Cell Host & Microbe
Resource

Functional Identification of the *Plasmodium* Centromere and Generation of a *Plasmodium* Artificial Chromosome

Shiroh Iwanaga,[1,*] Shahid M. Khan,[2] Izumi Kaneko,[1] Zoe Christodoulou,[3] Chris Newbold,[3] Masao Yuda,[1] Chris J. Janse,[2] and Andrew P. Waters[2,4,*]
[1]Mie University, School of Medicine, Tsu 514-0001, Japan
[2]Leiden Malaria Research Group, Centre of Infectious Diseases, Leiden University Medical Centre, Leiden 2333 ZA, The Netherlands
[3]Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK
[4]Present address: Division of Infection and Immunity, Institute of Biomedical Life Sciences and Wellcome Centre for Molecular Parasitology, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow G12 8TA, Scotland
*Correspondence: iwanaga@doc.medic.mie-u.ac.jp (S.I.), waters@bio.gla.ac.uk (A.P.W.)
DOI 10.1016/j.chom.2010.02.010

SUMMARY

The artificial chromosome represents a useful tool for gene transfer, both as cloning vectors and in chromosome biology research. To generate a *Plasmodium* artificial chromosome (PAC), we had to first functionally identify and characterize the parasite's centromere. A putative centromere (*pbcen5*) was cloned from chromosome 5 of the rodent parasite *P. berghei* based on a *Plasmodium* gene-synteny map. Plasmids containing *pbcen5* were stably maintained in parasites during a blood-stage infection with high segregation efficiency, without drug pressure. *pbcen5*-containing plasmids were also stably maintained during parasite meiosis and mitosis in the mosquito. A linear PAC (L-PAC) was generated by integrating *pbcen5* and telomere into a plasmid. The L-PAC segregated with a high efficiency and was stably maintained throughout the parasite's life cycle, as either one or two copies. These results suggest that L-PAC behaves like a *Plasmodium* chromosome, which can be exploited as an experimental research tool.

INTRODUCTION

The centromere in eukaryotic cells plays a fundamental role in the fidelity of chromosome segregation during nuclear division, through its physical association with the kinetochore. Centromere-associated functions include sister chromatid association and separation, microtubule attachment, chromosomal movement, and establishment of heterochromatin and mitotic checkpoint control (Cleveland et al., 2003; Morris and Moazed, 2007; Pluta et al., 1995). The centromere was first identified in the budding yeast *Saccharomyces cerevisiae*: a plasmid containing a yeast centromere was demonstrated to segregate stably during both mitosis and meiosis, providing experimental evidence for the function of the centromere (Clarke and Carbon, 1980). Following this functional identification, a linear DNA construct, including the centromere, telomeric DNA from *Tetrahymena* rDNA termini, and the autonomously replicating sequence, was generated and termed yeast artificial chromosome (YAC) (Murray and Szostak, 1983). YACs can be stably maintained in a linear form in cells during nuclear division and are widely utilized as a vector for transferring large DNA fragments in yeast (Burke et al., 1987). Several artificial chromosomes for other eukaryotes have since been constructed using this approach of combining the same three elements (Harrington et al., 1997; Ikeno et al., 1998). These artificial chromosomes are being used not only as gene-transfer and cloning vectors but also as genetic tools in chromosome function/biology research (Grimes et al., 2004; Okada et al., 2007).

The goal of this study was to generate a *Plasmodium* artificial chromosome (PAC) as a tool to assist in genome and genetic studies. To achieve this goal, we had first to functionally identify and characterize the centromeres of *Plasmodium*. Putative centromeres (PCENs) of the unicellular protozoan parasite *P. falciparum* were previously identified as 2–3 kb A/T-rich regions with no protein-coding potential, based on a sequence comparison between two *P. falciparum* chromosomes (Bowman et al., 1999). After completion of the genome sequence of *P. falciparum*, PCEN regions with similarly high A/T-rich content (>96%) were described for 13 of the 14 chromosomes (Gardner et al., 2002). Further comparisons of *P. falciparum* chromosomes with a whole-genome synteny map of three rodent malaria parasites (RMPs) revealed a high level of synteny between the genomes of different *Plasmodium* species that extends to the location of the PCEN (Kooij et al., 2005). A more recent study located PCENs on *P. falciparum* chromosomes by examining the distribution of etoposide-mediated topoisomerase-II cleavage sites (Kelly et al., 2006). The PCEN locations identified by this study agreed entirely with the locations predicted by whole-genome sequencing and confirmed the presence of A/T-rich domains with a strict size range of 2.3–3.5 kb. Although these studies have provided an insight into sequences and locations of PCENs, experimental evidence is lacking as to whether these highly A/T-rich DNA regions alone can function as *Plasmodium* centromeres. Based on the syntenic location of PCEN, we identified and cloned the PCEN of *P. berghei* chromosome 5

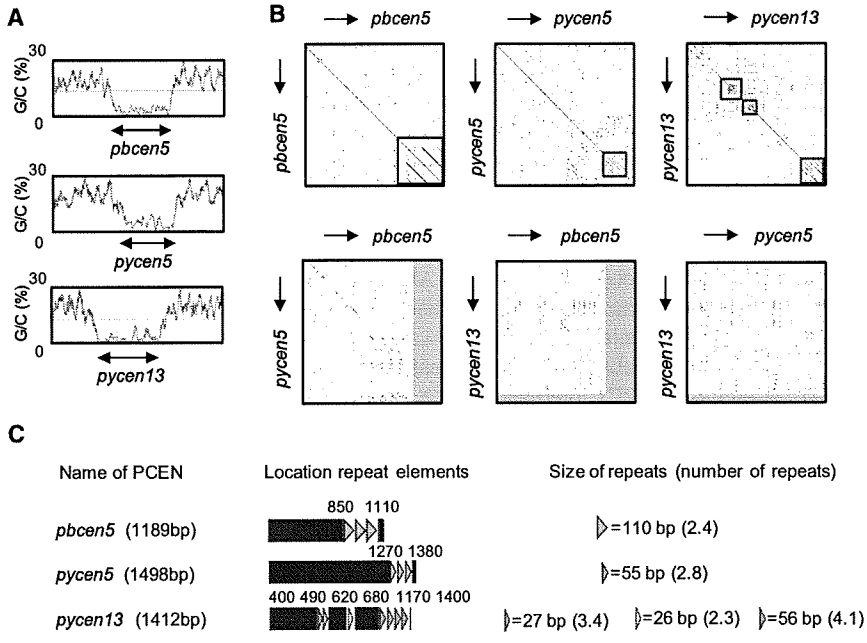


Figure 1. PCENs Are Highly A/T Rich and Consist of Core and Repetitive Regions

(A) The DNA AT content of centromeric regions of *P. berghei* chromosome 5 and *P. yoelii* chromosomes 5 and 13 was analyzed using Artemis 10. Arrows indicate the ultra-high A/T-rich regions termed PCENs (*pbcen5*, 1189 bp; *pycen5*, 1498 bp; and *pycen13*, 1412 bp).

(B) Dot matrix analysis of the three rodent malaria PCENs using the Dotlet program. In this analysis, only the PCEN regions identified in (A) were used. Panels depict the graphical results of the matrix analysis of the rodent PCENs aligned either against themselves or against the other rodent PCENs. The diagonal line within each analysis represents sequence identity, and parallel lines to the diagonal indicate repetitive regions within each PCEN. Boxes highlight repetitive regions used for further analysis. See also Figure S1A.

(C) Lengths and numbers of sequence motifs in the repetitive regions (boxes in B) identified in the dot matrix analyses were determined using the Tandem Repeats Finder program (see also Figure S2 and Table S1). Based on these analyses, the locations, lengths, and numbers of repetitive elements are schematically depicted for the three rodent PCENs analyzed. Each triangle represents a repetitive sequence motif/element. See also Figure S1B and Table S1.

(*pbcen5*) in plasmids. We show that plasmids containing *pbcen5* were both efficiently and stably maintained in transfected parasites during blood-stage multiplication, without drug pressure, demonstrating that *pbcen5* can confer the function of a *Plasmodium* centromere.

In order to generate a PAC, we combined both *pbcen5* and *Plasmodium* telomere sequences into a single plasmid. The third conventional element of a YAC, autonomous replication sequences, has not been identified in *Plasmodium*. However, much *Plasmodium* research has involved plasmid transfections, and despite the varying degrees of segregation efficiency of these plasmids into daughter parasites, they have all been able to replicate inside the parasite, indicating that a variety of DNA sequences can act as *Plasmodium* origins of replication. Telomere sequences of *P. berghei* have been previously identified and cloned (Pace et al., 2000; Ponzi et al., 1985). Evidence has been found that telomere lengthening occurs in *Plasmodium* chromosomes mediated by the activity of *Plasmodium* telomerase enzyme (Figueiredo et al., 2005; Pace et al., 2000; Ponzi et al., 1992). Here we provide evidence that a linear construct, L-PAC, containing both the identified centromere, *pbcen5*, and telomere sequences behaves like a *Plasmodium* chromosome. It segregates with high efficiency during mitosis and meiosis, all in the absence of drug selection pressure, and is stably maintained during the complete life cycle as a few (unconcatemerized) copies. Moreover, we find evidence for telomere elongation, in keeping with the L-PAC being maintained and recognized as a *Plasmodium* chromosome. We discuss the possibilities of using PACs in *Plasmodium* research and how it promises to significantly expand the currently limited set of genetic modification techniques available to understand and manipulate the parasite.

RESULTS

Cloning and Sequence Properties of Putative *Plasmodium* Centromeres

In our previous study, PCENs from chromosomes 5 and 13 of *P. yoelii* were cloned by PCR amplification based on the gene-synteny map of RMP (Kooij et al., 2005). DNA fragments of 4.5 and 4.1 kb were obtained from chromosomes 5 and 13, respectively, which included highly A/T-rich regions predicted as PCENs: *pycen5* (DQ054838.1, 1498 bp, A/T = 97.7%) and *pycen13* (DQ054839.1, 1412 bp, A/T = 98.1%) (Figure 1A). In this study, we amplified the PCEN of chromosome 5 from *P. berghei*. For this PCR we used the same primer pair as was used for *pycen5*, because there were no available sequences from the regions proximate to the PCENs of *P. berghei* based on the RMP gene-synteny map (see the Supplemental Experimental Procedures available online). The amplified 3.8 kb DNA fragment included a highly A/T-rich region, and thus it was predicted to be *pbcen5* (i.e., GU809989, 1189 bp and A/T content of 96.1%; Figure 1A).

To investigate the sequence properties of *pbcen5*, a dot matrix analysis was performed using the Dotlet program (Figure 1B and see also the Supplemental Experimental Procedures). In this analysis, a strict threshold was applied to determine sequence identity, as the sequence complexity of PCEN was low due to its high A/T content. Identity was only recorded if the sequences within a 15 bp sliding box were more than 80% identical. The dot matrix analysis showed that *pbcen5* included a repetitive region indicated by parallel lines in Figure 1B. Furthermore, similar analyses of *pycen5* and *pycen13* as well as 13 annotated PCEN regions from *P. falciparum* (sequence information for PFCEN on chromosome 10 is absent) showed that these PCENs contained

one to three repetitive regions. These results suggest that PCENs consist of the nonrepetitive region, termed the "core region," and the repetitive region, and this general sequence organization is conserved between different *Plasmodium* species (Figures 1B and 1C and Figure S1).

Reciprocal dot matrix analysis between various PCENs showed that sequence identity (indicated by the diagonal line) was found between *pbcen5* and *pycen5* (Figure 1B, the left panel in the lower row). A BLAST2 analysis between these two rodent malaria PCENs showed that they shared homology with a sequence identity of 79%. In contrast, no significant sequence identity was found in reciprocal analyses between the other rodent malaria and *P. falciparum* PCENs (data not shown). Interestingly, sequence identity between *pbcen5* and *pycen5* was restricted to the core regions, indicating that the rates of evolution of the core and the repetitive regions are different.

We further analyzed the possible existence of consensus repetitive motifs within the repetitive regions of the PCENs using the Tandem Repeats Finder program. As can be seen in Figure 1C, Figure S1, and Table S1, different lengths and numbers of repetitive motifs can be distinguished in the various PCENs; however, because of widely divergent sequences, no consensus repetitive repeats were detected (data not shown).

Segregation of Plasmids Containing a PCEN in *P. berghei* Transgenic Blood-Stage Parasites

During blood-stage asexual multiplication, *Plasmodium* chromosomes are segregated between daughter cells during a process termed schizogony, which differs in several aspects from mitosis in most other eukaryotes. After DNA replication, chromosome segregation occurs without observable chromosomal condensation or breakdown of the nuclear envelope (Aikawa, 1966; Rudzinska, 1969), resulting in polyploid syncytia after several (three to four) rounds of genome duplication. Only in the final step of mitosis when the individual daughter parasites (8–16 merozoites) are formed do the individual nuclei bud off from the syncytium. In contrast to this even segregation of chromosomes between the daughter cells, episomally maintained plasmids (introduced by transfection) segregate unevenly during schizogony, such that several of the daughter merozoites do not receive plasmids during each multiplication cycle (O'Donnell et al., 2001; van Dijk et al., 1997). This uneven distribution leads to a growth disadvantage and is believed to explain why plasmids are rapidly lost when transfected parasites are grown in the absence of drug pressure.

