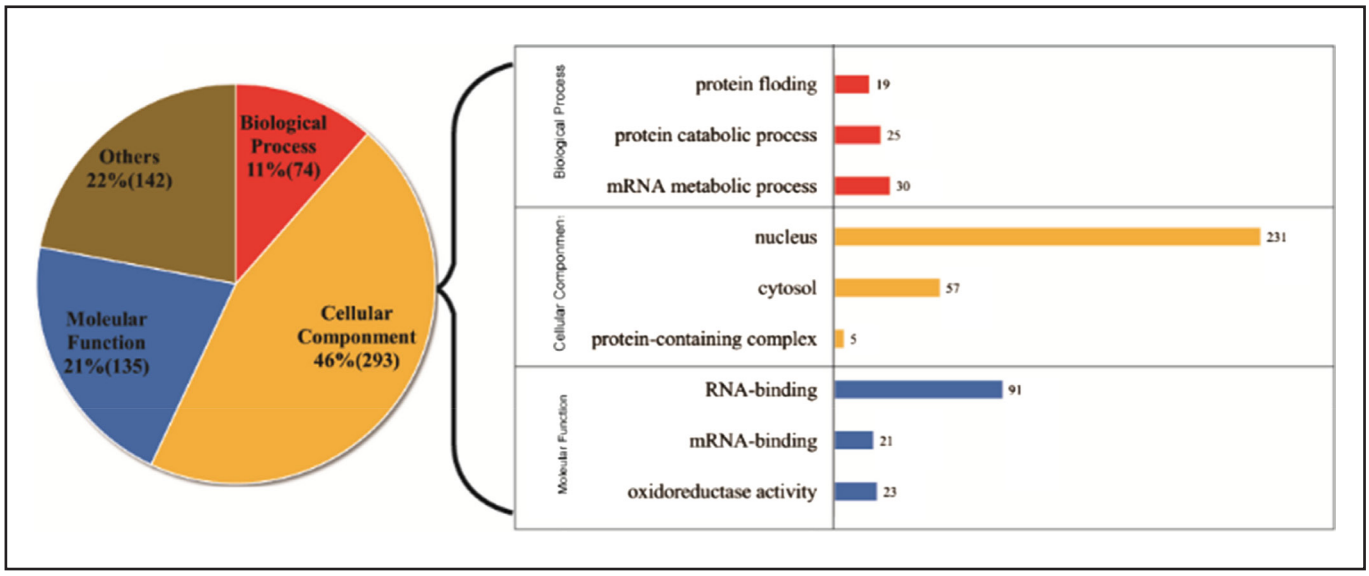

Figure 3 Gene Ontology (GO) annotation of novel potentially indispensable genes for gametocyte development with functional characteristic sites in the primary sequence to identify functional categories: cellular components, biological processes, and molecular functional levels. The left panel displays a pie chart of the overall percentage. The right panel displays the precise decomposition of subcomponents for each category.

were extracted; after removing and analyzing above genes, finally, 708 (~12%) novel potentially indispensable genes for gametocyte development were selected for functional characterization, using the downstream steps in the research pipeline.

Functional characterization of novel potentially indispensable genes for gametocyte development

Out of 708 novels potentially indispensable genes for gametocyte development, there were 644 protein coding genes, 56 ncRNA genes and 8 pseudogenes, which were defined here, based on at least one frame shift or premature termination codon. Using PlasmoDB (https://plasmodb.org/plasmo/app), a total of 191 genes exist in the transmembrane; there were 29 protein coding genes for exported protein; in addition, subcellular localization prediction shows that 58 genes were predicted to be present in apicoplast regions, followed by RBC membrane (Pexel motif) (43 genes) and RBC membrane (HT motif) (29 genes). 74 proteins were also found to have signal peptide.

In addition, all 644 protein coding genes were annotated in Gene Ontology (GO). Of these, 502 protein coding genes were classified as several categories: the largest cluster was cellular processes, followed by molecular function, while nucleus and cytosol were highest in cellular components. Similarly, in molecular function, binding and oxidoreductase activities were most abundant. For these genes, the representation of Gene Ontology annotation of all the categories was shown in Figure 3.

