Coordinated regulation of gene expression in *Plasmodium* female gametocytes by two transcription factors

Yuho Murata[1], Tsubasa Nishi[1], Izumi Kaneko[1], Shiroh Iwanaga[2], and Masao Yuda[1,*]

[1]Department of Medical Zoology, Mie University School of Medicine, Mie, Tsu 514-8507, Japan

[2]Department of Molecular Protozoology, Research Center for Infectious Disease Control, 3-1 Yamadaoka, Suita, Osaka 565-0871 Japan

* Masao Yuda

Email: m-yuda@med.mie-u.ac.jp

Competing Interest Statement: The authors declare no competing interest.

Keywords: *Plasmodium*, female gametocyte, transcriptional regulation

This PDF file includes:

Main Text
Figures 1 to 6

Figure S1
Captions of supplementary tables (Table S1 to S4)

Abstract

Gametocytes play key roles in the *Plasmodium* lifecycle. They are essential for sexual reproduction as precursors of the gametes. They also play an essential role in parasite transmission to mosquitoes. Elucidation of the gene regulation at this stage is essential for understanding these two processes at the molecular level and for developing new strategies to break the parasite life cycle. We identified a novel *Plasmodium* transcription factor (TF), designated as a partner of AP2-FG or PFG. In this paper, we report that this TF regulates the gene expression in female gametocytes in concert with another female-specific TF AP2-FG. Upon the disruption of *PFG*, majority of female-specific genes were significantly downregulated, and female gametocyte lost the ability to produce ookinetes. ChIP-seq analysis showed that it was located in the same position as AP2-FG, indicating that these two TFs form a complex. ChIP-seq analysis of PFG in *AP2-FG*-disrupted parasites and ChIP-seq analysis of AP2-FG in *PFG*-disrupted parasites demonstrated that PFG mediates the binding of AP2-FG to the ten-base motif and that AP2-FG binds another motif, GCTCA, in the absence of PFG. In promoter assays, this five-base motif was identified as another female-specific cis-acting element. Genes under the control of the two forms of AP2-FG, with or without PFG, partly overlapped; however, each form had target preferences. These results suggested that combinations of these two forms generate various expression patterns among the extensive genes expressed in female gametocytes.

Main Text

Introduction

Plasmodium gametocytes are erythrocytic sexual stages essential for parasite transmission to mosquito vectors. Within the blood meal of a mosquito, female and male gametocytes egress host erythrocytes and transform into gametes for fertilization. Zygotes undergo meiotic nuclear division and develop into motile ookinetes to escape from the blood meal to the mosquito midgut epithelium. The elucidation of gene regulation during this lifecycle stage is important for developing antimalarial strategies that block transmission and prevent the spread of the disease.

Gametocytes develop from a subset of asexual erythrocytic parasites that are committed to sexual reproduction. The transition from the asexual erythrocytic stage to gametocytes occurs in late trophozoites; a subpopulation of trophozoites differentiates into gametocytes for sexual reproduction, and the rest develop into schizonts to continue asexual proliferation. Early gametocytes display no sex-specific features, but after a period of time, which is different among *Plasmodium* species, they start differentiating into each sex and finally become mature male and female gametocytes that are ready to transform into gametes. In *P. berghei*, early gametocytes manifest morphologies distinguishable from asexual stages at 18 hpi (hours post-infection of the new red blood cell) and manifest sex-specific features at 20 hpi. They then complete a sex-specific maturation process, during which a sex-specific gene expression repertoire is established, at approximately 24-26 hpi (1). Sex-specific proteomic and transcriptomic analyses in *P. berghei* have revealed a number of genes specifically expressed in each sex (2)(3)(4). In females, genes necessary for fertilization and protein synthesis are expressed, and the number of transcripts is stored in an untranslated state in the cytoplasm to prepare for zygote development (5)(6). In males, genes for flagella-based motility and genome replication are expressed, as well as genes for fertilization.

Recently, it has become clear that transcription factors (TFs) belonging to the AP2 family play a central role in sexual development. AP2-G is essential(7). Disruption of this gene causes the parasite to lose its ability to generate gametocytes. In *P. berghei*, its expression starts from the late trophozoite at 14-16 hpi and peaks at 18 hpi when sexual stage–specific features first appear (8). *P. berghei* AP2-G induces hundreds of genes that contain several TFs important for the development of gametocytes, suggesting that induction of these TFs by AP2-G generates a driving force for gametocytogenesis (8).

AP2-FG is a member of the *Plasmodium* AP2 family of TFs and is a target genes of AP2-G (8)(9). AP2-FG is a female-specific TF and plays an essential role in the development of female gametocytes. In *P. berghei*, *AP2-FG* is expressed from early gametocytes to mature females, and by disruption of *AP2-FG*, the development of the female gametocytes is impaired, showing immature morphologies and resulting in complete loss of capability to mediate parasite transmission to mosquitoes. This TF binds to a ten-base female-specific *cis*-acting element and regulates variety of genes include those for fertilization, meiosis, and the development of ookinetes. This broad repertoire of target genes and their essential role in female development suggest that this TF is a master regulator of female development. However, *AP2-FG*-disrupted parasites can produce ookinetes with decreased numbers and abnormal morphologies, suggesting that

additional mechanisms for activating female-specific genes are still functional upon the disruption of *AP2-FG*.

We explored novel TFs for gametocyte development in the target genes of AP2-G and identified a novel female gametocyte-specific TF. In this paper, we report that this TF, designated as PFG, regulates the gene expression in female gametocytes in concert with AP2-FG.

Results

PFG is a target gene of AP2-G and is expressed in female gametocytes

We screened for novel sequence-specific TFs among target genes that were functionally unannotated in PlasmoDB (https://plasmodb.org/plasmo/app/) using highly conserved amino acid sequences among *Plasmodium* species as a criterion. We hypothesized that the amino acid sequences of DNA-binding domains of sequence-specific TFs would have been difficult to change during evolution because even a small change in the sequence-specificity of TFs could cause catastrophic effects by global changes in gene expression. Through this screening, a gene encoding a 2709 amino acid protein with two regions highly conserved among *Plasmodium* was identified (PBANKA0902300, designated *PFG*, Fig. 1A). The two conserved regions comprised 108- and 139-amino acids, respectively, and each region showed 94% and 91% conservation between *P. berghei* and *P. falciparum* (Fig. 1B). The Blastp search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using these regions revealed that genes homologous to *PFG* exist broadly among apicomplexan parasites and also in alveolates closely related to Apicomplexa, such as *Vitrella brassicaformis* (Fig.1C)(10), whereas a large inter-region between these two regions present in all *Plasmodium* homologs was not observed in these proteins (Fig. 1A). Based on domain prediction using the SMART database (http://smart.embl-heidelberg.de/), no functionally characterized domains were identified to be significantly similar to these two regions.

To investigate the expression of *PFG*, we generated parasites expressing GFP-fused PFG (*PFG::GFP* parasites; Fig. S1A). Fluorescence analysis showed that this gene is expressed only in female gametocytes and is localized in the nucleus (Fig. 1D), suggesting its involvement in the transcriptional regulation of female gametocytes. According to sex-specific transcriptome data (3), the PFG is specifically transcribed in female gametocytes. We examined the target genes of the female-specific transcriptional activator AP2-FG and found that it had been predicted as a target of AP2-FG and harbored AP2-FG peaks with binding motifs in the upstream region (Fig. 1E). These results suggest that *PFG* was activated in two steps, that is, by AP2-G and then by AP2-FG, during female development. Temporal profiling of the expression of *PFG* using *PFG::GFP* parasites showed that the expression became visible from 20 hpi (hours post-erythrocyte infection) (Fig. 1F). The timing of the expression was approximately four hours later than that of AP2-FG (9).

PFG is essential for female development

We generated *PFG*-disrupted (*PFG*(-)) parasites to assess the role of *PFG* (Fig. S1B). The asexual stages of *PFG*(-) parasites proliferated as wild-types (Fig. 2A), and morphologically mature male and female gametocytes were observed on Giemsa-stained blood smears (Fig. 2B). However, in mosquito transmission experiments, no oocysts were observed on the mosquito midgut wall 14 days after the infective blood meal in two independent clones (Fig. 2C). In ookinete cultures, only approximately 20% of the female gametocytes were converted into zygotes, and no banana-shaped or retort-form ookinetes were generated (Fig. 2D). Considering the female-specific expression of *PFG*, these results suggest that disruption of *PFG* significantly impaired the development of female gametocytes. In the cross-fertilization experiment, ookinetes were formed when crossing with normal females (*P48/45*(-)) but not when crossing with normal males (*P47*(-)), demonstrating that the abnormal phenotype observed in *PFG*(-) was indeed derived from female gametocytes (Fig. 2E).

Next, we generated *PFG*(-) parasites from transgenic *P. berghei* ANKA 820 cl1m1cl1 (here called 820) parasites, which express red fluorescent protein (RFP) under the control of the female-specific promoter (*CCP2*) and GFP under the control of the male-specific promoter (dynein gene)

(*PFG*(-)^{820} parasites, Fig. S1B)(11) (12). Although female gametocytes with mature morphology were observed on Giemsa-stained blood smears as the original 820 parasites, in fluorescence-activated cell sorting (FACS) analysis, no RFP-positive parasites were detected in the *PFG*(-)^{820} parasites (Fig. 2F). In contrast, GFP-positive parasites were observed in them as in the original 820 parasites (Fig. 2F). These results indicated that the promoter activity of the female-specific CCP2 gene used for RFP expression in 820 parasites was severely reduced by the disruption of *PFG*.

