High-Throughput Generation of P. falciparum Functional Molecules by Recombinational Cloning

João Carlos Aguiar, Joshua LaBaer, Peter L. Blair, Victoria Y. Shamailova, Malvika Koundinya, Joshua A. Russell, Fengying Huang, Wenhong Mar, Robert M. Anthony, Adam Witney, Sonia R. Caruana, Leonardo Brizuela, John B. Sacci, Jr., Stephen L. Hoffman and Daniel J. Carucci

Genome Res. 2004 14: 2076-2082
Access the most recent version at doi:10.1101/gr.2416604

Letter

High-Throughput Generation of *P. falciparum* Functional Molecules by Recombinational Cloning

João Carlos Aguiar,[1,3,8] Joshua LaBaer,[2] Peter L. Blair,[1] Victoria Y. Shamailova,[1] Malvika Koundinya,[2] Joshua A. Russell,[1] Fengying Huang,[1] Wenhong Mar,[2] Robert M. Anthony,[1,4] Adam Witney,[1,5] Sonia R. Caruana,[1] Leonardo Brizuela,[2] John B. Sacci Jr.,[1,6] Stephen L. Hoffman,[1,7] and Daniel J. Carucci[1]

[1]*Malaria Program, Naval Medical Research Center, Silver Spring, Maryland 20910, USA;* [2]*Harvard Institute of Proteomics, Cambridge, Massachusetts 02141, USA;* [3]*Pan American Health Organization, Washington, DC 20037, USA*

Large-scale functional genomics studies for malaria vaccine and drug development will depend on the generation of molecular tools to study protein expression. We examined the feasibility of a high-throughput cloning approach using the Gateway system to create a large set of expression clones encoding *Plasmodium falciparum* single-exon genes. Master clones and their ORFs were transferred en masse to multiple expression vectors. Target genes (n = 303) were selected using specific sets of criteria, including stage expression and secondary structure. Upon screening four colonies per capture reaction, we achieved 84% cloning efficiency. The genes were subcloned in parallel into three expression vectors: a DNA vaccine vector and two protein expression vectors. These transfers yielded a 100% success rate without any observed recombination based on single colony screening. The functional expression of 95 genes was evaluated in mice with DNA vaccine constructs to generate antibody against various stages of the parasite. From these, 19 induced antibody titers against the erythrocytic stages and three against sporozoite stages. We have overcome the potential limitation of producing large *P. falciparum* clone sets in multiple expression vectors. This approach represents a powerful technique for the production of molecular reagents for genome-wide functional analysis of the *P. falciparum* genome and will provide for a resource for the malaria resource community distributed through public repositories.

[Supplemental material is available online at www.genome.org. All entry clones from both Panels A and B have been deposited at the Malaria Research and Reference Reagent Resource Center (http://www.malaria.mr4.org) for distribution to researchers upon request. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: Y. Charoenvit and P. de la Vega.]

The entire genomic sequence of the human malaria parasite *Plasmodium falciparum* has been completed (Gardner et al. 2002); two of the 14 chromosomes were published earlier (Gardner et al. 1998; Bowman et al. 1999). In addition, the genome of the rodent malaria parasite *Plasmodium yoelii* has also been sequenced (Carlton et al. 2002), and sequencing of *Plasmodium vivax*, the other important human malaria species, is well underway (Carlton 2003). In addition, a significant number of genes and/or gene fragments have been sequenced through GST, EST, and whole-genome shotgun projects from other *Plasmodium* species including *Plasmodium berghei*, *Plasmodium c. chabaudi*, *Plasmodium gallinaceum*, *Plasmodium knowlesi*, and *Plasmodium reichenowi*.

For many years, research on malaria has focused on a small fraction of the *Plasmodium* genome, resulting in the identification of a few targets of protective immunity and fewer targets of new drug development. Despite the identification of >5200 parasite genes in the genome, new vaccines and drugs are unlikely to be developed solely from this set of data. Researchers are now focusing on large-scale functional genomics studies combined with powerful relational databases and informatics to "credential" the genome; to determine the characteristics of each encoded protein. A variety of high-throughput novel technologies are now being applied, such as DNA microarray (Ben Mamoun et al. 2001; Bozdech et al. 2003; Le Roch et al. 2003), protein microarrays (Bacarese-Hamilton et al. 2002), yeast two-hybrid (Y2H) protein interaction (University of Washington/Prolexys Pharmaceuticals/NMRC collaboration), and mass spectrometry (MS; Florens et al. 2002; Lasonder et al. 2002). A major focus of these efforts will undoubtedly be the development of new vaccines and drug targets (Hoffman et al. 1998, 2002).