To determine whether *pbcen5* contributed to efficient segregation of episomal plasmids during schizogony, we examined the stability of a plasmid including *pbcen5* in blood-stage parasites without drug pressure. As shown in Figures 2B and 2C, the GFP expression in parasites transfected with the control pbGFPcon plasmid lacking PCEN was rapidly lost due to its uneven segregation. However, GFP expression remained stable in the parasites transfected with the plasmid including *pbcen5*, termed pbCEN5 (Figure 2A), and approximately 90% of the parasites with the pbCEN5 plasmid continued to express GFP 21 days after drug pressure had been removed (Figures 2B and 2C). During parasite blood-stage development, we found no growth disadvantage conferred by retaining the pbCEN5 plasmid, as compared to either wild-type or pbGFPcon-transfected parasites. This is based on the observation of comparable courses of parasitemia after infection with standard dose of 10^4–10^5 of either wild-type or PCEN-containing parasites. These results clearly demonstrate that the pbCEN5 plasmid can be stably maintained in parasites independent of drug treatment during multiplication, indicating that pbCEN5 segregated evenly into daughter parasites (cells) during asexual blood-stage mitosis.

Similar results were obtained when the stability of a plasmid containing the 4.5 kb DNA fragment including *pycen5* was assessed in the parasite. The percentage of GFP-positive parasites with pyCEN5 was comparable to that of pbCEN5 in the absence of drug pressure as described above (Figures 2B and 2C). Interestingly, the pfCEN3 plasmid was not as stably maintained in blood-stage *P. berghei* as the plasmids containing a centromere of RMP origin. The percentage of GFP-positive parasites decreased to <60% within 21 days (Figures 2B and 2C), although it was higher than that of parasites transfected with the control plasmid pbGFPcon.

To confirm that the PCEN plasmids pbCEN5, pyCEN5, and pfCEN3 were retained as episomes and not integrated into the parasite genome during nuclear replication, we analyzed the genomic DNA of the parasites collected at the start and end of the 18–21 day multiplication period. HindIII-digested DNA was hybridized with a *Pbdhfr-ts* 5′UTR DNA fragment, as this probe recognizes both the endogenous *dhfr-ts* gene (4.9 kb fragment) and the plasmid. As seen in Figure 2D, only the endogenous *dhfr-ts* gene and the plasmid were detected in all parasite populations, both at the beginning and at the end of the multiplication period, indicating that the plasmids were not integrated into the parasite genome but were maintained in episomal form in the parasite nuclei. These results were supported by plasmid rescue experiments in which we were able to recover intact plasmids from all of the parasite populations (data not shown).

Using the percentages of GFP-expressing parasites at the end of the 18–21 day period, we calculated segregation efficiencies for each of the plasmids, based on the assumption that during schizogony the parasite nuclei undergo four nuclear divisions, resulting in the production of 16 daughter nuclei over a 24 hr period (see the Experimental Procedures). The segregation efficiencies of the plasmids pbCEN5, pyCEN5, pfCEN3, and pbGFPcon were calculated as 99.9%, 99.9%, 99.3%, and 93.9% per nuclear division, respectively. When the predicted percentages of GFP-positive parasites were plotted on the basis of the calculated segregation efficiencies over time, it was seen that these percentages fit closely with the observed percentages of GFP-positive parasites (Figure 2C and Figure S2).

Southern analysis of plasmid presence indicated that in parasites maintained in the presence of pyrimethamine, the hybridization signal of the pbGFPcon plasmid was significantly stronger than that of the PCEN plasmids (Figure 2D), indicating that the former plasmid is maintained at a higher number of copies in *P. berghei* (Figure 2D). Comparison of signal intensities for DNA extracted from parasites transfected with the various plasmids and maintained under drug pressure revealed that the pbGFPcon plasmid copy number is on average 28.3 (SE ± 0.3; Table 1), whereas the PCEN plasmid copy numbers are significantly lower (i.e., pbCEN5, 8.0 ± 0.4; pyCEN5, 6.6 ± 0.3; and pfCEN3, 5.4 ± 1.6 copies per parasite; Table 1). After

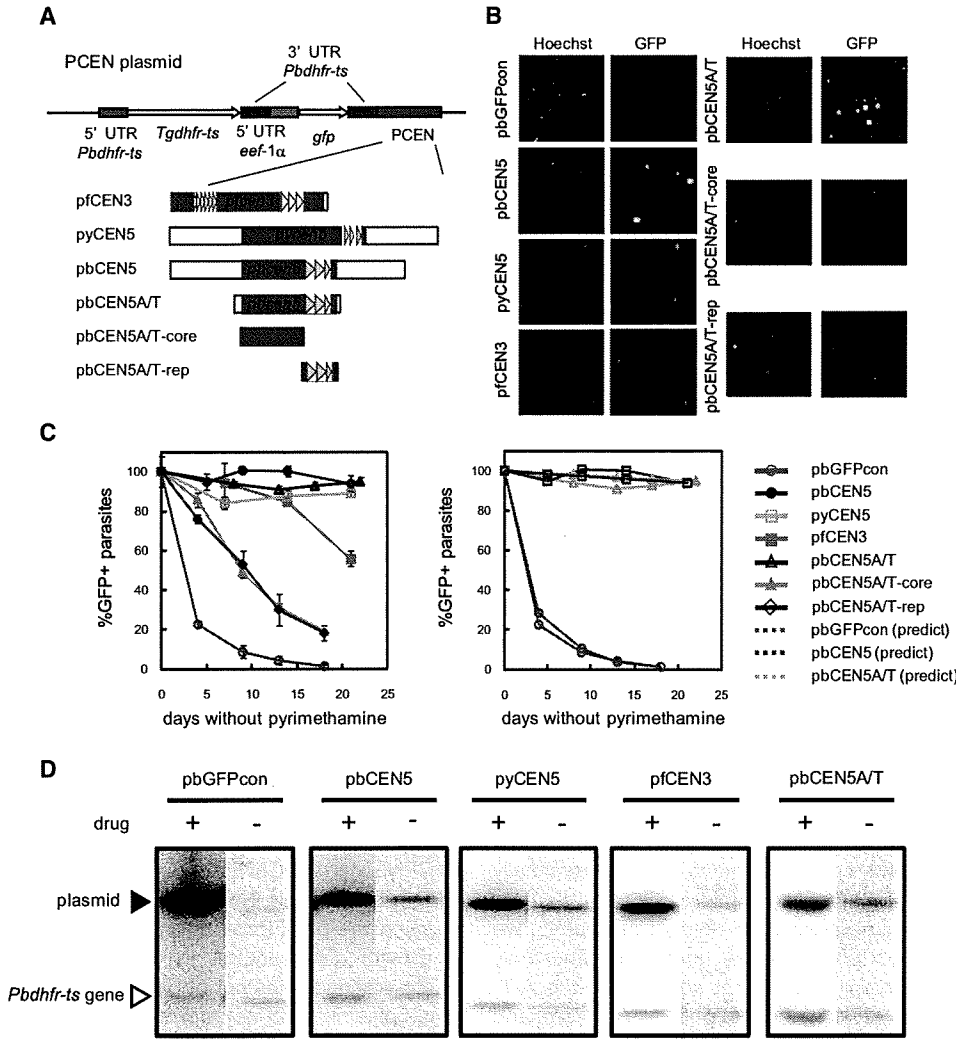


Figure 2. Addition of PCEN to Plasmids Improves Their Segregation Efficiency during *P. berghei* Blood-Stage Multiplication

(A) Schematic representation of the various PCEN plasmids. The plasmid contains the *Tgdhfr-ts* selectable marker and the GFP reporter protein under the control of the *eef1αa* promoter. pbGFPcon is a control plasmid without a PCEN. The names of the resulting PCEN plasmids are indicated to the left of the schematic drawings of the cloned PCEN DNA fragments.

(B) GFP expression of Hoechst 33258-stained (all parasites) blood-stage *P. berghei* transfected with the various PCEN plasmids, 18–21 days after drug withdrawal.

(C) The percentage of GFP-positive parasites during asexual multiplication of blood-stage *P. berghei* that have been transfected with the various PCEN-containing plasmids (left panel). Observed percentage (right panel) of GFP-positive parasites determined by fluorescence microscopy (solid lines) compared to the predicted percentage of GFP-positive parasites (dashed lines), based on the calculated segregation efficiencies of the various plasmids (see also Figure S2). The error bars represent standard error.

(D) Southern analysis of the presence of plasmids in the parasites either after 7 days of multiplication in the presence of pyrimethamine (+) or after 18–21 days in the absence of pyrimethamine (–). Blots were hybridized to the *P. berghei* 5′UTR *dhfr-ts* probe, which recognizes the plasmid (black triangle) and the endogenous single *dhfr-ts* genome copy (white triangle, 4.9 kb). The sizes of signals from each plasmid are as follows: pbGFPcon, 9.0 kb; pbCEN5, 12.4 kb; pyCEN5, 13.1 kb; pfCEN3, 11.5 kb; and pbCEN5A/T, 10.4 kb.

removal of drug pressure, the PCEN plasmid copy number 21 days later was reduced by 2- to 4-fold for the RMP centromere-containing plasmids and >5-fold for pfCEN3. However, there was a strong reduction (>47-fold) in the pbGFPcon plasmid copy number (Table 1). These results indicate that the PCEN elements not only confer greater segregation efficiency to the plasmids but also maintain their low copy number independent of drug treatment.

Only the Highly A/T-Rich Centromeric Sequences Are Required to Confer High Segregation Efficiency to Plasmid in Blood-Stage Parasites

We next examined whether only the highly A/T-rich region (consisting of both the core and repetitive regions) predicted to be *pbcen5* was sufficient to produce high segregation efficiency. First, the amplified 1.4 kb fragment including *pbcen5* (1189 bp) was cloned into the plasmid, thereby creating the plasmid

Table 1. The Copy Number of PCEN Plasmids and PAC in Parasites

	+Drug	−Drug
pbGFPcon	28.3 (0.3)	0.6 (0.1)
pbCEN5	8.0 (0.4)	1.9 (0.1)
pyCEN5	6.6 (0.3)	2.8 (0.3)
pfCEN3	5.4 (1.6)	0.9 (0.1)
pbCEN5A/T	2.1 (0.3)	1.3 (0.4)
C-PAC	1.8 (0.6)	1.3 (0.1)
L-PAC	2.1 (0.6)	1.3 (0.1)

Copy numbers of PCEN plasmids, C-PAC and L-PAC were calculated from the ratio of intensities obtained from Southern analysis as shown Figures 2D and 5D. The standard errors are indicated in the parenthesis.

pbCEN5A/T (Figure 2A). The pbCEN5A/T plasmid was introduced into *P. berghei*, and its stability during mitosis in the absence of drug pressure was examined. Indeed, pbCEN5A/T was stably maintained during mitosis of asexual blood-stage parasites similar to the pbCEN5 plasmid, indicating that only the A/T-rich region (i.e., *pbcen5*) was required to confer high segregation efficiency to episomal plasmids (99.9%) (Figures 2B and 2C). Southern analysis using DNA isolated from the transfected parasites showed that pbCEN5A/T was maintained as an episomal plasmid throughout parasite multiplication and, further, that a low copy number was maintained both in the presence and absence of the drug (Figure 2D, Table 1). These results demonstrate that plasmids containing only the highly A/T-rich region, consisting of the core and repetitive regions, behave essentially the same as the other RMP PCEN plasmids.

To determine the minimal functional element of *pbcen5*, we generated two additional plasmids that contained either only the core (nonrepeat) or only the repeat sections of the A/T-rich region of *pbcen5*, designated as pbCEN5A/T-core and pbCEN5A/T-rep, respectively (Figure 2A). In parasites transfected with these plasmids, the percentage of GFP-expressing blood-stage parasites gradually decreased in the absence of drug pressure (Figures 2B and 2C): segregation efficiencies were 97.7% for pbCEN5A/T-core and 97.3% for pbCEN5A/T-rep. These efficiencies were higher than that of the control pbGFPcon plasmid but lower than those of the plasmids containing the complete A/T-rich region. These results indicate that the entire A/T-rich region is the minimal functional unit of *pbcen5* necessary to permit efficient segregation and maintenance of plasmids during asexual blood-stage mitosis.

Plasmids Containing PCEN Elements Are Stably Maintained in Parasites during Meiosis and Mitosis inside the Mosquito

In addition to asexual multiplication during blood-stage development, malaria parasites have a sexual phase involving meiosis (within the zygote/ookinete) and additional phases of extensive asexual multiplication in the mosquito host (within the oocysts). During meiotic and mitotic division of parasites in the mosquito host, segregation of chromosomes also occurs without chromosome condensation or breakdown of the nuclear membrane, resulting in polyploid nuclei (polyploid syncytia) containing as few as 4 (i.e., the tetraploid ookinete, after meiosis) to as many as 10,000 haploid genomes (i.e., midgut sporozoites). To analyze the segregation behavior of the PCEN plasmids during meiosis and mitosis in the mosquito, parasites carrying pbCEN5, pbCEN5A/T, or pbGFPcon were transmitted to mosquitoes, and the GFP expression of their salivary gland sporozoites was examined. The results indicated that 89.2% ± 3.6% and 89.9% ± 3.3% of sporozoites infected with parasites carrying pbCEN5 and pbCEN5A/T plasmids, respectively, expressed GFP, whereas only 1.9% ± 1.0% of sporozoites with parasites carrying the pbGFPcon plasmid were GFP positive (Figure 3). A growth disadvantage to parasites carrying either of the PCEN plasmids was not observed, suggesting that retention of these plasmids did not affect on the development of parasites in the mosquito. These results showed that addition of *pbcen5* sequences to plasmids greatly improves segregation and

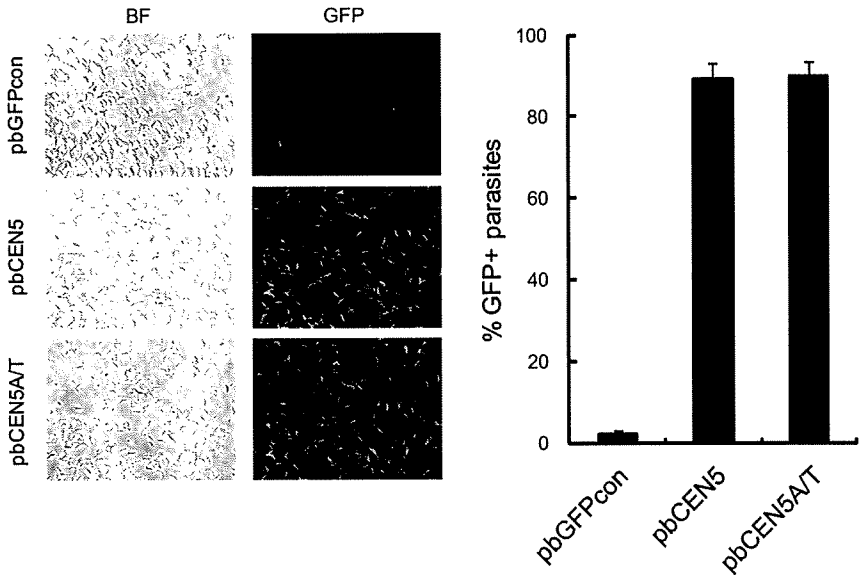


Figure 3. Retention of PCEN-Containing Plasmids in Parasites after Mosquito Passage

(Left panel) The GFP expressions in salivary gland sporozoites containing pbCEN5, pbCEN5A/T, or the control pbGFPcon are shown (BF, bright field image of the same sporozoites). (Right panel) The percentages of GFP-positive sporozoites (right) for parasites transfected with the various PCEN-containing plasmids were determined by fluorescence microscopic analysis. The error bars represent standard error.