Female-specific genes are globally downregulated in *PFG*(-) parasites

As the above results suggest that *PFG* encodes a TF that plays an important role in female-specific gene expression. To investigate the effect of *PFG* disruption on transcriptional regulation in female gametocytes, we performed RNA-seq analyses of the wild-type and *PFG*(-) parasites (Table S1). Mice were treated with phenylhydrazine prior to the passage of parasites, and after infection, asexual stages were killed by treatment with sulfadiazine in drinking water. Three independent samples were prepared from wild-type and *PFG*(-) parasites. In *PFG*(-) parasites, 279 genes were significantly downregulated ($\log_2$(fold change) < -2 and *p*-value adjusted for multiple testing with the Benjamini-Hochberg procedure < 0.001), more than 90% of which were female-specific genes according to sex-specific transcriptome data (3) (Fig. 2G). Consistent with the results of the gene targeting experiment in 820 parasites, transcripts of the CCP2 gene decreased over twenty times in *PFG*(-)parasites. Genes significantly downregulated in *PFG*(-) constituted more than half of the female-enriched genes (258/504, Table S1), suggesting that the expression of female-specific genes was globally downregulated in *PFG*(-) parasites.

PFG is co-localized with AP2-FG on ten-base motif

To examine whether PFG targeted these female-specific genes, ChIP-seq analysis was performed using *PFG::GFP* parasites and anti-GFP antibodies. In two biologically independent experiments, experiments 1 and 2, 1073 and 1,204 peaks were identified in the genome, respectively (Tables S2A and S2B), and 95% (1029 peaks) of the peaks identified in experiment 1 were common with those identified in experiment 2 (Fig. 3A). These peaks were observed upstream of the genes significantly downregulated in *PFG*(-) parasites, suggesting that PFG is directly involved in the transcriptional activation of these genes (examples of graphic images are shown in Fig. 3B). Statistical analysis of genomic sequences around the summits of peaks (common peaks, the same hereafter) showed that binding of PFG to the genome was associated with ten-base motif sequences, TGTRNNYACA (Fig. 3C), the female-specific *cis*-acting element identified in the ChIP-seq of AP2-FG (9). In comparison with the graphical views, the peaks identified in the ChIP-seq of PFG were co-localized with those of AP2-FG (Fig. 3B). An heat map positioned at the peak summit of PFG in the center showed that the position of the peaks was consistent with the ChIP-seq peaks of AP2-FG throughout the genome (Fig. 3D). These results strongly suggest that these two TFs form a complex on the ten-base motif and co-operatively activate their targets. On the other hand, graphic images also showed that some of the AP2-FG peaks lacked the corresponding ChIP-seq peaks of PFG (Fig. 3E).

PFG is essential for AP2-FG binding to ten-base motif

Some sequence-specific TFs form heterodimeric complexes that bind specific DNA sequences. To examine whether this was the case with PFG and AP2-FG, ChIP-seq analyses for each of these two TFs were performed using parasites in which the other was disrupted.

ChIP-seq of PFG in *PFG::GFP* parasites with the AP2-FG gene disrupted (*PFG::GFP*$^{AP2\text{-}FG(-)}$, Fig. S1C), 1077 and 881 peaks were obtained, and 847 peaks were common between them (Fig. 4A and Tables S2C and S2D). These peaks showed no apparent changes compared to those obtained for the original *PFG::GFP* parasites (Fig. 4B). The heat map showed that the positions of PFG peaks obtained in *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites (common peaks, the same hereafter) were consistent with those obtained for *PFG::GFP* parasites throughout the genome, and vice versa (**Fig.** 4C). Statistical analysis showed that the same ten-base motif obtained in the *PFG::GFP* parasites

was enriched around the summits of the PFG peak obtained in *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites (Fig. 4D). These results indicate that PFG binds to the ten-base motif even in the absence of AP2-FG.

ChIP-seq of AP2-FG obtained in *AP2-FG::GFP* parasites with the PFG gene disrupted (*AP2-FG::GFP*$^{PFG(-)}$, Fig. S1D), 575 and 492 peaks were obtained, and 457 peaks were common between them (Fig. 4E and Tables S2E and S2F). The graphical appearance of the AP2-FG peaks changed drastically from that of the AP2-FG peaks obtained for the AP2-FG::GFP parasites. Comparison with PFG peaks showed that upon disruption of *PFG*, AP2-FG peaks common with PFG peaks disappeared, and only AP2-FG peaks not common with PFG peaks remained (Fig. 4F). In line with this, statistical analysis showed that the ten-base motif disappeared and instead two distinct motifs, GCTCA and TGCACA, became the most enriched motifs around the summits of these peaks (common peaks, the same hereafter) with a *p*-value of 2.1×10^{-27} and 1.3×10^{-28}, respectively (Fig. 4G). These results suggest that PFG is essential for binding AP2-FG to the ten-base motif, that is, PFG mediates AP-FG binding to the ten-base motif. The results also suggest that AP2-FG binds to the genome in two distinct forms, with and without PFG, and that the former form binds to the ten-base motif, TGTRNNYACA, and the latter binds to other motifs, supposedly GCTCA or TGCACA. In the following, we refer to the form with PFG as cAP2-FG or the complex form, and the form without PFG as sAP2-FG or the single form.

Ten-base motif is essential for binding of PFG to the genome

ChIP-seq of PFG in *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites showed that PFG binds to the ten-base motif in the absence of AP2-FG. To demonstrate that the ten-base motif is essential for binding PFG to the genome, we performed a ChIP-qPCR assay using transgenic parasites in which the motif upstream of a target was mutated. A CPW-WPC family protein gene (PBANKA_1346300) (13) harbors a ChIP-seq peak for PFG and a ten-base motif under the summit of the peak. Three-point mutations were introduced into the motif using the CRISPR/Cas9 system, and *GFP* was fused to the PFG gene in these parasites (Fig. 4H). Introducing these mutations reduced the %input value to background levels (Fig. 4I), demonstrating that the motif is essential for PFG binding to the genome.

AP2-FG binds to the five-base motifs directly with its AP2 domain

The results of the ChIP-seq analysis of AP2-FG in *AP2-FG::GFP*$^{PFG(-)}$ parasites suggested that AP2-FG can bind to the genome directly through its AP2 domain. Statistical analysis showed that two distinct motifs, GCTCA and TGCACA, were enriched around the summits of ChIP-seq peaks (Fig. 4G). To investigate whether the AP2 domain of AP2-FG binds to these sequences, we performed DNA immunoprecipitation followed by high-throughput sequencing (DIP-seq) using the recombinant AP2 domain of *P. berghei* AP2-FG and *P. berghei* genomic DNA (Table S2G). Around the summits of the peaks obtained by DIP-seq analysis, the five-base motif sequence GCTCA, identical to one of motifs enriched around the summits of ChIP-seq peaks described above, was highly enriched with a *p*-value of 2.6×10^{-246} (Fig. 5A). In addition, two motifs, GATCA and ACTCA, which may be variants of the GCTCA motif, were also enriched, with *p*-value of 2.0×10^{-153} and 6.8×10^{-51}, respectively (Fig. 5A). In contrast, the six-base motif, TGCACA, was not enriched around DIP-seq peaks (*p*-value = 0.051), suggesting that it is not a binding motif of AP2-FG. Two minor motifs, GATCA and ACTCA, were also enriched around the summits of ChIP-seq peaks of sAP2-FG, with *p*-value of 2.0×10^{-53} and 1.0×10^{-13}, respectively. In the ChIP-seq of sAP2-FG, 84.2% of the peaks contained at least one of these five-base motifs within 300 bp from the summits, and the average distance from the summits was 62.4 bp (Fig. 5B). Collectively, these results suggested that AP2-FG binds directly to these motifs through its AP2 domain.

A ChIP-qPCR assay was performed to confirm that this motif is essential for the binding of sAP2-FG to the upstream of target genes. NEK2 (PBANKA_1240700) is a Nima-related protein kinase essential for parasite transmission to mosquito vectors (14). The NEK2 gene is one of the target genes of sAP2-FG, harboring one five-base motif sequence under the ChIP-seq peak. We prepared parasites with this motif mutated from *AP2-FG::GFP*$^{PFG(-)}$ parasites (Figs 5C) and performed a qPCR assay with anti-GFP antibodies. Upon addition of the mutation, the AP2-FG

associated with the locus decreased to background levels (Fig. 5D). These results demonstrated that AP2-FG binds to the five-base motif through its AP2 domain.

Five-base motif acts as a cis-activating element on the promoter

The function of the five-a base motif as *cis*-acting element in the female-specific promoter was determined by promoter assays using the centromere plasmid(15). The upstream region of the female-specific gene *P28* was used in this assay (Fig. 5E). The region harbors one ten-base motif and three five-base motifs. The mCherry gene on the centromere plasmid was expressed under the control of this promoter region, and mCherry expression was analyzed by FACS (Fig. 5F). Promoter activity was reduced by introducing mutations to the ten-base motifs, but the signals were still retained. The signals became almost undetectable by introducing mutations into the five-base motifs. These results demonstrate that the five-base motifs are female-specific *cis*-acting elements and function independently of the ten-base motif in the promoter.