Moreover, considering these terminologies, "unknown" function, and "putative" function (where the function is predicted but requires experimental verification) are now used for functional annotation for genes of *Plasmodium*. Then, as shown in Figure 4, using PlasmoDB, a text searched against 708 novels potentially indispensable genes for annotations of "putative function", yielding approximately 42% (300 genes) of all genes, and a text search for annotations of "unknown function", yielding approximately 28% (196 genes) of all genes.

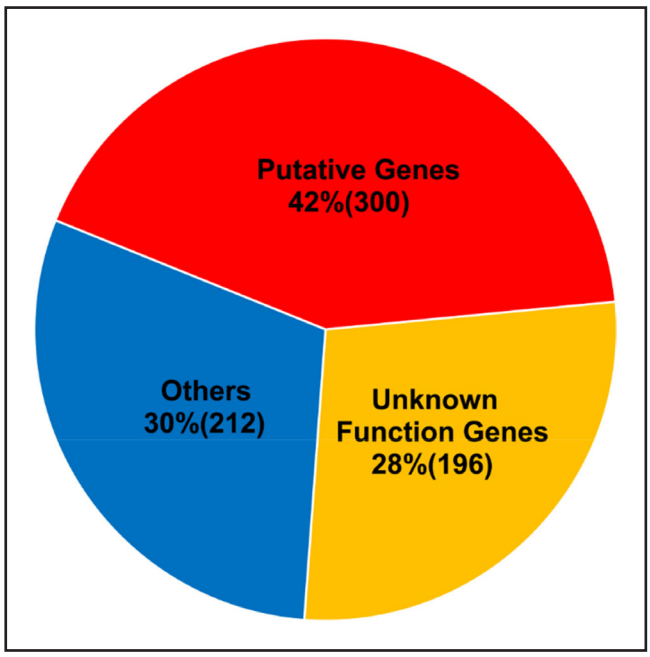

Figure 4. The text search against novel potentially indispensable genes for annotations of "putative" and for annotations of "unknown function" using PlasmoDB. Generally, terminologies, "unknown" function and "putative" function (where the function is predicted but requires experimental verification) are now used for functional annotation for genes of *Plasmodium* in PlasmoDB.

DISCUSSION

Malaria is an infectious disease transmitted by mosquitoes and considered as one of the diseases that has existed for centuries. It is crucial for maintaining transmission between mosquito vectors and vertebrate hosts (WHO, 2019). The sexual stage (gametocytes) of malaria parasites is crucial for the transmission of parasites from vertebrate hosts to mosquitoes and the completion of the sexual stage of the life cycle. In the midgut of mosquitoes, gametocytes are activated, and gametogenesis proceeds rapidly to form male gametes and female gametes, which are fertilized to form zygote (WHO, 2019; Elvis *et al*., 2022). These stages are key bottlenecks in the parasite's life cycle and targeting them through transmission blocking drugs/vaccines is crucial for achieving effective malaria transmission blocking strategies (Elvis *et al*., 2022). With the advancement of sequencing technology, annotation of the malaria parasite genome has promoted the discovery of various new drugs or vaccine development (Kissinger *et al*., 2002; Bahl *et al*., 2003). For example, based on proteomics and genomics, new technologies have been developed to understand the disease resistance mechanism of malaria parasites (Sardar *et al*., 2021; Sourabh *et al*., 2021).