Two recent genome-wide studies have elucidated the *P. falciparum* proteome and transcriptome of several parasite stages including sporozoites, merozoites, trophozoites, and gametocytes. By using large-scale proteomic approaches (Florens et al. 2002; Lasonder et al. 2002) and a high-density oligonucleotide array (HDA; Le Roch et al. 2003), several proteins and clusters of stage- and or function-specific have been identified for high-throughput studies.

Given that sterile protection can be conferred in humans via the bites of thousands of mosquitoes containing attenuated *P. falciparum* sporozoites and subunit vaccines have not yet conferred comparable protection, some hypothesize that an effective vaccine will require combining better protective antigens with vaccine delivery systems capable of eliciting the appropriate immune response(s). Together, the complexity and inaccessibility

Present addresses: [4]Microbiology and Immunology, Uniformed Services University, Bethesda, MD 20814, USA; [5]Department of Cellular and Molecular Medicine, St. George's Hospital Medical School, London SW17 0RE, UK; [6]Department of Microbiology and Immunology, The University of Maryland School of Medicine, Baltimore, MD 21201, USA; [7]Sanaria Inc., Rockville, MD 20852, USA.
[8]Corresponding author.
E-MAIL aguiarj@nmrc.navy.mil; FAX (301) 319-7545.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2416604.

14:2076–2082 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

of certain malaria life cycle stages, the poor understanding of mechanisms of immunity, the lack of algorithms to predict targets of protective immune mechanisms, and the power of DNA-based vaccines have provided a foundation for a genomes-to-vaccine program and have required the development of large-scale gene cloning technologies (Doolan et al. 2003a). Hence, a high-throughput DNA cloning technology is critical for the discovery of new vaccine targets based on the reverse vaccinology approach currently being applied to various genomes (Rappuoli and Covacci 2003).

To pursue such an approach, the establishment of a repository of cloned genes would be extremely helpful and would also provide an important and valuable resource to the research community. The exploitation of the *P. falciparum* genome sequences has led to the identification of novel biochemical pathways and new putative targets of heretofore unrecognized antimalarial compounds, that is, drugs traditionally used as antibacterial agents (Waller et al. 1998; Jomaa et al. 1999). Using the malaria genome sequence for vaccine development is less clear than for drug development. Researchers undertook large-scale cloning efforts to produce 350 recombinant proteins from *Neisseria meningiditis* as a resource for subsequent antibody production and identification of new vaccine candidates (Pizza et al. 2000). This "reverse vaccinology" approach has shown promise in spite of the "brute force" required to generate the hundreds of recombinant molecules for screening. As the numbers of genomes are completed that contain thousands of genes each, more cost-effective and less time-consuming efforts to produce these clones will be needed.

Traditional cloning of a single gene into a vector using restriction sites is a routine procedure for most research laboratories. However, the efficiency of this process varies greatly depending on the organism. For malaria research, particularly in the case of drug and vaccine development, candidate molecules are often cloned and assessed individually. Routine screening procedures, therefore, can be limited to the number of candidates available for testing. Because the highly (80%; A+T) rich genome of *P. falciparum* (Triglia and Kemp 1991) often results in a large number of nonrecombinant or rearranged clones, many clones must be screened before a single correct recombinant clone is obtained. The production of 5000 or more gene fragments in one or more plasmid vectors via these traditional methods for use in functional genomics studies in malaria is daunting.

To address this technical hurdle and as a means of providing large-scale reagents to the malaria research community at large, we investigated various methods of producing the large numbers of recombinant molecules for our functional genomics and vaccine studies. The method had to be cost effective, reproducible, scalable, and able to substantially overcome the problems associated with cloning the highly AT-rich *P. falciparum* genome. One such potential method, the Gateway, does not involve standard nucleic acid digestion and ligation, but instead relies on the well-characterized site-specific recombination process between bacteriophage λ and *Escherichia coli* (Hartley et al. 2000; Walhout et al. 2000). This recombinant-based cloning provides for highly efficient and accurate directional cloning. This technique has been used recently to study protein–protein interactions in *Caenorhabditis elegans* (Li et al. 2004) and protein subcellular localization (Simpson et al. 2000). As part of our vaccine development program, we have successfully used the Gateway system for the high-throughput cloning of *P. falciparum* open reading frames (ORFs) and the production of DNA vaccines and other reagents.

RESULTS AND DISCUSSION

Recombinational Cloning and Analysis

We selected a set of genes from *P. falciparum*, to clone into two Gateway-compatible master vectors: the pDONR/Zeo and pDONR207 plasmids, and three Gateway-compatible destination vectors: a modified DNA vaccine destination vector VR1020-DV and two protein expression destination vectors, pDEST17 and pMAL-2c. These studies were designed to determine the feasibility of developing a robust, high-throughput cloning system for *P. falciparum* with the intention of expanding the scope to all identified ORFs, cDNA sequences, or gene models. Figure 1 illustrates the overall flow used to accomplish these tasks, indicating specific procedures as well as the cloning efficiency achieved.