Genome-wide identification of cAP2-FG and sAP2-FG targets

Our ChIP-seq analyses suggested that the targets of AP2-FG were constituted by targets of cAP2-FG and sAP2-FG. To investigate which genes were commonly or uniquely regulated by these two forms of AP2-FG, we predicted the targets of cAP2-FG and sAP2-FG separately. As targets of cAP2-FG, 810 genes were predicted from ChIP-seq peaks of PFG obtained from *PFG::GFP* parasites (common peaks of two experiments, the same hereafter) (Fig. 6A and Table 3A). These targets broadly contained genes related to female gametocyte-specific functions, including genes for female gametocyte development, such as genes for egress from the erythrocyte and fertilization, and genes for zygote/ookinete development, such as meiosis-related genes, genes for pellicular and subpellicular structures, ookinete surface or secretory protein genes, and crystalloid protein genes (Table S3A). As target genes for sAP2-FG2, 321 were predicted by ChIP-seq in *AP2-FG::GFP*$^{PFG(-)}$ parasites (Fig. 6B and Table S3B). Targets of sAP2-FG also contained many genes for female gametocyte-specific functions, but in smaller numbers than those in cAP2-FG targets. As a result of these target predictions, 967 genes were identified as targets for either cAP2-FG or sAP2-FG, and 164 genes were common between them (Fig. 6C). Based on this classification, the cAP2-FG and sAP2-FG targets had the following features: First, genes known to be involved in gametocyte development are target genes of sAP2-FG, and most of them are also targets of cAP2-FG (thus belonging to common targets). Second, genes for zygote/ookinete development are targets of cAP2-FG, and some are also targets of sAP2-FG. Among these, most meiosis-related genes and crystalloid protein genes are the target genes unique to cAP2-FG.

Target prediction also showed that *PFG* is a target of sAP2-FG and cAP2-FG. Upstream of *PFG* harbors the binding sites of three different TFs: AP2-G, sAP2-FG (AP2-FG), and cAP2-FG (PFG) (Fig. 6D). These results suggest that PFG is activated in early females by sAP2-FG and then by cAP2-FG in later female gametocytes. Meanwhile, the zygote TF AP2-Z was predicted to be a target of cAP2-FG only (16). Together with the results obtained in previous studies(8)(16), a cascade of TFs starting from AP2-G could be summarized as in Fig. 6 E.

Impact of *PFG* disruption on the expression of AP2-FG target genes

To examine the independence of transcriptional activation by cAP2-FG from that of sAP2-FG, the impact of *PFG* disruption on the expression of AP2-FG target genes was assessed separately in the three groups described in Fig 5C, that is, unique for sAP2-FG, unique for PFG, and common for both, using RNA-seq data already presented in Fig. 2G. Among the 279 genes significantly downregulated in *PFG*(-) parasites, 165 genes were targets for PFG (unique for PFG or common for sAP2-FG and PFG), and only four genes were targets unique to sAP2-FG. The distribution of $\log_2$(fold change) in the three groups of target genes showed that the average value was significantly higher in the unique targets of sAP2-FG than in those of the other two, unique targets for cAP2-FG and common targets for both, with p-value of 1.3×10^{-10} and 1.4×10^{-5}, respectively, by two-tailed Student's t-test (Fig. 6F). In addition, the average $\log_2$(fold change) value in common target genes was relatively higher than that in targets unique to PFG, suggesting that transcriptional activation by sAP2-FG partly complements the impact of PFG disruption on these common targets.

These results suggest that the trans-activity of cAP2-FG and sAP2-FG is independent of each other.

Discussion

Genes transcribed in female gametocytes are involved in various steps of parasite transmission to the mosquito vector, including sex-specific differentiation in the vertebrate host and zygote/ookinete development that take place in the mosquito vector. In a previous study, we reported that this wide range of gene expression repertoires in females is regulated directly by a female-specific TF AP2-FG that binds to a ten-base *cis*-acting element upstream of these genes (17). In this study, we report two findings that change our previous view. First, the binding of AP2-FG to the ten-base motif is mediated by another female-specific TF, PFG. Second, there is another *cis*-acting element in female-specific promoters, and AP2-FG binds to this element directly as a *trans*-activating factor. These findings suggest that transcriptional regulation of female-specific genes is not as simple as previously thought but is regulated by various promoters composed of different combinations of the two *cis*-acting elements. Considering the different expression profiles of PFG and AP2-FG, it is assumed that different expression profiles are generated among female-specific genes by the combination of these two *cis*-acting elements.

The present study suggests that PFG is upregulated in two steps in female gametocytes; first by sAP2-FG, and then by cAP2-FG. Transcriptional activation of *PFG* by cAP2-FG constitutes a transcriptional positive feedback loop, suggesting that after activation by sAP2-FG, PFG maintains its transcription through this autoactivation mechanism. As the expression of *PFG* increases by this mechanism, AP2-FG recruited by PFG would increase and cAP2-FG would become predominant in transcriptional regulation in female gametocytes. Fig. 6G illustrates this model. Considering that targetome of cAP2-FG contains many ookinete genes, transcripts for ookinete development would gradually increase in transcriptome of females. In this model it is hypothesized that binding of AP2-FG contributes to the trans-activity of cAP2-FG. This is supported by the observation that expression of many unique targets of PFG deceases in *AP2-FG*-disrupted parasites (9). However, it remains elusive how AP2-FG contributes to this activity. As discussed below, PFG may have its own trans-activity for the ten-base element. It is necessary to investigate the role of AP2-FG in cAP2-FG as a next step.

In *AP2-FG*-disrupted parasites, while female gametocytes display immature morphologies, a few develop into retort-form ookinetes. Because PFG is essential for the parasite's capacity to produce ookinetes, this phenotype suggests that PFG retains trans-activity in the absence of AP2-FG and partly promote ookinete development. The observation also supports this assumption that RFP signals in female gametocytes could still be detected in *AP2-FG*(-)^{820} parasites by FACS (9) even though RFP signals were reduced to undetectable level in *PGP*(-)^{820} parasites under the same experimental conditions. This means that the promoter of the *CCP2* gene, which is a target of cAP2-FG only, is still active in *AP2-FG*(-)^{820} parasites. This putative trans-activity of PFG could also explain why the fact the expression of *PFG* was still observed in *AP2-FG*-disrupted parasites. It is assumed that PFG continues to activate its transcription through a positive feedback loop once it is induced by AP2-G and complements the absence of AP2-FG. Based on this assumption and the model presented in Fig. 5G, phenotypic differences between *PFG*-disrupted and *AP2-FG*-disrupted parasites can be explained as follows. *PGP*(-) parasites completely lose their ability to develop into ookinetes, because the expression of genes necessary for ookinete development is significantly reduced. However, they manifest mature morphologies in females owing to the induction of the genes necessary for female development by sAP2-FG. In contrast, *AP2-FG*(-) parasites show immature morphologies in females because the expression of genes involved in female development is downregulated in the early stage. Despite this immature morphology, they retain the ability to fertilize and develop into retort-form ookinetes because PFG is still expressed in *AP2-FG*(-) and activates genes for zygote/ookinete development.

In conclusion, our results suggest that the two forms of AP2-FG, which correspond to distinct female-specific *cis*-acting elements, play pivotal roles in female gametocyte development and can generate variations among female-specific promoters. The present findings provide the possibility to predict the expression profile of each gene by analyzing its promoter sequence and

further estimating its functions according to this regulation. In the next step, it will be necessary to compare the properties of the number of actual promoter activities *in vivo* and improve the model presented in this study. Such studies deepen our understanding of the sexual stage, including gametocyte development in erythrocytes, fertilization, meiosis, and development into ookinetes, and provide clues to interrupt parasite transmission from humans to mosquitoes.

Materials and Methods

Parasite preparations

The ANKA strain of *P. berghei* was maintained in female BALB/c mice (6–10 weeks of age). To examine the number of oocysts, infected mice were exposed to *Anopheles stephensi* mosquitoes. Fully engorged mosquitoes were selected and maintained at 20 °C. The number of oocysts was evaluated 14 days after the infective blood meal. To prepare mature schizonts, infected mouse blood was cultured in the medium for 16 h. Mature schizonts were purified using Nicoprep density gradient. To prepare gametocyte-enriched blood, mice were pretreated with hydrazine and infected with *P. berghei* by intra-abdominal injection of infected blood. After parasitemia increased to over 1%, they were treated with sulfadiazine for two days in drinking water (20 mg/L) to kill asexual parasites. After checking the exflagellation rate, whole blood was collected.

Preparation for transgenic parasites

Parasite transfection was performed as previously described (9). Briefly, cultured mature schizonts were purified using a density gradient of Nicoprep, transfected with DNA constructs by electropoletion, and injected intravenously into mice. These mice were treated with pyrimethamine to select the parasite integrated with the construct by double crossover homologous recombination. The parasites were cloned by limiting dilution to obtain transgenic parasite clones. To prepare the targeting construct, DNA fragments for recombination were annealed to each side of the DNA fragments containing *human DHFR* as a selectable marker gene by overlapping PCR. Constructs for parasites expressing the GFP-fused genes were prepared as follows. Briefly, DNA fragments corresponding to the 3´ portion of the gene were subcloned into the plasmid vector containing the GFP gene, the 3´-terminal portion of the *P. berghei* HSP70 gene, and a selectable marker cassette for expressing the *human DHFR* gene. The plasmid was separated from the vector backbone by digestion with restriction enzymes Not1 and BanH1.
Transgenic parasites with mutated motif sequences were prepared using the CRISPR/Cas9 method, as reported previously (18). Briefly, cultured mature merozoites of CAS9-expressing *P. berghei* parasites were transfected with a gRNA plasmid vector and a linear DNA template containing mutations in the motif sequences and injected intravenously into mice. These mice were treated with pyrimethamine for two days to deplete parasites in which the locus did not change. The parasites were cloned by limiting dilution. Mutations in these motifs were confirmed by direct genome sequencing.