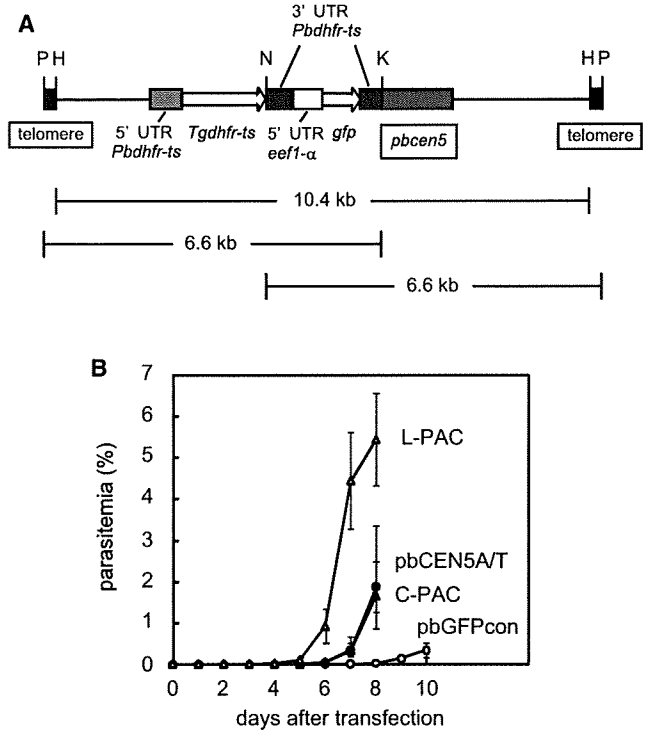


Figure 4. Transfection of Parasites with the Linear *Plasmodium* Artificial Chromosome Construct, Containing PCEN and Telomere Sequences

(A) Schematic representation of the L-PAC construct. The construct contains the *Tgdhfr-ts* selectable marker and the GFP reporter protein under the control of the *eef1αa* promoter. Restriction sites for PmeI (P), HindIII (H), NheI (N), and KpnI (K) are shown, which were used for the Southern analysis to determine the size of the DNA fragments digested with each enzyme.

(B) The course of parasitemia in mice infected with pbGFPcon (open circle), pbCEN5A/T (closed circle)-, L-PAC (open triangle)-, and C-PAC (closed triangle)-transfected parasites. All parasites (1×10^8) were transfected with 5 μg DNA, and mice were treated with pyrimethamine 1 day after i.v. injection of parasites. The error bars represent standard error.

maintenance during both meiosis in zygotes/ookinetes and mitosis in oocysts.

Generation of a *Plasmodium* Artificial Chromosome

The results reported above clearly indicate that the region of highly A/T-rich DNA predicted to be a *Plasmodium* centromere did indeed confer the function of a centromere to plasmids during parasite nuclear division. We therefore attempted to construct a PAC consisting of this *pbcen5* centromere as well as *P. berghei* telomeric DNA sequences. Briefly, a DNA insert containing two telomeric fragments, consisting of the previously characterized *P. berghei* telomeric CCCT(A/G)AA sequences (Pace et al., 1987) oriented head to head and separated by an ~500 bp spacer region, was cloned into the pbCEN5A/T plasmid, resulting in a plasmid designated circular *Plasmodium* artificial chromosome (C-PAC). The linear DNA construct (L-PAC) was generated through linearization of the C-PAC construct using PmeI, which removes the spacer fragment between the two telomeric fragments (Figure 4A). Unexpectedly, the transfection efficiency of the L-PAC was significantly higher than that of the circular plasmids. In multiple independent experiments, the mice injected with L-PAC-transfected parasites developed an parasitemia of ~1% 6 days after transfection, whereas a parasitemia of ~1% level was not observed until day 7 or 8 in mice injected with parasites transfected with the same amount of circular plasmid DNA, i.e., the C-PAC or pbCEN5A/T (Figure 4B).

The *Plasmodium* Artificial Chromosome Is Stably Maintained throughout the Complete Life Cycle

Segregation and maintenance of the C-PAC and the L-PAC constructs were investigated using the same methodologies as described for the PCEN plasmids. Approximately 85% of the C-PAC- and L-PAC-transfected blood-stage parasites expressed GFP 20 days after withdrawal of drug pressure (Figures 5A and 5B), and their segregation efficiencies at the blood stage were calculated as 99.8% and 99.9%, respectively. To confirm that neither the C-PAC nor the L-PAC was integrated into the parasite genome during replication, Southern analysis was carried out using KpnI-digested genomic DNA for the C-PAC and undigested genomic DNA for the L-PAC. A single 11.5 kb fragment was detected in the KpnI-digested genomic DNA for the C-PAC (Figure 5D), and a single fragment was detected in the undigested genomic DNA for the L-PAC (Figure 6, lanes 5 and 9), demonstrating that neither PAC was integrated into the genome. A similar analysis using HindIII-digested genome DNA for the C-PAC and the L-PAC determined copy numbers of 1.8 ± 0.6 and 2.1 ± 0.6 in drug-treated parasites, respectively, and 1.3 copies for both PAC vectors in the absence of drug treatment (Figure 5D and Table 1). These efficiencies of segregation and copy numbers in the blood stage were comparable to those of the pbCEN5A/T plasmid without the telomeric sequences. Next, we analyzed maintenance of the C-PAC and the L-PAC during mosquito-stage development and found that 83.2% ± 1.6% and 85.5% ± 1.6% of the sporozoites expressed GFP, respectively (Figures 5B and 5C). All of these results demonstrate that the addition of telomeric sequences and linearization did not alter the segregation efficiency and maintenance observed for plasmid pbCEN5A/T during mitosis and meiosis, suggesting that the centromere is able to function within either a circular or a linear DNA construct.

We also examined segregation of the C-PAC and the L-PAC in the dividing forms in the liver. After sporozoites are injected by a mosquito, they migrate to the liver, where they invade hepatocytes. It is in the hepatocytes where they undergo a rapid expansion: one infected hepatocyte can produce several thousand merozoites. For this analysis, it was not possible to estimate the segregation efficiency directly in liver cells; therefore, we assessed the percentage of GFP-positive (i.e., containing either L- or C-PAC) blood-stage parasites as soon as they were patent in the blood after liver stage development. As mentioned above, ~85% of the PACs-containing sporozoites were GFP positive, and the percentage of GFP-positive infected red blood cells after liver stage development was 62.2% ± 3.4% and 71.2% ± 3.4% for C-PAC and L-PAC, respectively (Figures 5B and 5C). The patent period in mice after infections established with sporozoites containing either C- or L-PAC was comparable to the patent period after infections initiated with the same number of wild-type sporozoites (data not shown). Those results would

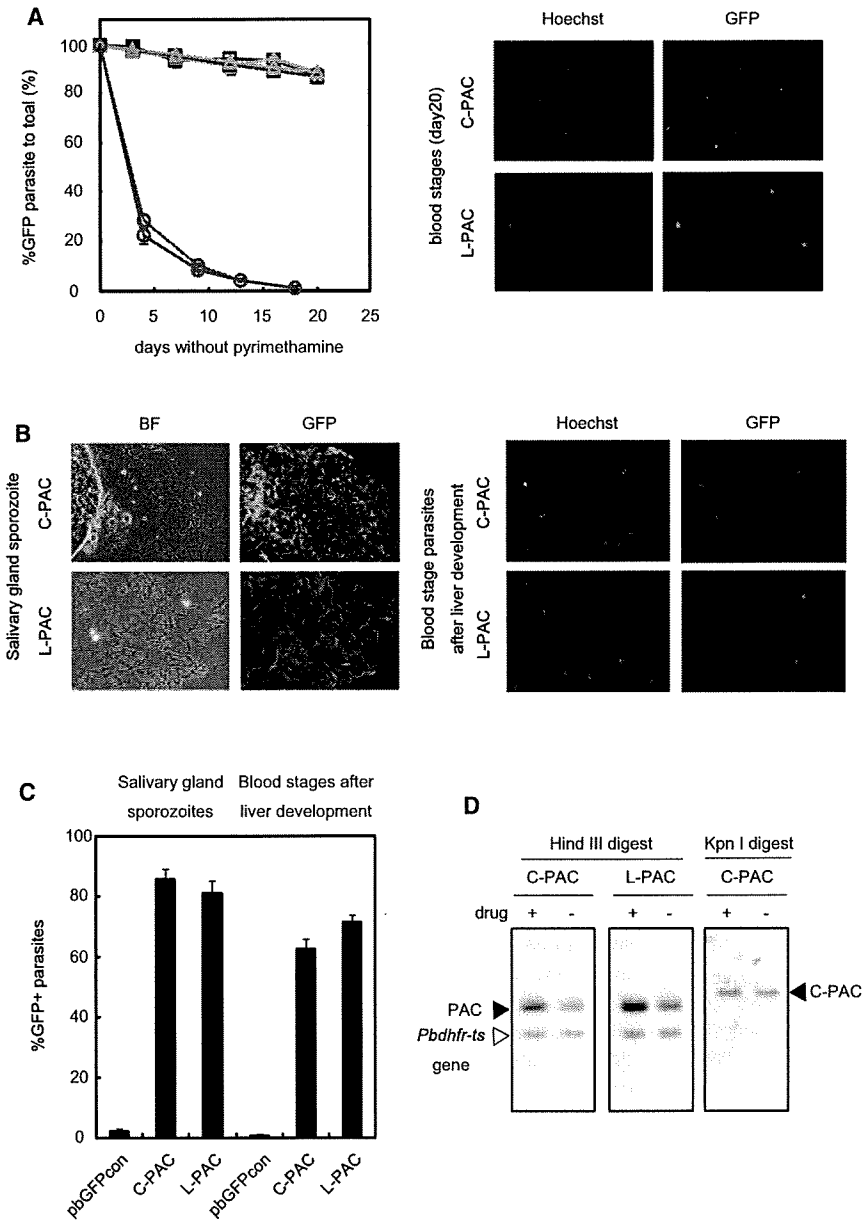


Figure 5. Efficient Segregation and Maintenance of Circular and Linear DNA Constructs Containing Both *pbcen5* and Telomere Sequences

(A) (Left panel) The percentages of GFP-positive parasites during blood-stage asexual multiplication of *P. berghei* transfected with the C-PAC (blue), L-PAC (green), or the control plasmid pbGFPcon (red) are shown. The observed percentages of GFP-positive parasites are determined by fluorescence microscopy (solid lines) compared to the predicted percentages of GFP-positive parasites (dashed lines), based on the calculated segregation efficiencies of the various constructs. The error bars represent standard error. (Right panel) The GFP expressions (C-PAC- and L-PAC-containing parasites) of Hoechst 33258-stained transfected parasites (all parasites) were observed by fluorescence microscopy after 18–20 days of asexual multiplication in the blood, without drug pressure.

(B) (Left panel) The GFP expressions in salivary gland sporozoite of parasites transfected with the C-PAC or the L-PAC are shown (BF, bright field image of the same sporozoites). (Right panel) GFP expression in the blood stages of parasites transfected with the C-PAC and L-PAC after first completing a passage through the mosquito and multiplication in the liver are shown.

(C) Percentage of GFP-positive, variously transfected (C-PAC, L-PAC, and pbGFPcon) salivary gland sporozoites after parasite development in the mosquito, as well as the percentage of GFP-positive blood stages after parasite passage through multiplications in the mosquito and in the liver. The error bars represent standard error.

(D) Southern analysis showing the presence of the DNA constructs in blood-stage parasites after either 7 days of multiplication in the presence of pyrimethamine (+) or 20 days in the absence of pyrimethamine (–). Southern analysis was performed on HindIII-digested C-PAC- or L-PAC-transfected parasite DNA with *P. berghei* 5′UTR *dhfr-ts* as a probe, which recognizes both the construct (black triangle) and the endogenous *dhfr-ts* genome copy (white triangle). Based on this result, the copy numbers are estimated and given in Table 1. Southern analysis of C-PAC-transfected parasite DNA digested with KpnI using the *gfp* gene as a probe identified the presence of a single fragment, indicating that the C-PAC is not integrated into the parasite genome.

indicate that despite the rapid expansion in parasite numbers in the liver, plasmids segregated with a high efficiency during the multiple rounds of mitotic division during liver-stage development.