This study constructed a systematic computational pipeline to achieve the genome-wide *in silico* analysis and find 708 novels potentially indispensable genes for gametocyte development. These genes consist of 644 protein coding genes, 56 ncRNA genes and 8 pseudogenes. Of these genes, it is estimated that a total of 191 genes existed in the transmembrane, 29 protein coding genes were considered as exported proteins, and 58 genes were predicted to be present in apicoplast regions. Furthermore, Gene Ontology (GO) analysis showed that the largest cluster was cellular processes, followed by molecular function, while nucleus and cytosol were highest in cellular components, among molecular function, binding and oxidoreductase activities were most abundant. At the same time, when a text searched against 708 novels potentially indispensable genes, there were ~42% (300 genes) of all genes with annotations of "putative", and a ~28% (196 genes) of genes with annotations of "unknown function". With the continuous revolutionization of next generation sequencing technologies, large-scale genomic or transcriptome data of several organisms can be generated in a single run. Although annotation methods for the analysis of generated data are available on a large scale, the most disconcerting aspect is that half of the generated data remains uncharacterized and comes under the "Hypothetical category". In recent years technical breakthroughs have made forward and reverse genome-scale screens in *Plasmodium* possible. Furthermore, the adaptation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-Associated protein 9 (CRISPR/Cas9) technology has dramatically improved gene editing ef ciency at the single gene level (Ishizaki *et al*., 2022). The implementation of genetic screens to the study of *Plasmodium* has signi cantly increased the number of genes associated with a phenotype and revealed a high degree of gene essentiality. These screens shine new light on parasite biology and present a catalogue of new potential drug targets, such as the parallel development of ef cient CRISPR/Cas9 gene editing tools for *Plasmodium* and CRISPR/Cas9 screens in the related apicomplexan *Toxoplasma gondii* forms a foundation for developing CRISPR/Cas9 screens in *Plasmodium*, which to date has not been feasible, because *Plasmodium* CRISPR/Cas9 screens must overcome the lack of non-homologous end joining. Enhancing the ef cacy of existing micro-homology mediated end joining could offer a way forward. Efforts towards developing conditional methods to understand the function of essential genes at scale would be of special merit. Our data could then be used by malaria researchers for future experimental validation to identify targets for transmission-blocking drugs/vaccines. In addition, this research pipeline can become an effective platform for analyzing potential new regulatory factors in other stages of the life cycle of *P. falciparum*, which will help provide potential targets for combating malaria.

ACKNOWLEDGMENT

This work was supported by Anhui Medical University (2021-ZQKY-131), Scientific Research of BSKY (XJ201807), and the Foundation of Education Department of Anhui province (KJ2021A0213).

Competing interests

No conflicts of interest existed.

REFERENCES

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., Ginsburg, H., Gupta, D., Kissinger, J.C., Labo, P. *et al*. (2003). PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Research* **31**: 212-215. https://doi.org/10.1093/nar/gkg081

Bennink, S., Kiesow, M.J. & Pradel, G. (2016). The development of malaria parasites in the mosquito midgut. *Cellular Microbiology* **18**: 905-918. https://doi.org/10.1111/cmi.12604

Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R.E. *et al*. (2017). Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell* **170**: 260-272.e8. https://doi.org/10.1016/j.cell.2017.06.030

de Koning-Ward, T.F., Gilson, P.R. & Crabb, B.S. (2015). Advances in molecular genetic systems in malaria. *Nature Reviews Microbiology* **13**: 373-387. https://doi.org/10.1038/nrmicro3450

Elvis, Q., Faustina, P., Jilong, S., Miao, L., Shijie, Y., Li, Y. & Chao, Z. (2022). ApiAP2 gene-network regulates gametocytogenesis in *Plasmodium* parasites. *Cellular Microbiology* online. https://doi.org/10.1155/2022/5796578

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S. *et al*. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498-511. https://doi.org/10.1038/nature01097

Hema, K., Ahamad, S., Joon, H.K., Pandey, R. & Gupta, D. (2021). Atomic resolution homology models and molecular dynamics simulations of *Plasmodium falciparum* tubulins. *ACS Omega* **6**: 17510-17522. https://doi.org/10.1021/acsomega.1c01988

Ijaq, J., Chandrasekharan, M., Poddar, R., Bethi, N. & Sundararajan, V. S. (2015). Annotation and curation of uncharacterized proteins- challenges. *Frontiers in Genetics* **6**: 119. https://doi.org/10.3389/fgene.2015.00119

Ishizaki, T., Hernandez, S., Paoletta, M.S., Sanderson, T. & Bushell, E.S.C. (2022). CRISPR/Cas9 and genetic screens in malaria parasites: small genomes, big impact. *Biochemical Society Transactions* **50**: 1069-1079. https://doi.org/10.1042/BST20210281

Janse, C.J., Kroeze, H., van Wigcheren, A., Mededovic, S., Fonager, J., Franke-Fayard, B., Waters, A.P. & Khan, S.M. (2011). A genotype and phenotype database of genetically modified malaria-parasites. *Trends in Parasitology* **27**: 31-39. https://doi.org/10.1016/j.pt.2010.06.016