Flow cytometric analysis

Flow cytometric analysis was performed using an LSR-II flow cytometer (BD Biosciences) as described previously (9). Briefly, in experiments using 820 parasites, the tail blood from an infected mouse was selected by gating forward, and side scattering, and then the gated population was analyzed for fluorescence of GFP and RFP (575/26 and 530/30, respectively). In the promoter assay using parasites transfected with a centromere plasmid, the tail blood from an infected mouse was selected by gating with side-scatter and staining with Hoechst 33342, followed by GFP fluorescence. The gated population was analyzed for mCherry fluorescence (575/26). Analysis was performed using the DIVER program (BD Biosciences).

ChIP-Seq

Two infected mice were used in each ChIP-seq experiment. ChIP-seq experiments were performed as previously described (19). The gametocyte-enriched blood was filtered using a Plasmodipur filter to remove white blood cells and fixed in 1% paraformaldehyde for 30 min with swirling. Red

blood cells were lysed in 0.84% NH_4Cl, and residual cells were lysed in a lysis buffer containing 1% SDS. The lysate was sonicated in Bioruptor 2. After centrifugation, the supernatant was diluted with dilution buffer, and chromatin was immunoprecipitated with Dynabeads protein A coated with an anti-GFP antibody. DNA fragments were recovered from the precipitated chromatin and used for library preparation for sequencing. The library was prepared using a Hyper Prep Kit (KAPA Biosystems). Sequencing was performed on an Illumina NextSeq sequencer.

Analysis of ChIP-seq data

Sequence data were mapped onto the *P. berghei* genome sequence (PlasmoDB, version 3) using BOWTIE2 software with default settings. Mapping data were analyzed using the MACS2 peak-calling algorithm. Conditions for peak calling were FDR < 0.01 and fold enrichment > 2 or 3. To detect large peaks constituted from multiple peaks, the option “Callsummit” was used. Mapped read data were visualized using IGV software. Sequences concentrated around the predicted summits of ChIP-seq peaks were investigated with Fisher’s exact test carried out between 200-bp regions that had summits in the center and 200-bp regions excised from the genome, excluding the former regions, to cover the entire genome sequence. Sequences with the least p-values were combined with the common sequence motifs. Genes were determined as targets when their 1.2-kbp upstream regions harbored the predicted summits of ChIP-seq peaks. When the upstream intergenic region was less than 1.2 kbp, the entire intergenic region was used for target prediction. Two biologically independent ChIP-seq experiments were performed. The ChIP-seq data were deposited in the Gene Expression Omnibus (GEO) with the accession number GSE226028. .

DIP-seq

The recombinant AP2 domain of AP2-FG was prepared as a GST-fused protein using the same procedure as reported previously(8). DIP was performed as reported previously(8). Briefly, recombinant GST-fused AP2 domain of AP2-FG and genomic DNA fragments were incubated at room temperature for 30 min in 100 µL of binding/washing buffer. The mixture was then incubated with 50 µL glutathione sepharose resin (Cytiva) at room temperature with rotation. Resin was washed three times with 150 µL of binding/washing buffer, and bound protein was eluted with 10 mM glutathione solution. Recovered DNA (5 ng) was sequenced using the same procedures as in ChIP-seq. Input genomic DNA fragments were also sequenced as a control. Analysis of the sequence data was performed using the same procedures as in ChIP-seq. Data were deposited in the GEO database under accession number GSE226028.

ChIP-qPCR

ChIP for ChIP-qPCR was performed using gametocyte-enriched mouse blood, using the same procedures as in ChIP-seq. One infected mouse was used in each ChIP experiment. qPCR was performed using the TB Green Fast qPCR Mix (Takara Bio) according to the manufacturer’s protocol. Three independent ChIP experiments were performed, and the percentage of input DNA (DNA extracted from lysate before ChIP) was compared between *P. berghei* parasites expressing GFP-fused TF and those with the mutated motifs. The primers used for qPCR are listed in Table S4.

RNA-seq analysis

Gametocyte-enriched blood was passed through a filter and subjected to erythrocyte lysis in 0.84% NH_4Cl. According to the manufacturer's protocols, total RNA was extracted from residual cells using Isogen II (Nippon Gene). Libraries were prepared using the RNA Hyperprep kit (KAPA Biosystems) and sequencing was performed using an Illumina NextSeq sequencer.

Analysis of RNA-Seq data

Three independent experiments were performed using *PFG*-disrupted parasites. Read data were counted using FeatureCount software. RPKM (Reads Per Kilobase of exon per million mapped reads) was calculated for each gene, and only genes with an RPKM maximum > 20 in wild-type parasites were used for the following analyses. Additionally, genes located in the sub telomeric

regions were excluded because of their variable expression among the clones. The ratio of read numbers and padj was calculated for each gene between wild-type and *gSNF2*-disrupted parasites using the DEseq2 software. Volcano plots of the obtained results were created using DDplot2 software. The RNA-seq data were deposited in GEO with the accession number GSE226028. The RNA-seq data deposited in GEO with accession number GSE198588 were used as data for the wild-type parasites.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 17H01542 to MY.

References

1. Mons B, Janse CJ, Boorsma EG, Van Der Kaay HJ. 1985. Synchronized erythrocytic schizogony and gametocytogenesis of Plasmodium berghei in vivo and in vitro. Parasitology 91:423–430.

2. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M, Waters AP. 2005. Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. Cell 121:675–687.

3. Witmer K, Fraschka SA, Vlachou D, Bártfai R, Christophides GK. 2020. An epigenetic map of malaria parasite development from host to vector. Sci Rep 10:1–19.

4. Yeoh LM, Goodman CD, Mollard V, McFadden GI, Ralph SA. 2017. Comparative transcriptomics of female and male gametocytes in Plasmodium berghei and the evolution of sex in alveolates. BMC Genomics 18.

5. Mair GR, Braks JAM, Garver LS, Wiegant LCAG, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP. 2006. Regulation of sexual development of Plasmodium by translational repression. Science (80-) 313:667–669.

6. Guerreiro A, Deligianni E, Santos JM, Silva PAGC, Louis C, Pain A, Janse CJ, Franke-Fayard B, Carret CK, Siden-Kiamos I, Mair GR. 2014. Genome-wide RIP-Chip analysis of translational repressor-bound mRNAs in the Plasmodium gametocyte. Genome Biol 15:493.

7. Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa A a, Bushell E, Graham AL, Cameron R, Kafsack BFC, Williams AE, Llinás M, Berriman M, Billker O, Waters AP. 2014. A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. Nature 507:253–7.

8. Yuda M, Kaneko I, Murata Y, Iwanaga S, Nishi T. 2021. Mechanisms of triggering malaria gametocytogenesis by AP2-G. Parasitol Int 84:102403.

9. Yuda M, Kaneko I, Iwanaga S, Murata Y, Kato T. 2019. Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. Mol Microbiol https://doi.org/10.1111/mmi.14334.

10. Oborník M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš J. 2012. Morphology, Ultrastructure and Life Cycle of Vitrella

brassicaformis n. sp., n. gen., a Novel Chromerid from the Great Barrier Reef. Protist 163:306–323.

11. Raine JD, Ecker A, Mendoza J, Tewari R, Stanway RR, Sinden RE. 2007. Female inheritance of Malarial lap genes is essential for mosquito transmission. PLoS Pathog 3.

12. Mair GR, Lasonder E, Garver LS, Franke-Fayard BMD, Carret CK, Wiegant JCAG, Dirks RW, Dimopoulos G, Janse CJ, Waters AP. 2010. Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development. PLoS Pathog 6.

13. Rao PN, Santos JM, Pain A, Templeton TJ, Mair GR. 2016. Translational repression of the cpw-wpc gene family in the malaria parasite Plasmodium. Parasitol Int 65:463–471.

14. Reininger L, Tewari R, Fennell C, Holland Z, Goldring D, Ranford-Cartwright L, Billker O, Doerig C. 2009. An essential role for the Plasmodium Nek-2 Nima-related protein kinase in the sexual development of malaria parasites. J Biol Chem 284:20858–20868.

15. Iwanaga S, Khan SM, Kaneko I, Christodoulou Z, Newbold C, Yuda M, Janse CJ, Waters AP. 2010. Functional Identification of the Plasmodium Centromere and Generation of a Plasmodium Artificial Chromosome. Cell Host Microbe 7:245–255.

16. Nishi T, Kaneko I, Yuda M. 2022. Identification of a novel AP2 transcription factor in zygotes with an essential role in Plasmodium ookinete development 1–24.

17. Yuda M, Kaneko I, Iwanaga S, Murata Y, Kato T. 2019. Female -specific gene regulation in malaria parasites by an AP2 -family transcription factor. Mol Microbiol https://doi.org/10.1111/mmi.14334.

18. Shinzawa N, Nishi T, Hiyoshi F, Motooka D, Yuda M, Iwanaga S. 2020. Improvement of CRISPR/Cas9 system by transfecting Cas9-expressing Plasmodium berghei with linear donor template. Commun Biol 3:1–13.

19. Kaneko I, Iwanaga S, Kato T, Kobayashi I, Yuda M. 2015. Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor. PLoS Pathog 11:e1004905.