To confirm whether the L-PAC was as an independent linear DNA construct throughout the complete parasite life cycle, we analyzed genomic DNA obtained from blood-stage parasites before and after mosquito transmission. The results of Southern analysis of undigested DNA hybridized with a probe against *gfp* demonstrated that the L-PAC is maintained as an extrachromosomal DNA construct during all stages of the parasite's life cycle (Figure 6). Southern hybridization with the same probe on the same DNA but now digested with the "single cutting" restriction enzymes, KpnI or NheI, revealed only a single band (~7 kb), which is again consistent with a linear, nonintegrated construct. Interestingly, both in the blood stages after 20 days of multiplication and after mosquito transmission, the size of both the KpnI and the NheI fragments was slightly larger than the expected 6.6 kb fragment of the original L-PAC. Moreover, the total size of the undigested construct was also slightly larger (~12 kb) than the size of the original L-PAC (10.8 kb). However, the internal fragment of the L-PAC remaining after HindIII digestion, which does not contain the telomeric sequences, was the same size as the original L-PAC. These results would therefore indicate that the increase in size of the undigested construct and the increase in the KpnI- or NheI-restricted fragments results from

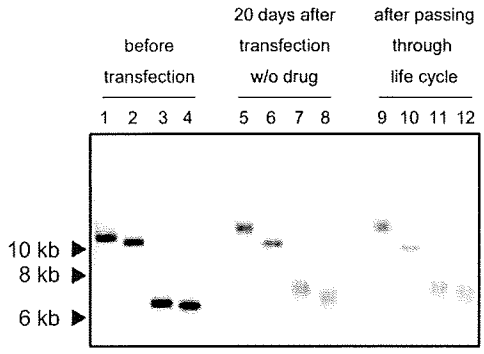


Figure 6. The L-PAC Was Maintained as a Linear Form in the Parasite throughout the Complete Life Cycle

Southern analysis to examine the L-PAC construct in transfected parasites after 20 days of multiplication in the blood in the absence of pyrimethamine, as well as in the re-emerging parasite in the blood after mosquito passage and multiplication in the liver. Genomic DNA was digested with HindIII (lanes 6 and 10), KpnI (lanes 7 and 11), or NheI (lanes 8 and 12) and hybridized with the *gfp* gene. In addition, undigested genomic DNA (lanes 5 and 9) was also resolved. Undigested L-PAC (lane 1) and HindIII (lane 2)-, KpnI (lane 3)-, and NheI (lane 4)-digested L-PAC were analyzed under the same conditions.

an increase in size of the two telomeric fragments, indicating that telomere lengthening occurs by addition of telomeric repeats and, further, that the L-PAC telomeres were maintained at a length of up to 1.4 kb.

DISCUSSION

After completing the sequencing of the *P. falciparum* genome, PCENs were predicted to be within 2–3 kb highly A/T-rich (>96%) gene-free regions. However, despite the recent increase in knowledge about the sequence and location of PCEN, no functional evidence demonstrating that PCEN regions function as centromeres has been reported. In this study, we demonstrated that the A/T-rich region predicted to be the centromere of chromosome 5 of *P. berghei* (*pbcen5*) conferred improved segregation efficiency to episomal plasmids. Transfected plasmids containing *pbcen5* were stably maintained in parasites in the absence of drug pressure. In *Plasmodium*, standard plasmids (e.g., without PCEN elements) are rapidly lost from transfected parasites during asexual multiplication when no drug selection is applied. The stable maintenance of the *pbcen5*-containing plasmids provides strong evidence that the *pbcen5* does indeed function as centromere and that these data constitute the functional characterization of a *Plasmodium* centromere.

The PCEN regions consist of a nonrepetitive region (core) and regions containing repetitive elements, as identified by dot matrix analysis. In an attempt to define a "minimal functional unit" within the *pbcen5* centromere, we analyzed the maintenance of plasmids that contained either only the regions containing repetitive elements or the core sequence. While some increase in segregation efficiency of plasmids is derived from having either one of these elements, high segregation efficiencies were only achieved when both the PCEN repeat and core regions are present, indicating that the entire A/T-rich sequence (i.e., both the core and the repetitive regions) is required for functioning of a centromere. Dot matrix analyses of the other available PCENs from *P. falciparum* and *P. yoelii* indicated a similar organization of all PCENs, consisting of the core and repetitive regions. Despite this conservation of the overall PCEN structure, we did not find conservation in location, sequence, or number of repeats between different chromosomal PCENs, neither within the same genome nor between different species. These results are in large part in agreement with centromere analyses of *P. falciparum* performed by Kelly et al. (Kelly et al., 2006). These observations suggest that the sequence organization and DNA composition of PCEN regions are important to centromere function.

In most strains of *Saccharomyces cerevisiae*, a 6.3 kb plasmid is present, the so-called "2 μm plasmid," which is replicated by a "rolling-circle" mechanism and is maintained in high copy numbers (50–100 copies/cell) (Broach and Hicks, 1980; Tschumper and Carbon, 1983). The addition of a yeast centromere to the 2 μm plasmid suppresses its rolling-circle replication, resulting in maintenance of only a few copies of the plasmid (one to five copies per cell) (Tschumper and Carbon, 1983). Standard *Plasmodium* plasmids, such as pbGFPcon, also replicate through a rolling-circle mechanism, and these ultimately form large concatameric multimers generated by both intra- and intermolecular recombination. In the present study, plasmids containing *pbcen5* are maintained in low copy numbers (i.e., one to two copies per cell) during blood-stage multiplication, indicating that *pbcen5* also prevents the formation of concatamers by suppressing a rolling-circle replication by the addition of a centromere.

The linear L-PAC construct, containing both the *pbcen5* and telomeric sequences, is stably maintained throughout the complete life cycle of the parasites, indicating that the addition of telomere sequences prevented the degradation of the ends of the linear construct. Attempts to generate transgenic parasites using a linear construct containing *pbcen5* but without telomeric sequences were unsuccessful (data not shown), supporting the importance of telomeric sequences in the maintenance of a linear construct. Our Southern analysis of L-PAC DNA obtained from parasites after successive rounds of replication in both host and vector indicates a lengthening of "telomeric ends" of the L-PAC plasmid and their maintenance within a certain length (1.4 kb). In *P. berghei*, telomere lengthening occurred after the introduction of a short (~500 bp) telomeric sequence at the end of a chromosome by a targeted terminal deletion event, resulting in a telomere extension to about a length of 1.2 kb (Pace et al., 2000), which is the average length of *P. berghei* telomeres (Dore et al., 1994). This excellent concordance between our present result and previous study suggested that telomere elongation and regulation of its length in L-PAC occurred by the same mechanism observed in parasite original chromosome.

In *P. falciparum*, multigene families encoding clinically important proteins such as the variant antigens PfEMP-1, RIFIN, and STEVOR are located in subtelomeric regions of chromosomes (Gardner et al., 2002). Genes of these multigene families are often expressed in a mutually (or partially mutually) exclusive manner (Scherf et al., 1998). Recent ChIP-on-chip analyses have demonstrated that the Lys9 of histone H3 is trimethylated

in both subtelomeric and telomeric regions, and the chromatin structure of these regions is believed to be an important factor in the mutually exclusive gene expression (Lopez-Rubio et al., 2009). In other *Plasmodium* species, the subtelomeric regions contain other multigene families (Janssen et al., 2004), and their transcription may be regulated by similar epigenetic mechanisms that are important in *P. falciparum*. The trimethylation of Lys9 of histone H3 is involved in the regulation of telomeric length. Given that the telomeric length of L-PAC is controlled, it is entirely plausible that L-PAC has a similar chromatin modification that is present in *Plasmodium* chromosomal telomeres. Therefore, this suggested that L-PAC reporter constructs should permit us to examine the mechanism regulating epigenetically controlled gene expression of subtelomerically located multigene families.

Various mutant parasites exhibiting clinically important phenotypes (e.g., drug resistance) have been isolated from field or generated in laboratories thus far. Currently, linkage analysis using progeny clones from genetic crosses between mutant and wild-type parasites is the only way to identify the specific genes responsible for these observed phenotypes. These methods are often complicated, technically challenging, and time consuming. We propose here an alternative approach with gene libraries generated by using L-PAC: genomic DNA or cDNA libraries of mutant parasites are constructed by using L-PAC and then directly introduced into wild-type parasites. The transformed parasites acquiring the mutant phenotype can then be selected by screening (e.g., drug selection), and then target genes incorporated in the L-PAC are identified from the selected parasites. This approach to surveying target genes using the gene library will provide fresh avenues for the analysis of mutant parasites.

Genetic modification of *Plasmodium* using episomally maintained plasmids was first described about 15 years ago (van Dijk et al., 1997), and the use of this method of genetic analysis has contributed significantly to studies of the parasite's biology. However, genetic modification using episomally maintained plasmids has two limitations that make it unsuitable for a number of parasite transformation studies. First, maintenance of the plasmids in parasites requires drug selection to prevent loss of the plasmids during mitosis. Consequently, the use of this technique is mainly restricted to the blood stage and is less suited for studies on parasites in either the mosquito or the liver. Second, episomal plasmids often form undesirable concatamers, resulting in high and variable copy numbers of the plasmid. As a result, the control or estimation of transcription activity of the genes introduced into the plasmid is difficult, and therefore plasmids are not suitable for precisely assessing gene expression levels in reporter assays. These technical limitations can be overcome by using the PAC and PCEN plasmids, since they are stably maintained throughout the life cycle and do not form multimeric concatamers. For example, accurate reporter assays at the mosquito and liver stages will only be achieved by using modified PAC and the PCEN plasmids or equivalents. Indeed, we have recently reported the use of such reporter assays (Yuda et al., 2009). In this study, we generated reporter PCEN plasmids containing several promoter regions (i.e., either containing or not containing *cis*-acting elements) and were able to precisely determine the transcriptional activities of those promoter regions controlled by the ookinete-specific transcription factor.

In conclusion, we describe in this paper the functional characterization of a *Plasmodium* centromere and the generation of PAC that appears to behave like a true *Plasmodium* chromosome. The study not only will expand the number of molecular tools available to malarial research but also sets out a template that should permit the generation of *P. falciparum* PAC.

EXPERIMENTAL PROCEDURES

Construction of Plasmids Containing PCEN

The various (complete and partial) PCEN-containing constructs were cloned into the plasmid vector pbGFPcon, which contains a selectable marker cassette encoding the *Toxoplasma gondii* pyrimethamine-resistant *dhfr-ts* gene (*Tgdhfr-ts*) as well as a *gfp* expression cassette under the control of the *P. berghei eef1αa* promoter. pbGFPcon was digested with EcoRI, thereby removing its *d-ssu-rrna* sequence. Subsequently, the PCEN-containing DNA fragments were digested with EcoRI and cloned into the EcoRI-digested pbGFPcon. All PCEN-containing plasmids are schematically shown in Figure 2A.

Construction of the *Plasmodium* Artificial Chromosome

The DNA fragment containing the two telomeric sequences, oriented head to head with an ~500 bp spacer region, was digested with HindIII and then cloned into the HindIII-digested pbCEN5A/T plasmid. This resulted in a circular plasmid, termed C-PAC. When we digested the C-PAC with PmeI, removing the spacer region between the two telomeres, the remaining linearized fragment was designated L-PAC.

Assays to Determine the Efficiency of Segregation and Maintenance of PCEN Plasmids and PACs in Transfected Parasites

Transfected parasites were first maintained for a period of 1–2 weeks in Swiss mice under pyrimethamine drug pressure. After this period, when the parasitemia had reached 5%–10%, the parasites were transferred to naive mice. The parasites were then maintained in these mice and subsequent mice for 18–21 days without drug treatment. During this period, mechanical passage of the parasites was repeated three to five times once the parasitemia in the infected mice had reached 5%–10%. Mechanical passage was performed by intraperitoneal injection of 2–4 × 10^4 infected erythrocytes. At each mechanical passage, 10 μl of blood was collected from each mouse in 1 ml of culture medium (RPMI1640 medium [pH 7.3], containing 10% fetal calf serum) to determine the percentage of parasites retaining PCEN constructs. This was done by assessing the percentage of infected erythrocytes that were GFP positive using a fluorescence microscope. Specifically, to determine the percentage of GFP-positive parasites, the samples of infected blood were incubated at 37°C for 5 min in the presence of Hoechst 33258 (10 μM final concentration) to stain all parasite nuclei. According to the percentages of GFP-positive parasites at the end of the 18–21 day multiplication period, we calculated the segregation efficiencies of the PCEN-containing constructs in the blood stage based on the assumption that during blood-stage schizogony the parasite nuclei undergo four nuclear divisions, resulting in the production of 16 daughter nuclei over a 24 hr period.

To evaluate segregation and maintenance of the PCEN-containing constructs during mosquito development, transfected parasites were first maintained for a period of 1 week in Swiss mice under pyrimethamine drug pressure. To obtain mosquito-stage parasites, *Anopheles stephensi* mosquitoes (3–4 days old) were fed for 10 min on anesthetized, transgenic parasite-infected mice. Mosquitoes were maintained at 20°C and 80% humidity and fed daily on 10% sucrose. To isolate salivary gland sporozoites, salivary glands were dissected at day 20–22 p.i., collected in ice-cold saline, and homogenized in 200 μl of medium 199. The percentage of GFP-positive sporozoites was determined by fluorescence microscopy.