Kissinger, J.C., Brunk, B.P., Crabtree, J., Fraunholz, M.J., Gajria, B., Milgram, A.J., Pearson, D.S., Schug, J., Bahl, A., Diskin, S.J. *et al*. (2002). The *Plasmodium* genome database. *Nature* **419**: 490-492. https://doi.org/10.1038/419490a

Miller, L.H., Baruch, D.I., Marsh, K. & Doumbo, O.K. (2002). The pathogenic basis of malaria. *Nature* **415**: 673-679. https://doi.org/10.1038/415673a

Russell, A.J.C., Sanderson, T., Bushell, E., Talman, A.M., Anar, B., Girling, G., Hunziker, M., Kent, R.S., Martin, J.S., Metcalf, T. *et al*. (2023). Regulators of male and female sexual development are critical for the transmission of a malaria parasite. *Cell Host & Microbe* **31**: 305-319.e10. https://doi.org/10.1016/j.chom.2022.12.011

Sanderson, T. & Rayner, J.C. (2017). PhenoPlasm: a database of disruption phenotypes for malaria parasite genes. *Wellcome Open Research* **2**: 45. https://doi.org/10.12688/wellcomeopenres.11896.2

Sardar, R., Katyal, N., Ahamad, S., Jade, D.D., Ali, S. & Gupta, D. (2021). *In-silico* profiling and structural insights into the impact of nSNPs in the *P. falciparum* acetyl-CoA transporter gene to understand the mechanism of drug resistance in malaria. *Journal of Biomolecular Structure & Dynamics* **39**: 558-569. https://doi.org/10.1080/07391102.2020.1711807

Singh, G. & Singh, V. (2018). Functional elucidation of hypothetical proteins for their indispensable roles toward drug designing targets from *Helicobacter pylori* Strain HPAG1. *Journal of Biomolecular Structure and Dynamics* **36**: 906-918. https://doi.org/10.1080/07391102.2017.1302361

Smith, R.C., Vega-Rodrɒguez, J. & Jacobs-Lorena, M. (2014). The *Plasmodium* bottleneck: malaria parasite losses in the mosquito vector. *Memorias do Instituto Oswaldo Cruz* **109**: 644-661. https://doi.org/10.1590/0074-0276130597

Sourabh, S., Chauhan, M., Yasmin, R., Shehzad, S., Gupta, D. & Tuteja, R. (2021). *Plasmodium falciparum* DDX17 is an RNA helicase crucial for parasite development. *Biochemistry and Biophysics Reports* **26**: 101000. https://doi.org/10.1016/j.bbrep.2021.101000

Uwimana, A., Legrand, E., Stokes, B.H., Ndikumana, J.M., Warsame, M., Umulisa, N., Ngamije, D., Munyaneza, T., Mazarati, J.B., Munguti, K. *et al*. (2020). Emergence and clonal expansion of *in vitro* artemisinin-resistant *Plasmodium falciparum* kelch13 R561H mutant parasites in Rwanda. *Nature Medicine* **26**: 1602-1608. https://doi.org/10.1038/s41591-020-1005-2

Wassmer, S.C., Taylor, T.E., Rathod, P.K., Mishra, S.K., Mohanty, S., Arevalo-Herrera, M., Duraisingh, M.T. & Smith, J.D. (2015). Investigating the pathogenesis of severe malaria: a multidisciplinary and cross-geographical approach. *American Journal of Tropical Medicine and Hygiene* **93**: 42-56. https://doi.org/10.4269/ajtmh.14-0841

WHO (World Health Organization) (2019). The "World malaria report 2019" at a glance. https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019. Accessed 4 December 2019.

WHO (World Health Organization) (2023). Fact sheets: Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria. Accessed 29 March 2023.

Young, J.A., Fivelman, Q.L., Blair, P.L., de la Vega, P., Le Roch, K.G., Zhou, Y., Carucci, D.J., Baker, D.A. & Winzeler, E.A. (2005). The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Molecular and Biochemical Parasitology* **143**: 67-79. https://doi.org/10.1016/j.molbiopara.2005.05.007

Zhang, M., Wang, C., Otto, T.D., Oberstaller, J., Liao, X., Adapa, S.R., Udenze, K., Bronner, I.F., Casandra, D., Mayho, M. *et al*. (2018). Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**: eaap7847. https://doi.org/10.1126/science.aap7847