20. Tremp AZ, Carter V, Saeed S, Dessens JT. 2013. Morphogenesis of Plasmodium zoites is uncoupled from tensile strength. Mol Microbiol 89:552–564.

21. Bergman LW, Kaiser K, Fujioka H, Coppens I, Daly TM, Fox S, Matuschewski K, Nussenweig V, Kappe SHI. 2003. Myosin A tail domain interacting protein (MTIP) localizes to the inner membrane complex of Plasmodium sporozoites. J Cell Sci 116:39–49.

22. Kariu T, Ishino T, Yano K, Chinzei Y, Yuda M. 2006. CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. Mol Microbiol 59:1369–1379.

23. Andreadaki M, Pace T, Grasso F, Siden-Kiamos I, Mochi S, Picci L, Bertuccini L, Ponzi M, Currà C. 2020. Plasmodium berghei Gamete Egress Protein is required for fertility of both genders. Microbiologyopen 9:1–14.

24. Talman AM, Lacroix C, Marques SR, Blagborough AM, Carzaniga R, M??nard R, Sinden RE. 2011. PbGEST mediates malaria transmission to both mosquito and vertebrate host. Mol Microbiol 82:462–474.

25. Olivieri A, Bertuccini L, Deligianni E, Franke-Fayard B, Currà C, Siden-Kiamos I, Hanssen E, Grasso F, Superti F, Pace T, Fratini F, Janse CJ, Ponzi M. 2015. Distinct properties of the egress-related osmiophilic bodies in male and female gametocytes of the rodent malaria parasite Plasmodium berghei. Cell Microbiol 17:355–368.

26. Nishi T, Kaneko I, Iwanaga S, Yuda M. 2023. PbAP2-FG2 and PbAP2R-2 function together as a transcriptional repressor complex essential for Plasmodium female development. PLoS Pathog 19:1–28.

27. Ecker A, Bushell ESC, Tewari R, Sinden RE. 2008. Reverse genetics screen identifies six proteins important for malaria development in the mosquito. Mol Microbiol 70:209–220.

28. Ukegbu C V., Christophides GK, Vlachou D. 2021. Identification of Three Novel Plasmodium Factors Involved in Ookinete to Oocyst Developmental Transition. Front Cell Infect Microbiol 11:1–17.

29. Sansam CL, Pezza RJ. 2015. Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. FEBS J 282:2444–57.

30. Pezza RJ, Voloshin ON, Vanevski F, Camerini-Otero RD. 2007. Hop2/Mnd1 acts on two critical steps in Dmc1-promoted homologous pairing. Genes Dev 21:1758–1766.

31. Jenwithisuk R, Kangwanrangsan N, Tachibana M, Thongkukiatkul A, Otsuki H, Sattabongkot J, Tsuboi T, Torii M, Ishino T. 2018. Identification of a PH domain-containing protein which is localized to crystalloid bodies of Plasmodium ookinetes. Malar J 17:1–15.

32. Kangwanrangsan N, Tachibana M, Jenwithisuk R, Tsuboi T, Riengrojpitak S, Torii M, Ishino T. 2013. A member of the CPW-WPC protein family is expressed in and localized to the surface of developing ookinetes. Malar J 12:129.

33. Dessens JT, Saeed S, Tremp AZ, Carter V. 2011. Malaria crystalloids: Specialized structures for parasite transmission? Trends Parasitol 27:106–110.

34. Tremp AZ, Saeed S, Sharma V, Lasonder E, Dessens JT. 2020. Plasmodium berghei LAPs form an extended protein complex that facilitates crystalloid targeting and biogenesis. J Proteomics 227:103925.

35. Santos JM, Duarte N, Kehrer J, Ramesar J, Avramut MC, Koster AJ, Dessens JT, Frischknecht F, Chevalley-Maurel S, Janse CJ, Franke-Fayard B, Mair GR. 2016. Maternally supplied S-acyl-transferase is required for crystalloid organelle formation and transmission of the malaria parasite. Proc Natl Acad Sci U S A 113:7183–7188.

36. Santos JM, Kehrer J, Franke-Fayard B, Frischknecht F, Janse CJ, Mair GR. 2015. The Plasmodium palmitoyl-S-acyl-transferase DHHC2 is essential for ookinete morphogenesis and malaria transmission. Sci Rep 5:1–10.

37. Wetzel J, Herrmann S, Swapna LS, Prusty D, Peter ATJ, Kono M, Saini S, Nellimarla S, Wong TWY, Wilcke L, Ramsay O, Cabrera A, Biller L, Heincke D, Mossman K, Spielmann T, Ungermann C, Parkinson J, Gilberger TW. 2015. The role of palmitoylation for protein

recruitment to the inner membrane complex of the malaria parasite. J Biol Chem 290:1712–1728.

38. Saeed S, Tremp AZ, Sharma V, Lasonder E, Dessens JT. 2020. NAD (P) transhydrogenase has vital non - mitochondrial functions in malaria parasite transmission . EMBO Rep 21:1–11.

Figures and Tables

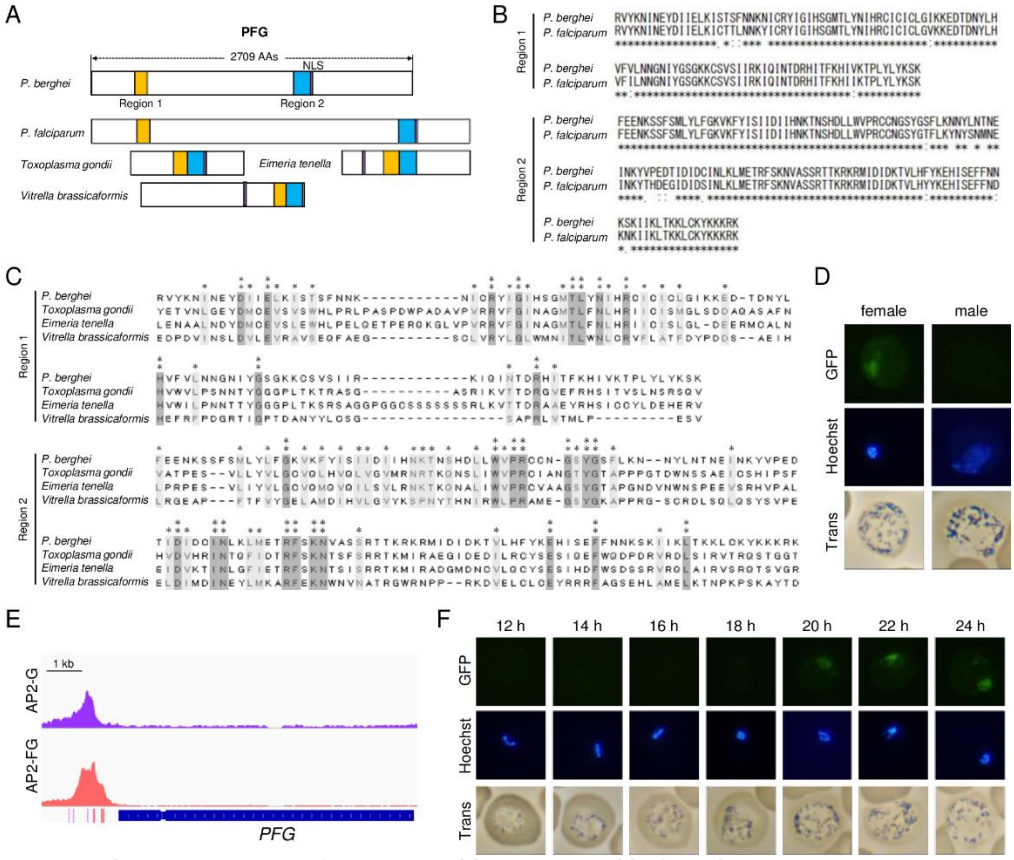

Figure 1. PFG is a target gene of AP2-G and is expressed in female gametocytes

A. Schematic diagram of PFG from *P, berghei* and its homologs in apicomplexan parasites. Regions homologous to Region 1 and Region 2, which are highly conserved among *Plasmodium* species, are shown as yellow and blue rectangles, respectively, and nuclear localization signals, which were predicted using the cNLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi).
B. Alignment of amino acid sequences of Region 1 and Region 2 from *P. berghei* and *P. falciparum* PFGs was performed using the ClustalW program (https://www.genome.jp/tools-bin/clustalw).
C. Alignment of the amino acid sequences of Region 1 and Region 2 from *P. berghei* PFG and its homologs from other apicomplexan parasites in (A). Positions at which all sequences have identical amino acids are indicated by two asterisks, and positions with amino acid residues of the same properties are indicated by one asterisk.
D. Expression of *PFG* in mature male and female gametocytes of *PFG::GFP* parasites. Nuclei were stained with Hoechst 33342.
E. Integrative Genomics Viewer (IGV) images from the ChIP-seq data of AP2-G and AP2-FG in the upstream region of *PFG*. Data were obtained from previous papers (8)(9). The purple and red bars indicate the binding motifs of AP2-G and AP2-FG in the peak region.
F. Time-course observations of *PFG* expression during gametocyte development using *PFG::GFP* parasites. Parasites were observed every two hours from 12 hpi onwards. Nuclei were stained with Hoechst 33342.