To analyze segregation and maintenance of the plasmids within the parasites during liver-stage development, 3 × 10^4 salivary gland sporozoites were injected intravenously into young Wistar rats (3 weeks old). Four to eight

days after infection of the rats, the development of blood-stage parasitemia was determined using Giemsa-stained blood smears. At a parasitemia of 5%–10%, tail blood was collected and stained with Hoechst as described above, and the percentage of GFP-positive to Hoechst-stained parasites was determined by fluorescence microscopy.

Southern Analysis of DNA Collected from *P. berghei* Blood Stages Transfected with PCEN Plasmids and PACs

Genomic DNA was isolated from blood-stage parasites transfected with the various PCEN plasmids and PACs and maintained for 18–21 days without pyrimethamine. Parasite DNA containing the episomal plasmid DNA was digested with HindIII, which cleaves a single site within all PCEN plasmids and two sites in both the C- and L-PAC constructs. Southern hybridization was performed using a 5′UTR DNA fragment of the *P. berghei dhfr-ts* gene as a probe, which is present in all PCEN plasmids and PACs and on chromosome 7 of the *P. berghei* genome. The hybridization intensity of the signal in each sample was quantified using Quantity One software (Bio-Rad). By comparing the intensity of the hybridization signal for the plasmid and the hybridization of the probe to the endogenous *dhfr-ts* gene in the genome, the copy number of each plasmid was determined. In addition, to examine the integration of the C-PAC, genomic DNA was digested with KpnI, which cleaves at a single site within the C-PAC. The digested DNA was processed and analyzed by Southern analysis as described above.

Genomic DNA from parasites transfected with the L-PAC was collected both from blood stages, as described above, and from blood stages after mosquito transmission. This genomic DNA from parasites transfected with the L-PAC was digested with either HindIII, which recognizes two sites within the L-PAC, or with NheI or KpnI, which each recognize a single site within the L-PAC. Hybridization was performed using a fragment of the *gfp* gene as a probe (present exclusively in the L-PAC plasmid), which specifically recognized fragments from the L-PAC vector.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, one table, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.chom.2010.02.010.

ACKNOWLEDGMENTS

We would like to thank Tomomi Kato for her expert technical assistance. This work was supported by the Ministry of Education, Science, Culture, and Sports of Japan (grants 21022031 and 21790406 to S.I., and grants 20249023 and 21022019 to M.Y.) and by a Wellcome Trust Post-Genome Initiative Grant (C.N.) and Principal Research Fellowship (A.P.W.), European Network of Excellence, BioMalPar (A.P.W., C.J.J., C.N.).

Received: November 10, 2009
Revised: January 12, 2010
Accepted: February 24, 2010
Published: March 17, 2010

REFERENCES

Aikawa, M. (1966). The fine structure of the erythrocytic stages of three avian malarial parasites, Plasmodium fallax, P. lophurae, and P. cathemerium. Am. J. Trop. Med. Hyg. *15*, 449–471.

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., et al. (1999). The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum. Nature *400*, 532–538.

Broach, J.R., and Hicks, J.B. (1980). Replication and recombination functions associated with the yeast plasmid, 2 mu circle. Cell *21*, 501–508.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science *236*, 806–812.

Clarke, L., and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature *287*, 504–509.

Cleveland, D.W., Mao, Y., and Sullivan, K.F. (2003). Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell *112*, 407–421.

Dore, E., Pace, T., Picci, L., Pizzi, E., Ponzi, M., and Frontali, C. (1994). Dynamics of telomere turnover in Plasmodium berghei. Mol. Biol. Rep. *20*, 27–33.

Figueiredo, L.M., Rocha, E.P., Mancio-Silva, L., Prevost, C., Hernandez-Verdun, D., and Scherf, A. (2005). The unusually large Plasmodium telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. Nucleic Acids Res. *33*, 1111–1122.

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.

Grimes, B.R., Babcock, J., Rudd, M.K., Chadwick, B., and Willard, H.F. (2004). Assembly and characterization of heterochromatin and euchromatin on human artificial chromosomes. Genome Biol. *5*, R89.

Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K., and Willard, H.F. (1997). Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat. Genet. *15*, 345–355.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. (1998). Construction of YAC-based mammalian artificial chromosomes. Nat. Biotechnol. *16*, 431–439.

Janssen, C.S., Phillips, R.S., Turner, C.M., and Barrett, M.P. (2004). Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites. Nucleic Acids Res. *32*, 5712–5720.

Kelly, J.M., McRobert, L., and Baker, D.A. (2006). Evidence on the chromosomal location of centromeric DNA in Plasmodium falciparum from etoposide-mediated topoisomerase-II cleavage. Proc. Natl. Acad. Sci. USA *103*, 6706–6711.

Kooij, T.W., Carlton, J.M., Bidwell, S.L., Hall, N., Ramesar, J., Janse, C.J., and Waters, A.P. (2005). A Plasmodium whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. PLoS Pathog. *1*, e44. 10.1371/journal.ppat.0010044.

Lopez-Rubio, J.J., Mancio-Silva, L., and Scherf, A. (2009). Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe *5*, 179–190.

Morris, C.A., and Moazed, D. (2007). Centromere assembly and propagation. Cell *128*, 647–650.

Murray, A.W., and Szostak, J.W. (1983). Construction of artificial chromosomes in yeast. Nature *305*, 189–193.

O'Donnell, R.A., Preiser, P.R., Williamson, D.H., Moore, P.W., Cowman, A.F., and Crabb, B.S. (2001). An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected Plasmodium falciparum parasites. Nucleic Acids Res. *29*, 716–724.

Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B controls centromere formation depending on the chromatin context. Cell *131*, 1287–1300.

Pace, T., Ponzi, M., Dore, E., and Frontali, C. (1987). Telomeric motifs are present in a highly repetitive element in the Plasmodium berghei genome. Mol. Biochem. Parasitol. *24*, 193–202.

Pace, T., Scotti, R., Janse, C.J., Waters, A.P., Birago, C., and Ponzi, M. (2000). Targeted terminal deletions as a tool for functional genomics studies in Plasmodium. Genome Res. *10*, 1414–1420.

Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G., and Earnshaw, W.C. (1995). The centromere: hub of chromosomal activities. Science *270*, 1591–1594.

Ponzi, M., Pace, T., Dore, E., and Frontali, C. (1985). Identification of a telomeric DNA sequence in Plasmodium berghei. EMBO J. *4*, 2991–2995.

Ponzi, M., Pace, T., Dore, E., Picci, L., Pizzi, E., and Frontali, C. (1992). Extensive turnover of telomeric DNA at a Plasmodium berghei chromosomal extremity marked by a rare recombinational event. Nucleic Acids Res. *20*, 4491–4497.

Rudzinska, M.A. (1969). The fine structure of malaria parasites. Int. Rev. Cytol. *25*, 161–199.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998). Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intraerythrocytic development in Plasmodium falciparum. EMBO J. *17*, 5418–5426.

Tschumper, G., and Carbon, J. (1983). Copy number control by a yeast centromere. Gene *23*, 221–232.

van Dijk, M.R., Vinkenoog, R., Ramesar, J., Vervenne, R.A., Waters, A.P., and Janse, C.J. (1997). Replication, expression and segregation of plasmid-borne DNA in genetically transformed malaria parasites. Mol. Biochem. Parasitol. *86*, 155–162.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G.R., Janse, C.J., Waters, A.P., Kato, T., and Kaneko, I. (2009). Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. Mol. Microbiol. *71*, 1402–1414.

Molecular Microbiology (2009) 71(6), 1402–1414 ■ doi:10.1111/j.1365-2958.2009.06609.x
First published online 11 February 2009

Identification of a transcription factor in the mosquito-invasive stage of malaria parasites

Masao Yuda,[1]* Shiroh Iwanaga,[2] Shuji Shigenobu,[3] Gunnar R. Mair,[4] Chris J. Janse,[4] Andrew P. Waters,[5] Tomomi Kato[1] and Izumi Kaneko[1]
[1]*Department of Medical Zoology, Mie University School of Medicine, Mie, Tsu, 514-0001, Japan.*
[2]*Department of Medical Zoology, Tottori University School of Medicine, Tottori, Yonago, Japan.*
[3]*Okazaki Institute for Integrative Bioscience, National Institute for Basic Biology, National Institutes of Natural Sciences, Higashiyama, Myodaiji, Okazaki, Japan.*
[4]*Department of Parasitology, Leiden University Medical Centre, 2333 ZA Leiden, the Netherlands.*
[5]*University of Glasgow, Glasgow Biomedical Research Centre, Glasgow, Scotland, UK.*

Summary

Gene expression in *Plasmodium* parasites undergoes significant changes in each developmental stage, but the transcription factors (TFs) regulating these changes have not been identified. We report here a *Plasmodium* TF (AP2-O) that activates gene expression in ookinetes, the mosquito-invasive form, and has a DNA-binding domain structurally related to that of a plant TF, Apetala2 (AP2). AP2-O mRNA is presynthesized by intraerythrocytic female gametocytes and translated later during ookinete development in the mosquito. The *Plasmodium* TF activates a set of genes, including all genes reported to be required for midgut invasion, by binding to specific six-base sequences on the proximal promoter. These results indicate that AP2 family TFs have important roles in stage-specific gene regulation in *Plasmodium* parasites.

Introduction

Elucidation of gene regulation mechanisms in the parasitic protozoan *Plasmodium* could provide important information for understanding parasite developmental programmes and its interactions with both host and vector during infection at each infection stage and could support the development of new antimalarial strategies. However, at present, little is known about the mechanisms of gene regulation in *Plasmodium* and only a small number of transcription factors (TFs), which may bind to specific DNA promoter sequences and control gene expression, have been predicted in the *Plasmodium* genome (Iyer *et al.*, 2008).

At present the Apetala2 (AP2) family is the only known gene family predicted to encode TF candidates in the *Plasmodium* genome (Balaji *et al.*, 2005). The AP2 family TFs were first identified in plants (Jofuku *et al.*, 1994) and then shown to form a large family in the genome of several plants (Gutterson and Reuber, 2004). In *Arabidopsis thaliana*, the AP2 family contains at least 140 members, some of which are involved in stress responses and others in regulation of reproductive and vegetative organ development (Gutterson and Reuber, 2004). Plant AP2 family TFs are named for their possession of at least one common DNA-binding domain of approximately 60 amino acids, the AP2 domain. A bioinformatic analysis of the complete *Plasmodium falciparum* genome revealed that AP2-related genes are present in apicomplexan genomes, including *Plasmodium* (Balaji *et al.*, 2005), where in the human malaria parasite *Plasmodium falciparum*, 26 AP2-related genes were predicted. The predicted *Plasmodium* AP2-related genes encode proteins with one to four AP2 domains; however, their function remains to be elucidated.

Plasmodium has three host-invasive stages: merozoites, ookinetes and sporozoites, which are erythrocyte-, mosquito midgut- and salivary gland/Kuppfer cell/hepatocyte-invasive stages respectively. These stages produce various stage-specific proteins involved in subsequent host invasion, some of which have been identified through various transcriptome and proteome analyses (Kaiser *et al.*, 2004; Hall *et al.*, 2005; Raibaud *et al.*, 2006). Each invasive form of the parasite produces a distinct repertoire of proteins that are secreted from apical secretory organelles such as micronemes and rhoptries and associated with invasion of host cells or tissues (Dubremetz *et al.*, 1998) or associated with the cell surface, suggesting that they participate in specific host–parasite interactions. However, regulatory mechanisms of individual or groups of stage-specific genes remain elusive in all these stages.

Accepted 12 January, 2009. *For correspondence. E-mail m-yuda@doc.medic.mie-u.ac.jp; Tel. (+81) 59 231 5430; Fax (+81) 59 231 5430.

During asexual replication in animal host erythrocytes, some *Plasmodium* parasites differentiate into sexual forms, called gametocytes. When ingested by a mosquito with a blood meal, gametocytes generate gametes that are fertilized and form zygotes in the mosquito midgut lumen. The zygotes develop into motile ookinetes that invade the midgut epithelium and differentiate into oocysts on the basal side of the midgut. The developing ookinetes produce large amount of stage-specific proteins for subsequent midgut invasion (Vlachou *et al.*, 2006) including microneme proteins such as circumsporozoite protein thrombospondin-related anonymous protein (TRAP)-related protein (CTRP), a protein belonging to the TRAP family and essential for gliding motility (Dessens *et al.*, 1999; Yuda *et al.*, 1999a); chitinase, a chitinolytic enzyme necessary for ookinete crossing of the mosquito chitin-containing peritrophic matrix (Vinetz *et al.*, 2000); membrane attack ookinete protein (MAOP), essential for rupture of mosquito midgut epithelial cell membrane (Kadota *et al.*, 2004); a protein kinase CDPK3 (calcium-dependant protein kinase 3) that stimulates invasion (Ishino *et al.*, 2006); cell surface-associated proteins, P28 and P25, which are necessary for ookinete development and are transmission-blocking vaccine targets (Tomas *et al.*, 2001; Saul, 2007); and other secreted proteins (Yuda *et al.*, 2001; Dessens *et al.*, 2003). The expression of some ookinete-specific proteins is regulated at the translational level where the female *Plasmodium* gametocyte represses specific transcripts through a protein complex composed of DOZI (development of zygotes inhibited), a conserved DDX6-class RNA helicase, and other proteins (Mair *et al.*, 2006). After fertilization, these gene transcripts are released from the complex and enter the translation pathway. On the other hand, *de novo* mRNA synthesis of invasion-related genes also occurs after fertilization (Raibaud *et al.*, 2006), and full-scale production of invasion-related proteins begins several hours after fertilization in developing ookinetes.