The two panels of genes described here were limited to those between 200 and 5500 bp in length to maximize the cloning efficiency. All genes from Panel A were selected prior to the full annotation of the *P. falciparum* genome, and therefore only single-exon genes were used. The second set of genes, Panel B, was selected later using bioinformatics and comparative genomics tools aimed at identifying antigens involved in pre-erythrocytic immunity mainly expressed in sporozoite and/or liver stages (Doolan et al. 2003a). Both spliced and single-exon ORFs were selected; however, because a cDNA source for these two stages is scarce, we decided to clone the largest exon for most of the spliced genes to allow for amplification from genomic DNA. Panel B comprises 79 single-exon (50%) and 80 multiexon genes (50%) as indicated in Supplemental Table 2. The actual amino acid sequences cloned for each ORF are also indicated in the same table. Future studies are planned to clone full-length

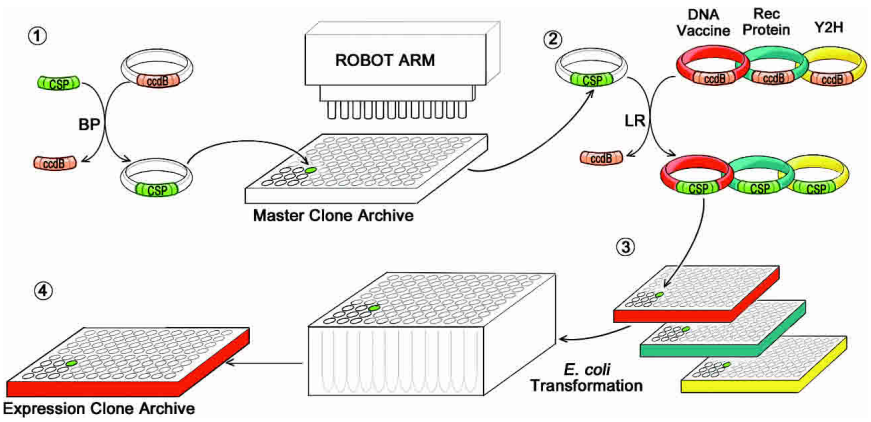

Figure 1 Flowchart of high-throughput recombinational cloning using the Gateway system. The schematic follows a single malaria ORF (CSP, green) as one example for the workflow. (1) The establishment of master clones is generated in a 96-well format including: ORF amplification, PCR purification, BP reaction, *E. coli* transformation, and recombinant clone screening. On average, we achieved 84% cloning efficiency for the BP reaction. Quality control is monitored and documented using PCR and DNA sequencing followed by archiving the master clones as purified DNA and glycerol stocks. (2) The simultaneous subcloning of ORFs into multiple expression vectors is achieved though the LR reactions. The illustration shows three destination vector examples: DNA vaccine (red), recombinant protein (green), and Y2H (yellow). Following the LR reaction and *E. coli* transformation (3), recombinant clones are screened by either single-colony (100% efficiency) or bulk-culture (93% efficiency) PCR as represented by the deep-well plate. (4) After screening and DNA purification, expression clones are then processed directly into functional assays.

spliced ORFs with sporozoite cDNA using the optimized cloning protocol described here. We have also optimized the Gateway system in the *P. yoelii* rodent model with similar cloning efficiency and have used this method to assist in the rapid and large-scale identification and characterization of *P. yoelii* antigens that can protect against a parasite challenge (Haddad et al. 2004).