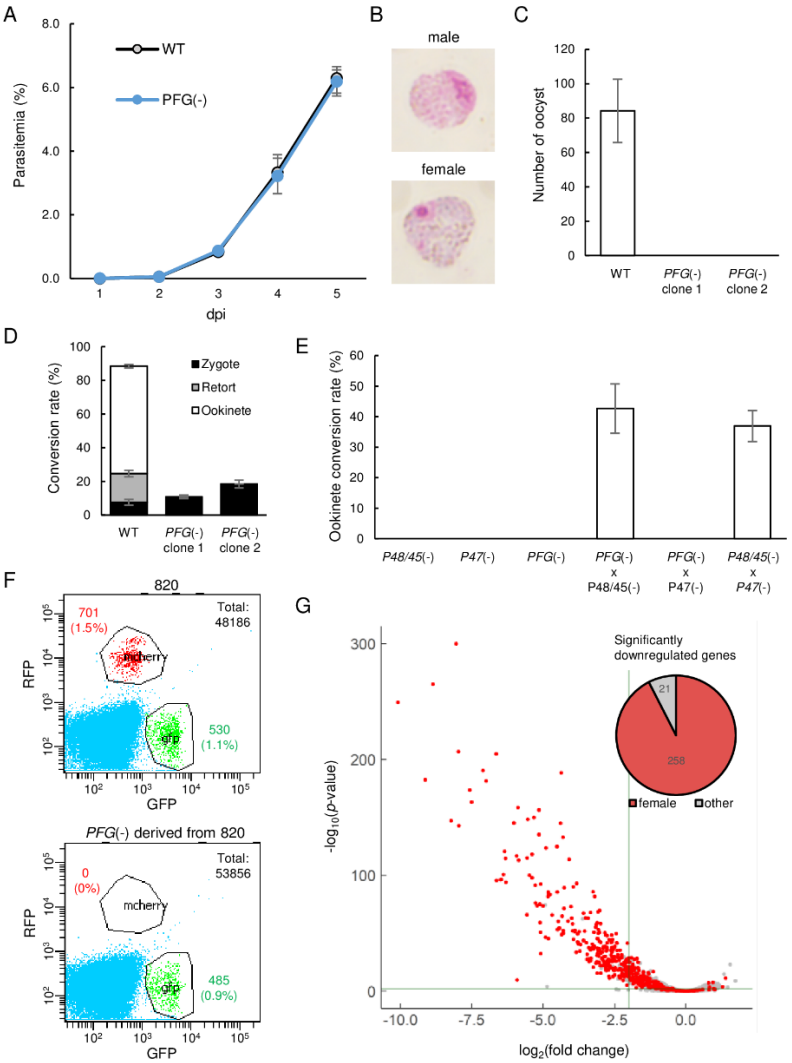

Figure 2. Development of female gametocytes was impaired in *PFG*(-)parasites.

A. Proliferation of blood-stage parasites in wild-type and *PFG*(−) parasites. Parasitemia was calculated by counting infected red blood cells using Giemsa staining of blood smears.
B. Giemsa-stained images of mature male and female gametocytes in *PFG*(-) parasites.
C. Number of oocysts in the mosquito midgut 14 days after infective blood meal of wild-type and *PFG*(-) parasites.
D. Ookinete cultures of wild type and *PFG*(-) parasites. Ratios of zygotes, retort-form, and banana-shaped ookinetes to all female-derived cells are shown in black, grey, and white, respectively.
E. A cross-fertilization experiment was performed between *PFG*(-), *P48/45*(-) (males are infertile), and *P47*(-) (females are infertile) parasites. The rate of conversion of females to mature ookinetes is shown.
F. Flow cytometry analysis of 820 parasites and *PFG*(-) parasites derived from 820 parasites.
G. Volcano plot showing differential expression of genes in *PFG*(-) compared to wild-type parasites. Red dots represent female-enriched genes, and horizontal and vertical lines indicate *p*-value of 0.001 and $\log_2$(fold change) of -2, respectively. A pie graph on the top right of the plot area shows the number of female-enriched genes among the significantly downregulated genes.

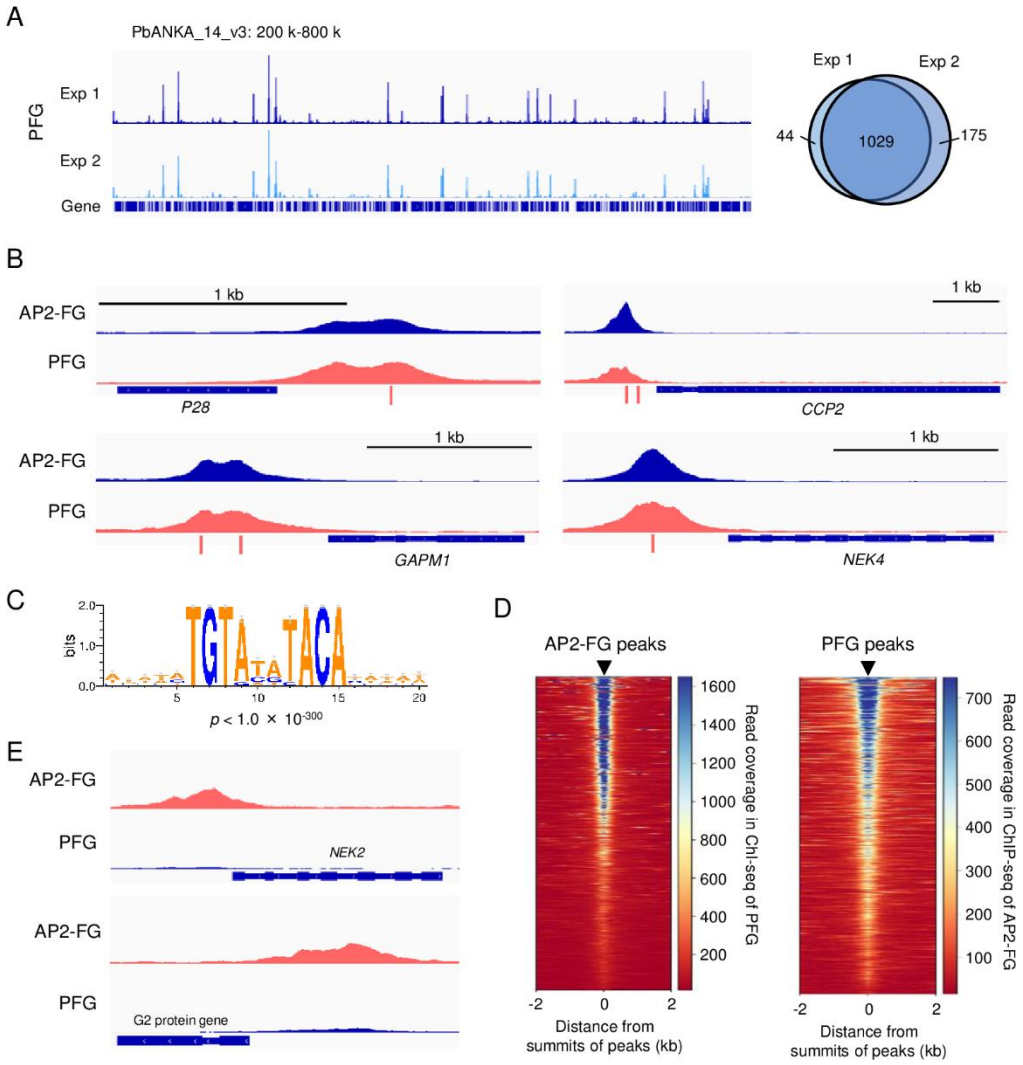

Figure 3. PFG is co-localized with AP2-FG

A. IGV images from ChIP-seq experiments 1 and 2 of PFG on chromosome 14. Histograms show row read coverage in the ChIP data at each base. The Venn diagram on the right shows the number of peaks common in ChIP-seq experiments 1 and 2.
B. Representative peak images of PFG and AP2-FG upstream of genes that were significantly downregulated in *PFG*(-) parasites. Positions of ten-base motifs are indicated by red bars. ChIP-seq data for AP2-FG were obtained from a previous paper (17).
C. The ten-base motif was enriched around the PFG peaks. Sequence logos were constructed using WebLogo 3 (http://weblogo.threeplusone.com/create.cgi).
D. Heat maps showing coverage in ChIP-seq of PFG (left) and AP-FG (right) with positioning summits of PFG peaks at the center.
E. IGV images of ChIP-seq peaks of AP2-FG that lacked the corresponding PFG peak. Histograms show the row read coverage in the ChIP data at each base.