During a systematic investigation of the predicted AP2 factors that are not expressed in asexual blood-stage parasites we discovered one we designate AP2-O (AP2 in ookinetes) that is essential for the formation of invasive ookinetes and the expression of ookinete invasion-related genes. Here, we report that AP2-O is the ookinete TF that directly activates the invasion-related genes by binding to a specific motif found upstream of them.

Results

Expression of AP2-O *is regulated at translational level*

Plasmodium berghei AP2-O (PB000572.01.0) encodes a protein with a single AP2 domain and has an orthologue in *P. falciparum* (PF11_0442, Fig. 1A) and other *Plasmodium* spp. AP2-O has two conserved regions which include the AP2 domain near the C-terminus. The amino acid sequences of the AP2-O AP2 domain are almost identical among *Plasmodium* spp. (Fig. S1).

To investigate *AP2-O* protein expression, a transgenic line of *P. berghei* that expressed green fluorescent protein (GFP) fused to the endogenous allele of *AP2-O* (AP2-O::GFP parasites) was generated (Fig. 1B). The fusion did not appear to affect the parasite's ability to infect mosquitoes, as the AP2-O::GFP parasite generated a normal number of oocysts in mosquitoes (Table 1). In the blood stages of AP2-O::GFP parasites, fluorescence of the GFP-fused protein was not observed (Fig. 1C). However, weak fluorescent signals were observed in retort-form ookinetes beginning 8 h after fertilization (Fig. 1C). The signals were localized in the nucleus and signal intensities increased with ookinete development.

We also compared the amount of *AP2-O* transcripts between gametocytes and ookinetes with real-time reverse transcription PCR (RT-PCR) analysis using wild-type (WT) asexual-free parasite preparations (Fig. 1D). Unexpectedly the analysis showed that *AP2-O* transcripts are most abundant in the gametocyte stage and markedly decreased in the subsequent zygote and ookinete stages. This expression profile was in contrast with that of the typical ookinete microneme protein, secreted ookinete adhesive protein (SOAP), whose transcripts are mainly synthesized in the zygote and ookinete stages (Dessens *et al.*, 2003). These results suggest that *AP2-O* expression is translationally repressed in female gametocytes and subsequently translated in ookinetes (Mair *et al.*, 2006).

We examined whether *AP2-O* transcripts form physical complexes with DOZI in gametocytes using immunoprecipitation (IP) experiments on DOZI::GFP gametocytes (Fig. 1E). *AP2-O* transcripts were co-precipitated with GFP-tagged DOZI with anti-GFP antibodies. These results demonstrated that the translation repression system of the DOZI complex controls expression of AP2-O mRNA.

AP2-O *disruption decreases expression of several mosquito invasion-related genes*

To investigate the function of *AP2-O*, we prepared *P. berghei* mutant parasites with the *AP2-O*-disrupted [*AP2-O* (–)] (Fig. 2A). The *AP2-O* (–) parasites formed morphologically normal female and male gametocytes and exhibited normal exflagellation rates (Table 1). However, they lacked the ability to infect mosquitoes as no oocysts or sporozoites were found in the mosquito midgut. In culture, *AP2-O* (–) parasites formed zygotes at normal conversion rates and differentiated into retort forms by 9 h after fertilization, as in WT parasites (Fig. 2B, top). Subtle morphological differences between WT and

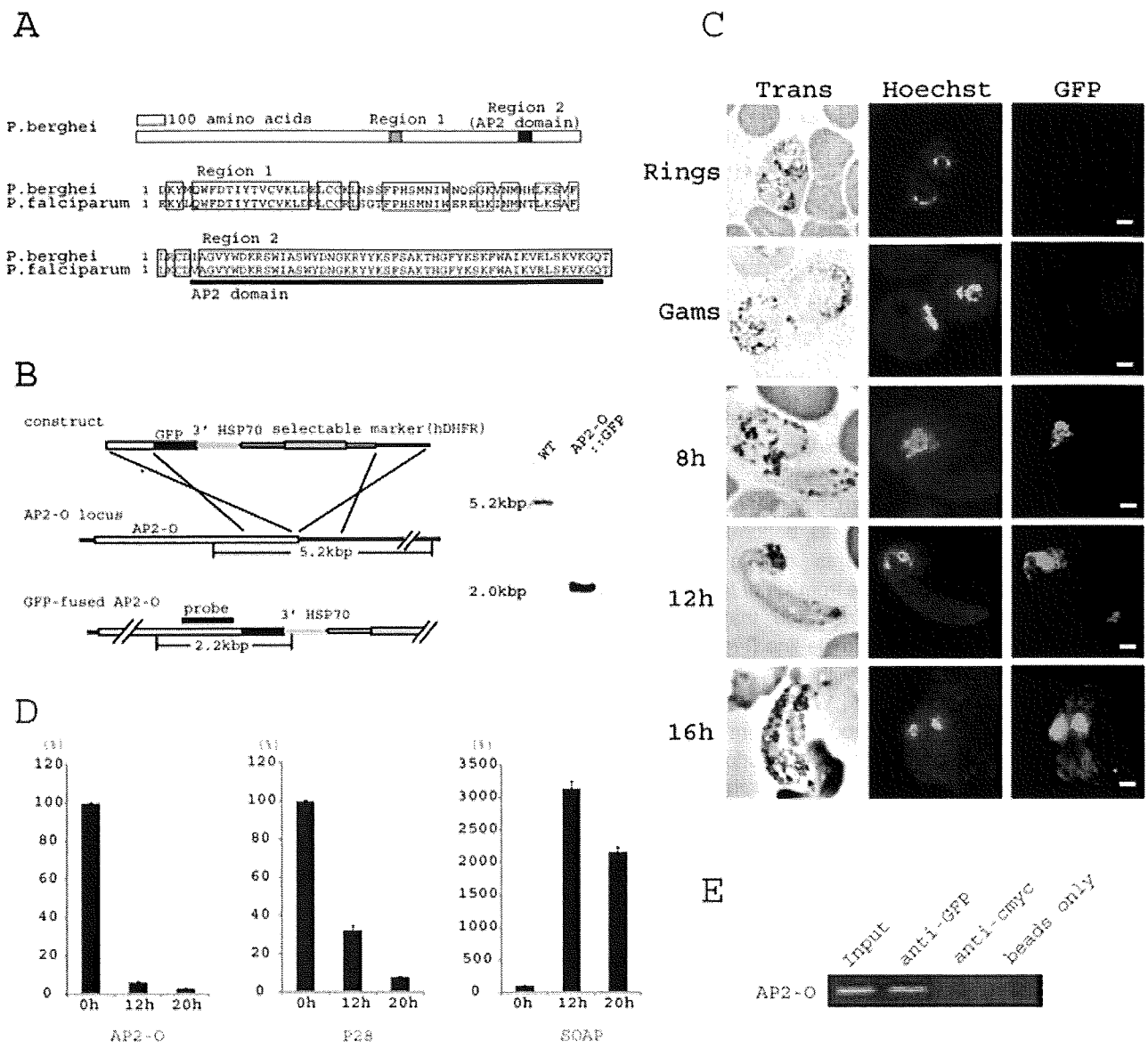


Fig. 1. Expression of AP2-O is controlled at the translational level.
A. The amino acid sequences around the single AP2 domain and another conserved region (region 1) were aligned between *Plasmodium berghei* and *P. falciparum*. Identical amino acids are highlighted by rectangles.
B. To construct AP2-O::GFP parasites, the targeting vector containing a marker gene (left top), human DHFR (hDHFR), was integrated into the *AP2-O* locus (left middle) by double-cross-over recombination, adding the GFP gene to the C-terminal portion of *AP2-O* and conferring pyrimethamine resistance to the AP2-O::GFP parasites (left bottom). For Southern hybridization (right), WT and AP2-O::GFP genomic DNA was digested with MunI and hybridized with a probe (solid bar; bottom left). Integration decreased detected fragment size from 5.2 to 2.0 kb.
C. To localize AP2-O *in vivo*, AP2-O::GFP parasites were cultured to the ookinete stage. At 8, 12 and 16 h after fertilization, nuclei were stained with Hoechst 34580. Images of blood-stage parasites were also shown. Rings, ring forms; Gams, gametocytes. Scale bars, 2 μm.
D. Total RNA was prepared from asexual-free blood-stage parasites (0 h), retort-form ookinetes (12 h after fertilization) and mature ookinetes (20 h after fertilization) cultured from the same infected blood. Quantitative RT-PCR was performed with primers for *AP2-O*, *P28* and *SOAP*. Each result is the mean of three independent experiments with standard error bars. For each gene, the amount of mRNA at each time was expressed as the per cent of mRNA relative to that at 0 h after fertilization. P28 mRNA is abundant in gametocytes, while SOAP mRNA is mainly synthesized in ookinetes. *AP2-O* showed an expression pattern different from that of *SOAP*.
E. For analysis of complex formation of *AP2-O* mRNA and the translation repression complex, IP was performed on gametocyte lysate of DOZI::GFP parasites using anti-GFP antibodies, anti-myc antibodies and protein G Sepharose beads only. RNAs recovered from IPs were analysed by RT-PCR with primers for *AP2-O*.

Table 1. *AP2* (–) malaria parasites lack the ability to infect mosquitoes.[a]

Genotype	Exflagellation rate	Conversion rate[b]	Number of oocysts per mosquito[c]	Number of sporozoites per mosquito[d]
WT	44.6	63.8 (6.8)	76.1 (14.73)	40 604 (6957)
AP2 (–) 1	37.5	56.8 (11.9)	0	0
AP2 (–) 2	24.3	71.0 (2.3)	0	0
AP2-O::GFP	35	70.5 (1.12)	71.6 (20.85)	42 940 (6797)

a. Infected mice were subjected to mosquito bites when exflagellation rates became more than 20 per 10^5 red blood cells. Fully engorged mosquitoes were dissected after 14 days, and the numbers of oocysts and sporozoites in the midgut were counted.
b. Conversion rates (%) of female gametocytes to retort-form ookinetes (standard errors in parentheses).
c. Twenty mosquitoes were dissected, and the number per mosquito was calculated (standard errors in parentheses).
d. Ten mosquitoes were dissected, and the number per mosquito was calculated (standard errors in parentheses).

AP2-O (–) parasites, however, became visible 12 h after fertilization with a swelling protrusion in the retort forms (Fig. 2B, centre). Final development was aberrant ookinetes with an pear-shaped form, different from the sword-like form of WT ookinetes (Fig. 2B, bottom).

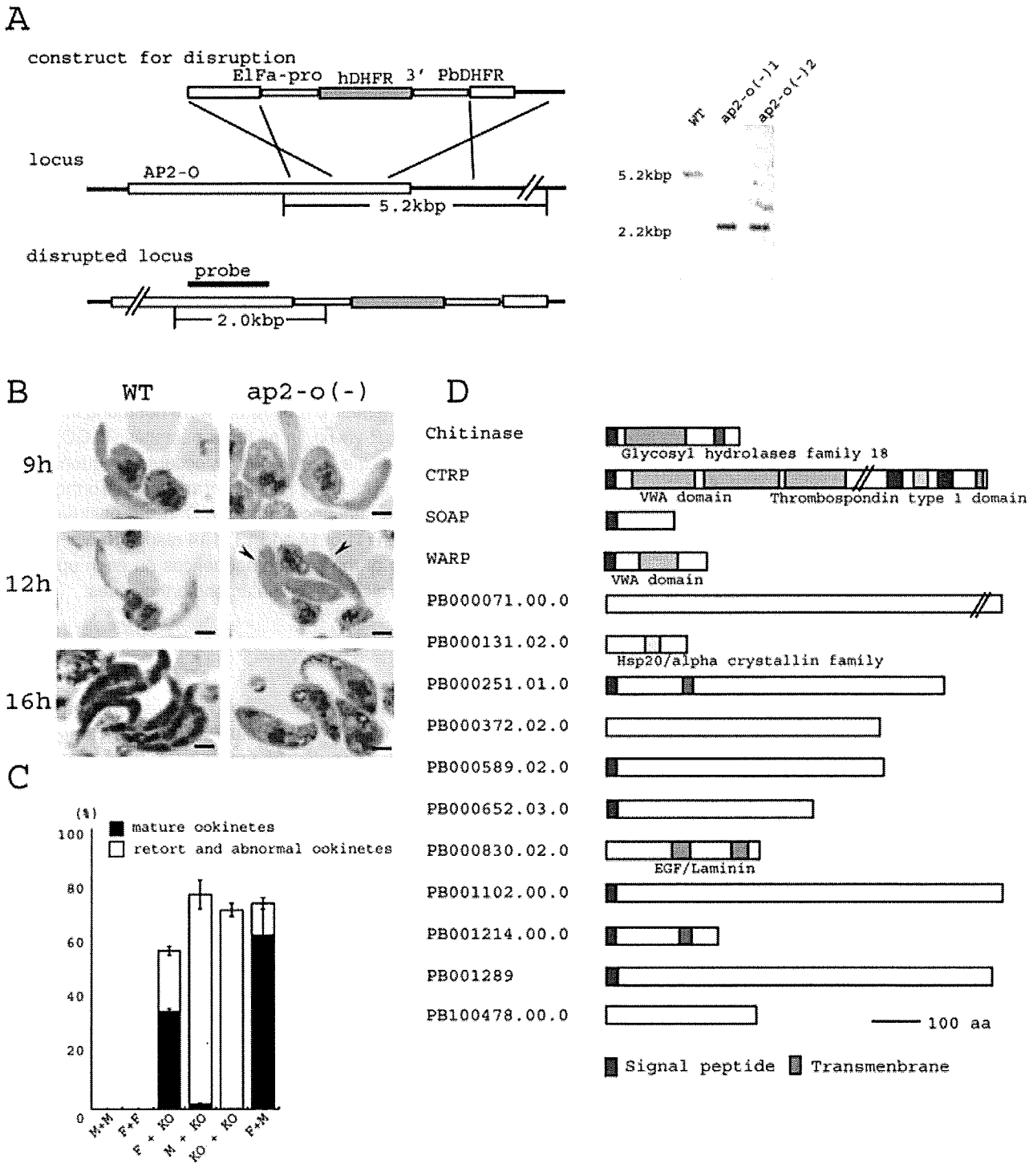
Cross-fertilization experiments mating *AP2-O* (–) mutant gametocytes with those from either female or male gametocyte-defective lines indicated that the *AP2-O* defect is inherited from the female line (Fig. 2C). *AP2-O* (–) females when fertilized with WT males produced deformed ookinetes. However, WT females when fertilized by *AP2-O* (–) males produced normal ookinetes. These results suggest that *AP2-O* transcripts in zygotes/ookinetes are derived from female gametocytes and essential for zygote development into normal infective ookinetes.

DNA microarray analyses were performed to investigate whether *AP2-O* disruption affects gene expression in ookinetes. Samples were obtained from WT and *AP2-O* (–) retort-form ookinetes at 12 h after fertilization, when slight morphological differences between them started to appear (Fig. 2B, centre). Genes were selected whose expression was decreased at least fivefold (Fig. 2D). A total of 15 genes were identified with this analysis, of which four were ookinete-specific genes (all encoding microneme proteins) (Yuda *et al.*, 1999b; 2001; Vinetz *et al.*, 2000; Dessens *et al.*, 2003), six were genes reported to be transcribed during ookinete development after zygote formation (Raibaud *et al.*, 2006) and five were novel genes not ascribed to. We performed real-time RT-PCR analysis in the five novel genes and showed that they were transcribed in the zygote/ookinete stage (Fig. S2). Very recently two of them (PB000652.03.0 and PB001214.00.0) were reported to be involved in ookinete midgut invasion (Ecker *et al.*, 2008). As shown in Table S1 and Fig. 2D, 14 of these 15 proteins have orthologous genes in *P. falciparum* and *P. vivax*, and 10 have a putative N-terminal signal peptide sequence, suggesting that they are microneme or cell surface-associated proteins that are available for subsequent midgut invasion.

AP2-O directly activates several genes by binding to a six-base sequence, TAGCTA

To investigate which of the 15 genes AP2-O regulates directly, chromatin immunoprecipitation (ChIP) assays were performed on AP2-O::GFP parasites (Fig. 3A and Fig. S3). The upstream promoter region of all genes was significantly enriched by ChIP with anti-GFP antibody. Enrichment was not observed in identical assays performed on WT parasites. These results indicated that AP2-O regulates all 15 genes directly and that there is at least one binding site for AP2-O in the upstream region of each gene. Therefore, we looked for putative AP2-O binding sites in these 15 genes by determining high-frequency sequences in the 1 kb immediate upstream regions of these genes using a computational analysis (Table S2). The sequence conservation pattern (Fig. 3B) shows a 6 bp sequence, TAGCTA, as the most frequently occurring sequence. This sequence was present in the upstream region of 14 of the 15 genes, with two or more TAGCTA sequences observed in 13 genes (Fig. 3C). In addition to TAGCTA, four other closely related sequences, TGGCTA/TAGCCA and CAGCTA/TAGCTG, are possible binding sites for AP2-O (see Fig. 3E). Including the less frequent binding elements, TGGCTA/TAGCCA, two or more binding sites were present in all 15 genes. Frequencies of CAGCTA/TAGCTG were significantly low compared with the other three binding sequences (Table S2). The majority of the binding sequences were adjacent (approximately 100–400 bp upstream) to the putative transcription start site as estimated from the most upstream expressed sequence tags (ESTs) of each gene (Fig. 3C).

Electrophoretic mobility shift assays (EMSA) were performed to examine whether the 6 bp sequences act as binding sites for AP2-O. The ~300 bp *SOAP* upstream region, which contains two TAGCTA sites, was used as the probe (Fig. 3D). The AP2 domain of AP2-O bound to the probe and produced a single band shift. Addition of a point mutation (TAGCTA to TAGGTA) in one of the two TAGCTA sites significantly reduced the intensity of the



shifted band, and addition of the mutation in both sites eliminated the band shift. The same results were obtained using the ~300 bp upstream region of the chitinase gene (Fig. S4A).

To further analyse the binding specificity of the AP2 domain, we performed EMSA using probes with various point mutations in the TAGCTA sequence. Of the 18 mutations examined, 14 nearly eliminated binding and 4 (TGGCTA/TAGCCA and CAGCTA/TAGCTG) moderately reduced band shift intensity (Fig. 3E). These results demonstrated that AP2-O preferentially binds to TAGCTA, but also binds with weaker affinity to the similar sequences.

Fig. 2. AP2-O is necessary for normal development of ookinetes and expression of invasion-related genes.
A. To construct *AP2-O* (–) parasites, a targeting vector (left top) containing a selectable marker gene was integrated into the *AP2-O* locus (left middle) by double-cross-over recombination resulting in disruption of the *AP2-O* gene and conferring pyrimethamine resistance to the *AP2-O* (–) mutants (left bottom). For Southern hybridization (right), WT and *AP2-O* (–) genomic DNA was digested with MunI and hybridized with a probe (solid bar; left bottom). Integration decreased fragment size from 5.2 to 2.2 kb. Independently prepared *AP2-O* (–) mutants, *AP2-O* (–) 1 and *AP2-O* (–) 2, exhibited the same phenotype.
B. Phenotype of *AP2-O* (–) parasites in Giemsa-stained cultured ookinetes at different times after fertilization. A morphological difference between WT and *AP2-O* (–) parasites apparent 12 h after fertilization was the swelling protrusion (arrowheads) in *AP2-O* (–) retort-form ookinetes. *AP2-O* (–) parasites developed into pear-shaped ookinetes 16 h after fertilization (lower right). Scale bars, 2 μm.
C. Cross-fertilization experiments were performed with *AP2-O* (–) parasites, a male-deficient line (*P48/45*$^-$) and a female-deficient line (*P47*$^-$). Conversion rates of female gametocytes to deformed ookinetes and mature normal ookinetes are shown. The abnormal ookinetes observed in the combination of *AP2-O* (–) mutants and a male-defective line (*P48/45*$^-$) may have derived from self-fertilization of *AP2-O* (–) mutants. M, female-deficient line (*P47*$^-$); F, a male-deficient line (P48/45$^-$); KO, *AP2-O* (–) parasites.
D. Names or IDs of 15 *P. berghei* genes whose expression was decreased at least fivefold in *AP2-O* (–) ookinetes compared with WT (left), and the structure of their encoded proteins (right).

TAGCTA as a cis*-acting element in the ookinete stage*

Reporter assays were carried out to investigate whether the TAGCTA sequence functions as a *cis*-acting element *in vivo*. For these assays a plasmid containing a *P. berghei* centromere was used (Fig. S4B). This plasmid acts as an artificial chromosome and is maintained stably through several cell divisions in transfected parasites (S. Iwanaga *et al.*, unpubl. results). The upstream region of the SOAP gene, containing three TAGCTA sequences, was inserted upstream of the GFP reporter gene, and *P. berghei* blood-stage parasites were transfected with this construct (Fig. 4). GFP signals were not observed in the intraerythrocytic stages, including gametocytes. In ookinete cultures, GFP signals were detected 8 h after fertilization and signal intensity increased with ookinete development (data not shown), indicating that the upstream region acts as a stage-specific promoter. When one of the three TAGCTA sequences was disrupted by addition of a point mutation (TAGCTA to TAGGTA), promoter activity reduced to 40% of that of the non-mutated sequence; mutations in two of the TAGCTA sequences reduced promoter activity to < 10% of the non-mutated sequences. Similar results were obtained with the ~300 bp upstream region of the chitinase gene (Fig. S4C). These results indicate that the TAGCTA sequence functions as a *cis*-acting element *in vivo* and that the number of elements is an important factor in determining promoter activity.

cis*-acting elements are conserved among* Plasmodium *spp.*

The amino acid sequence of AP2 domain of AP2-O is highly conserved among several *Plasmodium* spp. (Fig. S1). Conservation of the amino acid sequence in the AP2 domain suggests that sequences recognized by this domain would be conserved among *Plasmodium* spp., and thus the same *cis*-acting elements would be observed in the orthologues of the 15 genes. To examine this, we performed computational analysis on *P. falciparum* and *P. vivax* orthologues (Tables S3 and S4, and Fig. 5A; 14 orthologues were used because orthologues of one *P. berghei* gene, PB100478.00, were not found in *P. falciparum* and *P. vivax*). In all three parasites, TAGCTA was conserved as the most frequently occurring sequence. The great majority of the *P. falciparum* and *P. vivax* orthologues had binding sequences for AP2-O in the upstream promoter region (11 of 14 and 13 of 14 respectively; Fig. 5B). Distribution of the binding sequences on each upstream region was also conserved among orthologues.

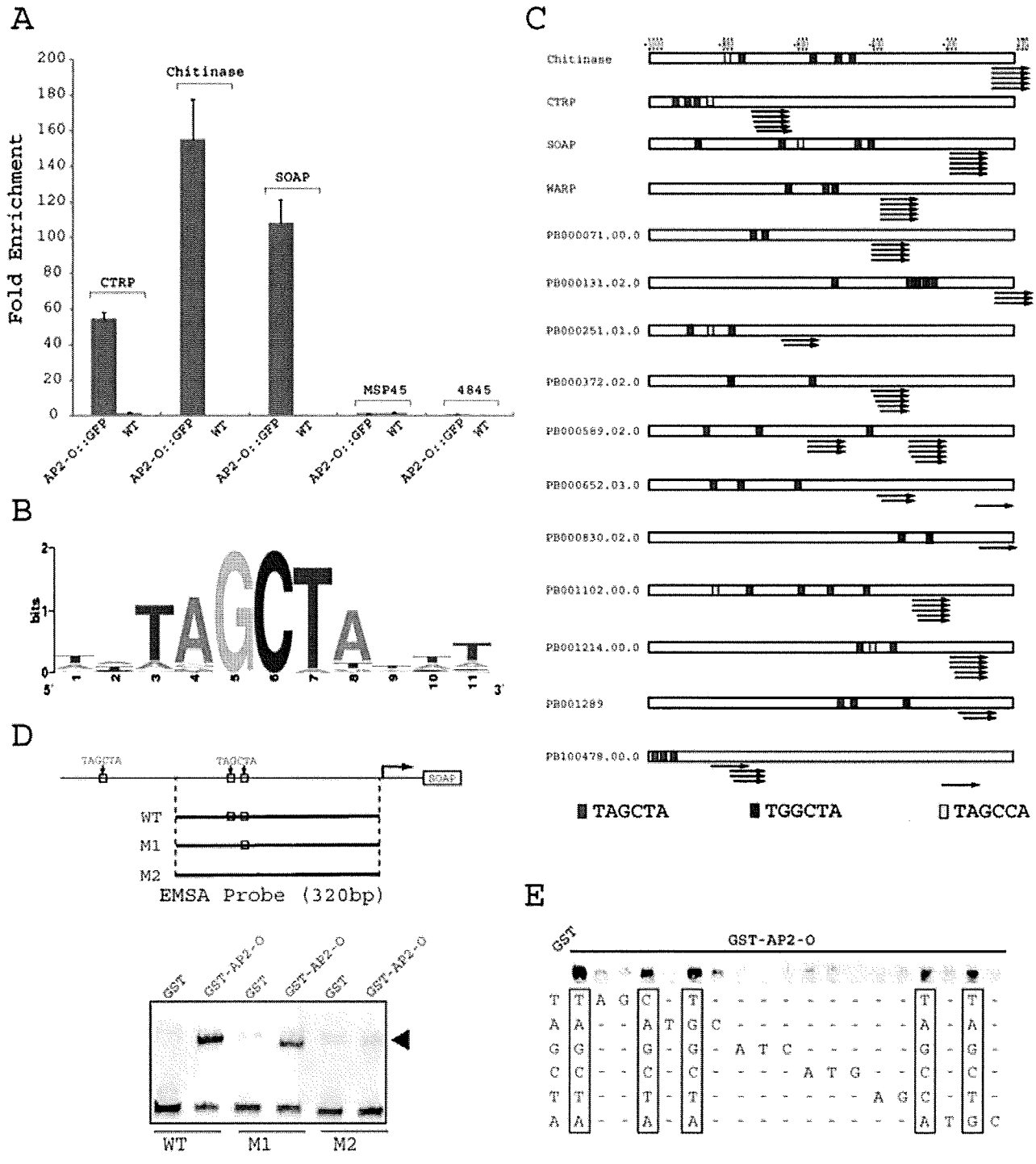
AP2-O activates all reported ookinete stage-specific genes

We analysed 15 genes whose expression significantly decreased in *AP2-O* (–) mutants. However, other genes, demonstrated or suggested to be involved in midgut invasion, *P25*, *P28*, *CDPK* and *MAOP* (Tomas *et al.*, 2001; Kadota *et al.*, 2004; Ishino *et al.*, 2006), were not in that initial group. To investigate whether these genes are regulated by AP2-O in the ookinete stage, we performed ChIP assays (Fig. 6A). The results demonstrated that AP2-O is associated with the upstream region in each of the four genes. Furthermore, two or more AP2-O binding sites were identified in the upstream region of each gene (Fig. 6B). These results indicate that these genes are activated by AP2-O in the ookinete stage.