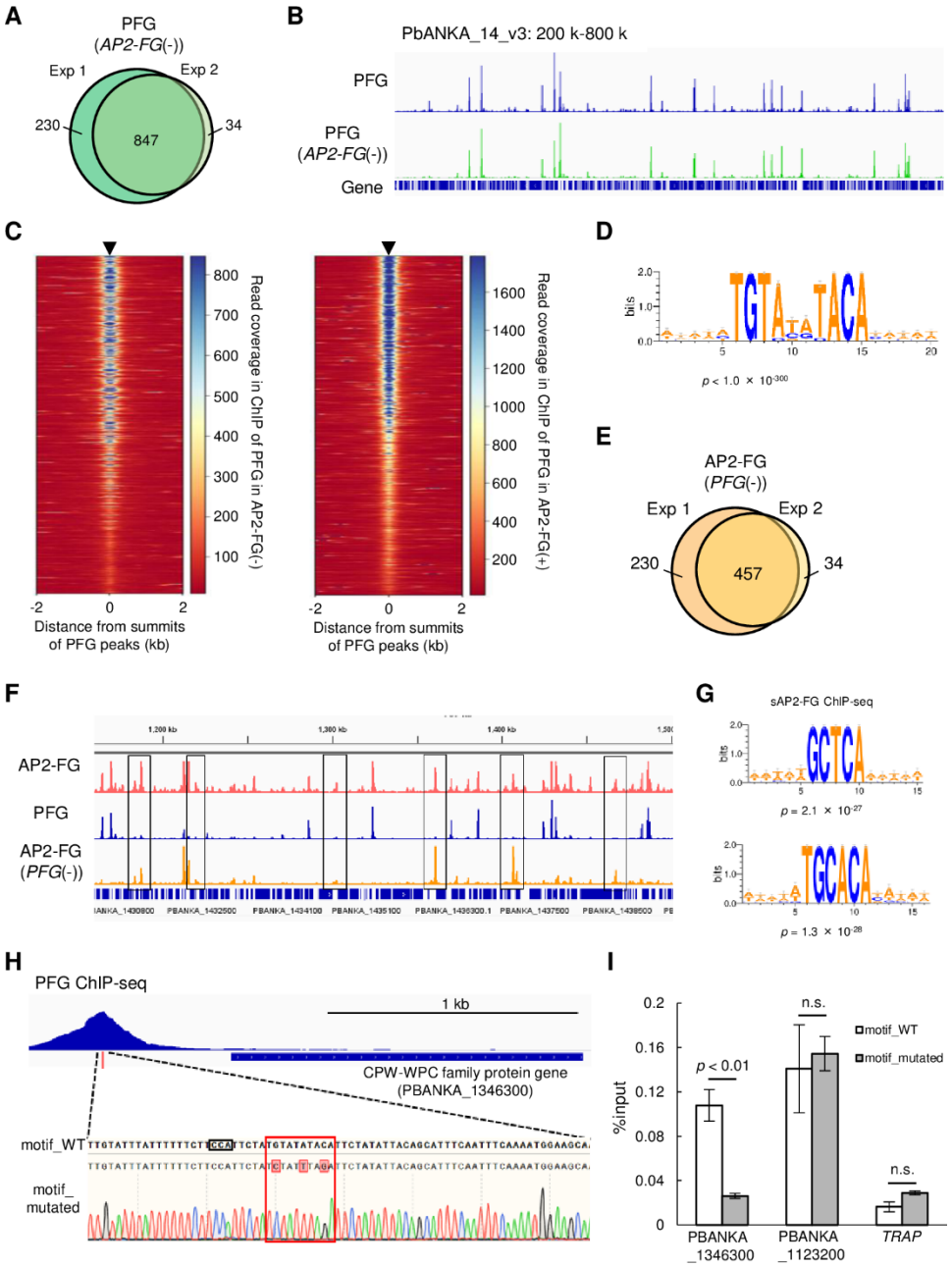

Figure 4. PFG mediates AP2-FG binding to the ten-base motif

A. Venn diagram showing the number of common peaks between experiments 1 and 2 in ChIP-seq of PFG using *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites.

B. IGV images from the ChIP-seq peaks of PFG on part of chromosome 14. The top panel shows the ChIP peaks using *PFG::GFP* parasites, and the bottom panel shows those using

PFG::GFP$^{AP2\text{-}FG(-)}$ parasites. Histograms show row read coverage in the ChIP data at each base.

C. Heat maps showing coverage in ChIP-seq of PFG using *PFG::GFP* parasites (left) and PFG using *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites, with positioning summits of peaks obtained in the other ChIP-seq at the center.

D. Motifs enriched around the summits of peaks identified in ChIP-seq of PFG using *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites. The sequence logo was depicted using WebLogo 3.

E. Venn diagram showing the number of common peaks between experiments 1 and 2 in ChIP-seq of AP2-FG using *AP2-FG::GFP*$^{PFG(-)}$ parasites.

F. IGV images of ChIP-seq peaks of AP2-FG and PFG in a region on which several AP2-FG peaks lack the corresponding PFG peak (chromosome 14:1,150 kb-1,500 kb). The top panel shows ChIP peaks of AP2-FG using *AP2-FG::GFP*. The middle and bottom panels show the ChIP-seq peaks of PFG using *PFG::GFP* and the ChIP peaks of AP2-FG using *AP2-FG::GFP*$^{PFG(-)}$ parasites, respectively. The AP2-FG peaks lacking their corresponding PFG peaks are highlighted by rectangles. Histograms show row read coverage in the ChIP data at each base.

G. Motifs enriched around the summits of peaks identified in the ChIP-seq of AP2-FG using *AP2-FG::GFP*$^{PFG(-)}$ parasites. Sequence logos were depicted using WebLogo 3.

H. Peak image of PFG and the ten-base motif under the summit upstream of a CPW-WPC family protein gene (PBANKA_1346300). The motif was mutated using the CRISPR/Cas9 system, and the mutation was confirmed by Sanger sequencing (the lowest panel; mutated nucleic acid residues are highlighted). The protospacer adjacent motif (PAM) sequence used for targeting is also highlighted by a rectangle.

I. ChIP-qPCR analysis of PFG in the upstream region of PBANKA_1346300. Gray and white bars indicate the results using wild-type and motif-mutated parasites, respectively. Three independent biological experiments were performed. The results are shown as a percentage input. Error bars indicate standard error. Experiments with another CPW-WPC family protein gene (PBANKA_1123200) and *TRAP* were performed as positive and negative controls, respectively.

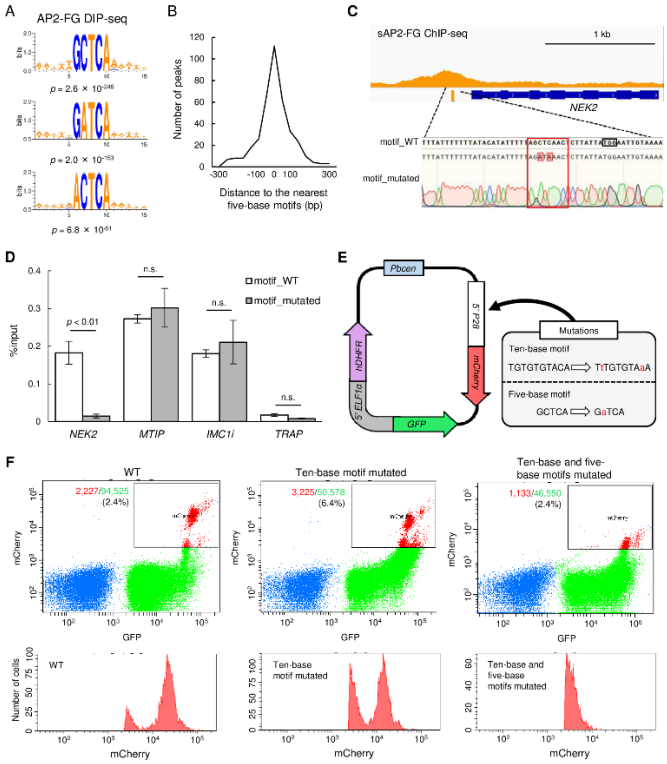

Figure 5. sAP2-FG binds to five-base motifs that function as a *cis*-acting element independent of the ten-base motif.

A. Motifs enriched around peaks identified in DIP-seq of the AP2 domain of AP2-FG. Sequence logos were depicted using WebLogo 3.
B. Histogram showing the distance between the peak summit and the nearest five-base motifs in (A) for each ChIP peak of sAP2-FG.
C. Peak image of sAP2-FG upstream of *NEK2* and the five-base motif under the summit. The motif was mutated, and the sequence was confirmed by Sanger sequencing (lowest panel). The protospacer adjacent motif (PAM) sequence used for targeting is highlighted by a rectangle. Mutated nucleic acid residues have also been highlighted.
D. ChIP-qPCR analysis of sAP2-FG in the upstream region of *NEK2*. Gray and white bars indicate the results obtained for wild-type and motif-mutated parasites, respectively. The results are shown as mean %input values from three biologically independent experiments. Error bars indicate standard error. Experiments with *MTIP* and *IMC1i* were performed as positive controls, and that in *TRAP* as a negative control.
E. Schematic diagram of a *Plasmodium* centromere plasmid used to assessing promoter activity of *P28*. Mutations introduced into the ten-base and five-base motifs of the *P28* promoter are described in a grey box. Pbcen, a sequence of the *P. berghei* chromosome 5 centromere, 5'-ELF1α, a bidirectional promoter of the elongation 1α gene for conferring constitutive expression
F. FACS analysis of parasites harboring the *P28*-reporter plasmid (E). Parasites were gated on forward-scatter and staining with Hoechst 33342, and then on GFP. The percentages of mCherry-positive parasites in all gated cells are shown in the left panel. The histogram on the right shows the number of mCherry-positive parasites at different signal strengths.