Discussion

Global surveys of gene expression indicated that a high proportion of *Plasmodium* genes are expressed in a stage-specific manner (Le Roch *et al.*, 2003). However, how these changes are regulated in the life cycle remains elusive. In this study, we showed that specific gene expression in the ookinete stage is regulated by a sequence-specific TF belonging to the AP2 family, AP2-O. This study provided the first experimental evidence that gene expression in plasmodium is regulated by specific TF.

Ookinetes express several genes necessary for midgut invasion during development in mosquito midgut lumen.



Gene expression begins at 8–10 h after fertilization and continues for approximately 10 h until midgut invasion starts (Yuda *et al.*, 1999b; 2001). Our results showed that AP2-O is expressed during this period and activates several genes by directly binding to their upstream region. The activated genes encode different types of proteins including secretory proteins, surface-associated proteins and cytoplasmic proteins and include all known ookinete-specific genes involved in midgut invasion. Therefore, AP2-O would be a master activator that induces a cluster of different genes necessary for midgut invasion.

Our reporter assays indicated that the number of binding sites for AP2-O on the upstream promoter may be a factor defining promoter activities in the ookinete stage.

Fig. 3. AP2-O binds specifically to six-base sequences on the proximal promoter region of ookinete stage-specific genes.
A. For analysis of AP2-O complexes with the genes in Fig. 2D, ChIP was performed in AP2-O::GFP ookinetes. Of the 15 genes listed in Fig. 2D, the results for *CTRP*, chitinase and *SOAP* and for the negative controls, *P48/45* and *MSP4/5*, which are not expressed in ookinetes, are shown. Values are fold enrichments relative to IPs with control antibodies. Control experiments with wild-type parasites (WT) were also performed. Three biologically independent ChIP experiments were performed, and essentially the same results were obtained. The results of one of these experiments were shown here. Each result is the mean of three independent quantitative PCR experiments with standard error bars. The results for several other *Plasmodium* spp. genes are shown in Fig. S2.
B. WebLogo diagram of sequences occurring at high frequency in the 1 kb upstream regions of the *P. berghei* genes listed in Fig. 2D. Graphical representations created by WebLogo (Crooks *et al.*, 2004).
C. AP2-O binding sequences and the furthest upstream five 5′-ESTs (arrows) in the 1 kb upstream regions of the *P. berghei* genes listed in Fig. 2D. In each gene, more than two AP2-O binding sequences are present 100–400 bp upstream from the most upstream EST cluster. Coloured rectangles indicate binding sequence locations.
D. Sequence specificity of AP2-O binding was detected by EMSA using a GST-fused AP2-O AP2 domain. The *SOAP* upstream region containing two TAGCTA sequences was used as the probe. Probes with one (M1) or two (M2) mutations in their TAGCTA sequence (TAGCTA to TAGGTA) were also used. TAGCTA sequences in the upstream region of WT (no mutation) and the two mutations are illustrated (top). GST was used as a negative control. In the EMSA gel (bottom), the arrowhead indicates the band shift.
E. Other sequence possibilities in the AP2 domain binding site. Various point mutations were added to the TAGCTA sequence, and binding to AP2-O was determined by EMSA. The added point mutations are indicated under each lane. Binding ability was lost or greatly reduced in most mutated probes. Those mutations retaining binding activity are highlighted by a rectangle.

The AP2-O binding sites are mainly located within a short 100–400 bp region from the transcription start site (Fig. 3C), suggesting that to activate target genes, AP2-O must bind to the region just upstream of the transcription initiation site.

The merozoite and sporozoite stages also produce large amounts of invasion-related proteins during the development to mature invasive forms. Our results suggest that AP2 family TFs regulate invasion-related genes in these stages of *Plasmodium* spp. In fact, this possibility is supported by some preceding studies. For example, in the asexual blood stages, genes for erythrocyte invasion are mainly expressed in the mid- to late-schizont stage, i.e. the stage preceding merozoites (Bozdech *et al.*, 2003). The expressed genes encode proteins necessary for erythrocyte invasion such as microneme and rhoptry proteins as well as several surface-associated proteins. Because at this period some AP2 family genes exhibit peak expression, these AP2 family genes might be TFs inducing these invasion-related genes (Balaji *et al.*, 2005). Recent study showed that one of such AP2-related protein binds to a specific sequence and that many invasion-related genes have this motif in the upstream (De Silva *et al.*, 2008), suggesting that the AP2-related gene encodes an activator of these genes. Although this suggestion should be tested further, it is compatible with our present finding that an AP2-related protein directly controls stage-specific expression of multiple genes in ookinetes. Also in sporozoites, several invasion-related genes are expressed and some of them are homologous to genes regulated by AP2-O (Yuda and Ishino, 2004). A recent *in silico* analysis reported that sporozoite-specific genes have a common six-base motif CATGCA in the upstream (Young *et al.*, 2008). Therefore, it is possible that this motif is a binding sequence for another activator belonging to the AP2 family. It remains for further study to explore which AP2 family genes are expressed during the formation of infective sporozoites.

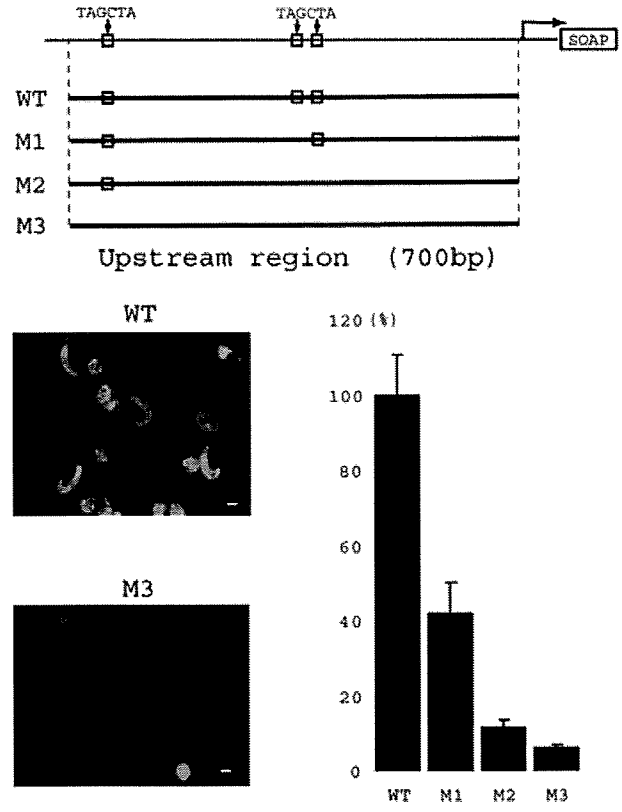


Fig. 4. The TAGCTA sequence as a *cis*-acting element in the ookinete stage. Promoter activity of the TAGCTA sequence in the ookinete stage *in vivo* was analysed using a reporter system with the *P. berghei* centromere plasmid Pcen (see also Fig. S3B). Mutations were added to three TAGCTA motifs in the SOAP upstream region (top). Mutations in M1 and M2 are the same as in M1 and M2 in Fig. 3D. M3 contains mutations in all three TAGCTA sequences. Each result is the mean of three independent experiments with standard error bars and is shown as the activity relative to that of the non-mutated WT. Representative images of GFP expression in ookinetes transfected with WT and M3 Pcen are shown at the bottom left. Round cells with strong fluorescence are mouse reticulocytes. Scale bars, 2 μm.

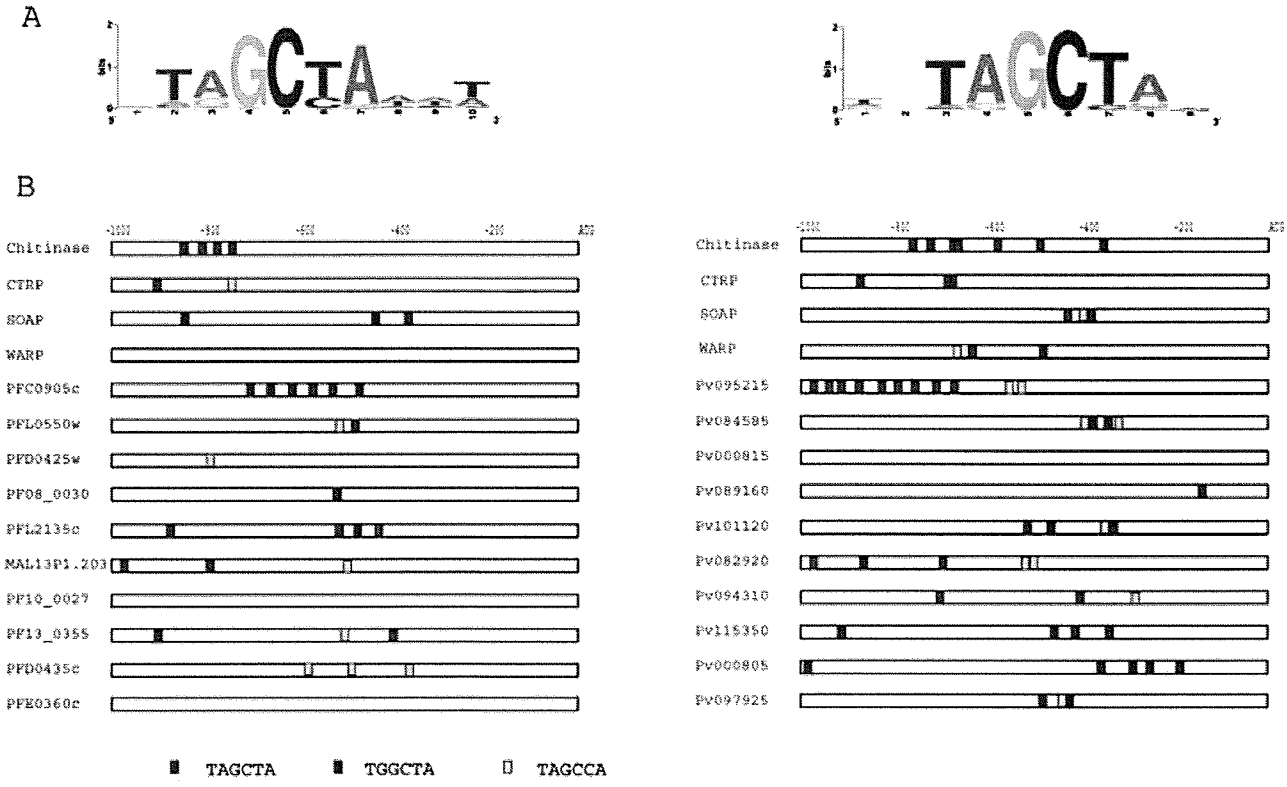


Fig. 5. Binding sequences for *Plasmodium berghei* AP2-O are conserved in *P. falciparum* and *P. vivax*.
A. WebLogo diagrams of sequences occurring at high frequency in the 1 kb upstream regions of the 14 orthologues in Fig. 2D in *P. falciparum* (left) and *P. vivax* (right). Graphical representations created by WebLogo (Crooks *et al.*, 2004).
B. AP2-O binding sequences in the 1 kb upstream regions of the 14 orthologues in Fig. 2D in *P. falciparum* (left) and *P. vivax* (right). Coloured rectangles indicate binding sequence locations.

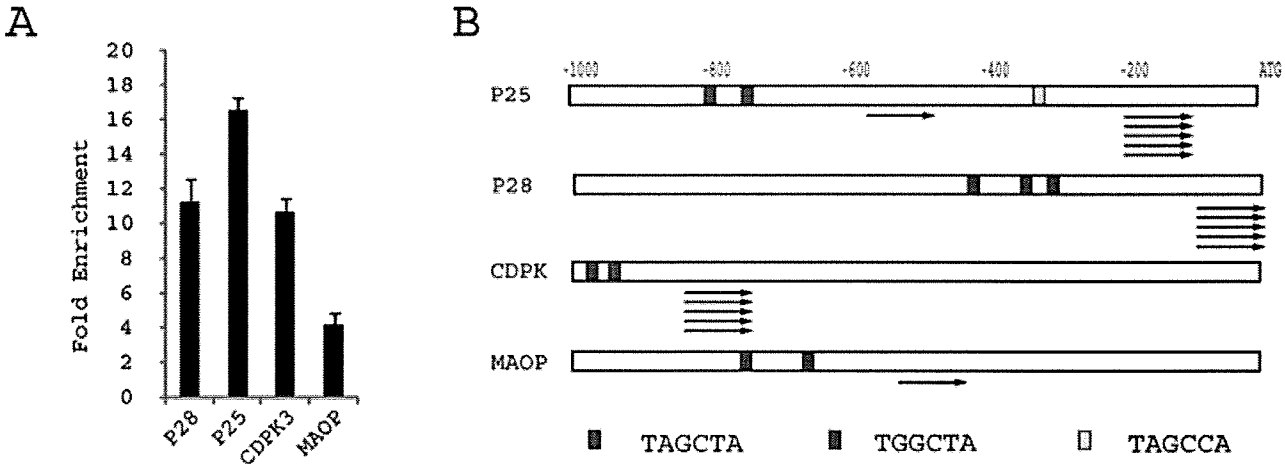


Fig. 6. Four *P. berghei* genes involved in ookinete invasion that are regulated by AP2-O.
A. ChIP was performed in *P28*, *P25*, *CDPK3* and *MAOP*, and DNA fragments were subjected to quantitative PCR with primers for upstream regions of the respective genes. Values are fold enrichments relative to IPs with control antibodies. Each result is the mean of three independent experiments with standard error bars.
B. AP2-O binding sequences and the most upstream 5′-ESTs (arrows) in the 1 kb upstream regions of four *P. berghei* genes are shown here. Each gene had more than two AP2-O binding sequences present 100–300 bp upstream from the most upstream EST cluster. Coloured rectangles indicate binding sequence locations.