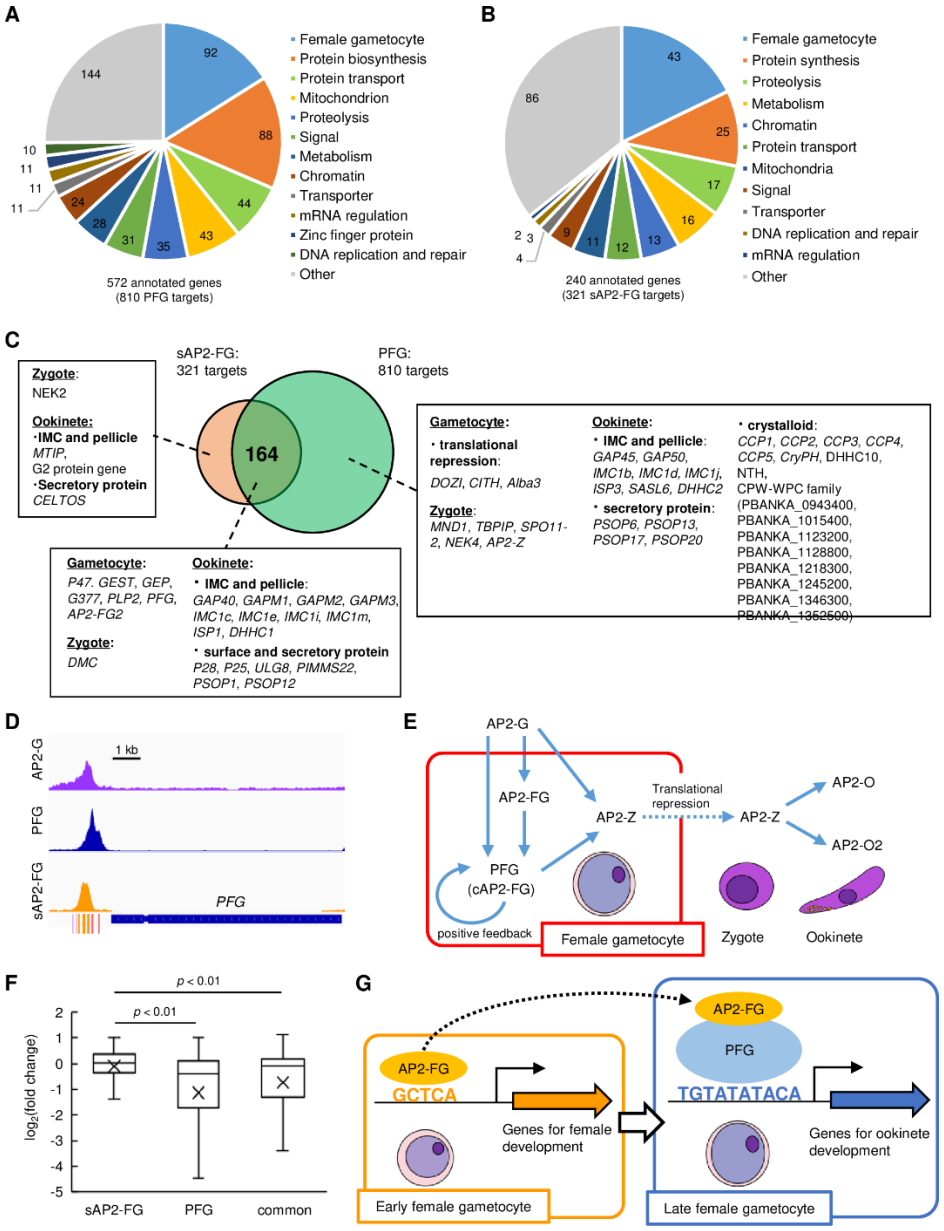

Figure 6. Target genes of PFG (cAP2-FG) and sAP2-FG contain different sets of genes.

A. Functional classification of the PFG target genes (cAP2-FG) (572 annotated genes in total). In this graph, different subgroups related to female gametocyte-specific functions are collectively shown as a group 'female-specific functions' (see also Table S3A). Hypothetical protein genes were not included. The number of members in each group is shown in the chart.

B. Functional classification of the target genes of sAP2-FG (240 annotated genes in total). Hypothetical protein genes were not included.

C. Venn diagram showing an overlap between the target genes of sAP2-FG and PFG (cAP2-FG). The target genes in 'female-specific functions' are written according to their belongings: unique to sAP2-FG, common, and unique to cAP2-FG. The assignment of these genes into

each subgroup was based on functional annotation in *Plasmo*DB and the following references(14)(20)(21)(22)(23)(24)(25)(26)(27)(13)(28)(29)(16)(30)(31)(32)(33)(34)(35)(36)(37)(38).

D. IGV images from the ChIP-seq data of AP2-G, PFG, and sAP2-FG in the upstream region of *PFG*. The data of the ChIP-seq peaks for AP2-G were obtained from a previous study (8). The peak region's binding motifs of AP2-G, ten-base and five-base motifs are indicated by purple, red, and orange bars.

E. A putative cascade of TFs starting from AP2-G is suggested in this and previous studies.

F. Box-and-whisker plots showing $\log_2$(fold change) for target genes unique to sAP2-FG and PFG (cAP2-FG) and common for both. Cross marks in the boxes indicate the average values.

G. A model of transcriptional regulation during female development. Female-specific genes harbor either or both of the female-specific *cis*-activating elements, 5-base and ten-base elements in the upstream. In early female gametocytes, AP2-FG binds to a five-base *cis*-activating element via its AP2 domain and activates genes for gametocyte development and some genes for zygote/ookinete development. In the later stage, when PFG is highly expressed, AP2-FG is predominantly recruited to the PFG on the ten-base *cis*-activating element, and the AP2-FG/PFG complex activates a full repertoire of genes for zygote/ookinete development.

Supporting information

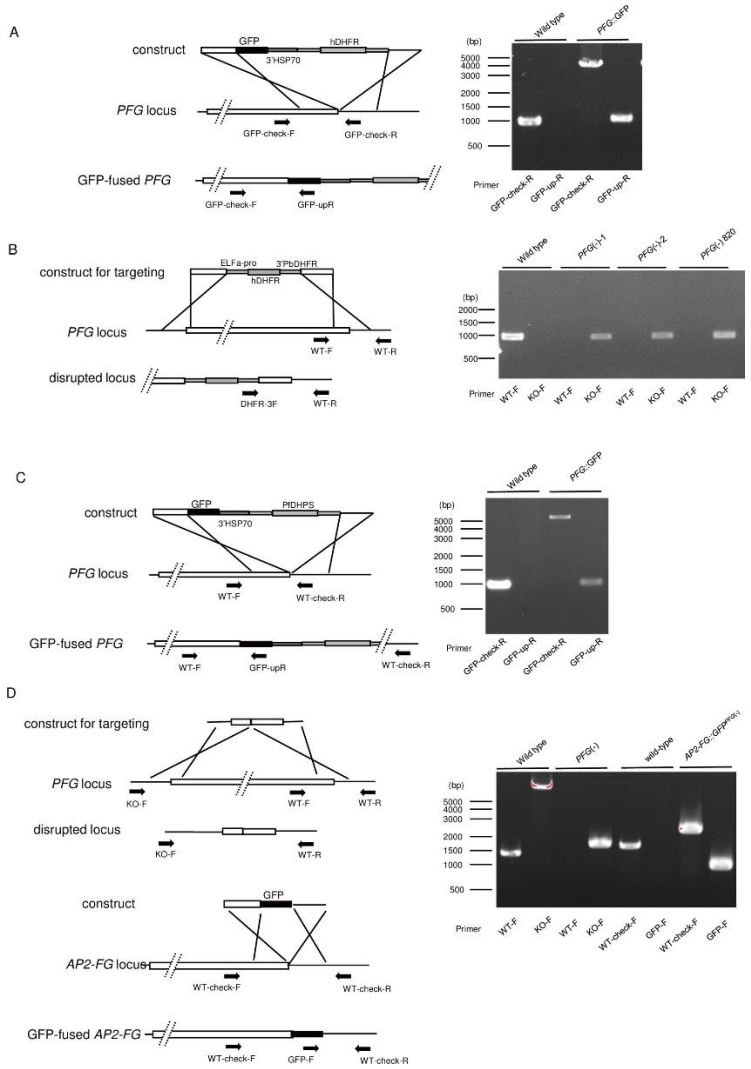

Fig. S1. Preparation of transgenic parasites.

A. Preparation of *PFG::GFP* parasites.

B. Preparation of *PFG*(-) and *PFG*(-)^{820} parasites.

C. Preparation of *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites. *PFG::GFP*$^{AP2\text{-}FG(-)}$ parasites were prepared from *AP2-FG*(-) parasites using the sulfadiazine-resistant dihydropteroate synthase (DHPS) gene as a selectable marker gene.

D. Preparation of *AP2-FG::GFP*$^{PFG(-)}$ parasites. *AP2-FG::GFP*$^{PFG(-)}$ parasites were prepared using CRISPR/Cas9 in two steps. First, *PFG* was disrupted in pbcas9, and then *GFP* was fused to the 3'-portion of the AP2-FG gene.

Captions of supplementary tables

Table S1 (separate file). Differential expression analysis between wild-type and PFG(-) parasites.

Table S2 (separate file). Peaks identified in ChIP-seq experiments.

Table S2A. Peaks identified in ChIP-seq of PFG (experiment 1).

Table S2B. Peaks identified in ChIP-seq of PFG (experiment 2).

Table S2C. Peaks identified in ChIP-seq of PFG using *AP2-FG*(-) parasites (experiment 1)

Table S2D. Peaks identified in ChIP-seq of PFG using AP2-FG(-) parasites (experiment 2).

Table S2E.Peaks identified in ChIP-seq of sAP2-FG (experiment 1).

Table S2F. Peaks identified in ChIP-seq of sAP2-FG (experiment 2).

Table S2G. Peaks identified in DIP-seq using recombinant AP2 domain of AP2-FG and P. berghei genomic DNA.

Table S3 (separate file). Target genes predicted from ChIP-seq peaks.

Table S3A. PFG target genes predicted from ChIP-seq peaks.

Table S3B. sAP2-FG target genes predicted from ChIP-seq peaks.

Table S4. List of primers used in this study.