To maintain maximum throughput, we screened recombinant clones by colony PCR and DNA sequencing. Four colonies per BP reaction were picked for screening to increase the chances of obtaining clones without mutations. A typical agarose gel of the PCR screening is shown in Figure 2. The screening of Figure 2A showed that of the 111 ORFs screened, 95 were successfully cloned (86%), that is, they had at least one clone positive, with 65% of the reactions generating four or more positive clones (Table 1). Similar cloning efficiency (83%) was observed for Figure 2B (159/192). Negative clones in the *P. falciparum* system were mainly caused by no colonies produced or by internal deletions as shown in Figure 2 (lane b, clone PFC0275w) and/or no recombination. DNA sequencing of the master clones was optimized to avoid the predicted secondary structures known to occur with Gateway *att* recombination sequences. DNA sequencing from both 5′- and 3′-junctions were successfully obtained for all master clones confirming the identities of all cloned ORFs. To evaluate the frequency of potential mutations introduced during cloning, we have completely sequenced 23 genes from Panel B. The analysis of the double-strand sequences indicated that no nucleotide variation was observed in 18 genes when compared with the published *P. falciparum* 3D7 sequences. However, we detected mutations on the remaining five genes: one gene contained a single silent mutation whereas the other four had a single missense mutation each. The sequencing of these 23 genes totaled 14,319 nucleotides, which yielded a 1 in 2864 discrepancy rate. Therefore, the use of high-fidelity PCR enzyme, the lack of major DNA deletions observed during PCR screening of all master clones, and the absence of frameshift mutations in the completely sequenced subset suggest that this clone set is accurate. The size range of uncloned ORFs (687–5556 bp) and cloned ORFs (198–3450 bp) overlapped (Fig. 3). Statistical analysis indicated, however, that the mean size of the cloned genes was significantly smaller than the mean size of the uncloned genes (independent samples *t*-test, two-tailed, $p < 0.0009$). We surmise that the Gateway system is better at cloning genes smaller than 5 kb although feasible attention may need to be paid toward developing methods for cloning larger genes. All entry clones from both Panels A and B have been deposited at the Malaria Research and Reference Reagent Resource Center (http://www.malaria.mr4.org) for distribution to researchers upon request. Further entry clones will be subsequently deposited at MR4 as they are generated. The identity of the available clones will also be displayed at the PlasmoDB Web site (http://www.PlasmoDB.org).

Once each of the master clones was obtained, generation of the subsequent destination clones was highly efficient and straightforward. Large-scale subcloning LR reactions were successfully set up in a high-

Table 1. Screening Efficiency of Gateway Cloning

	Panel of selected ORFs	
	A	B
ORF size range (bp)	500–3000	200–5500
Number of ORFs screened	111	192
4/4 positive colonies	72	108
3/4 positive colonies	11	28
2/4 positive colonies	06	15
1/4 positive colonies	06	08
0/4 positive colonies	16	33
Cloning efficiency (%)	86	83

The screening for recombinant master clones was performed by colony PCR on 4 colonies per each BP (gene) reaction. A positive colony represented the correct amplified gene size analyzed on agarose gel. DNA sequencing using plasmid-specific primers later confirmed the identity of the cloned genes.

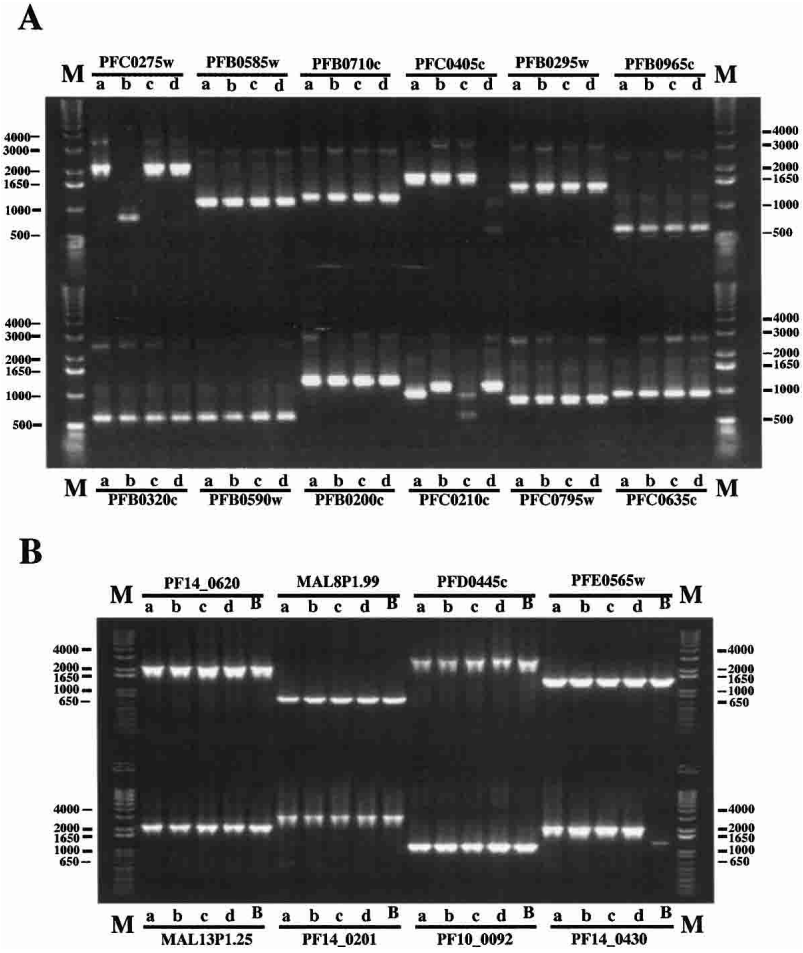

Figure 2 PCR screening BP and LR reactions. (*A*) Recombinant clones in the pDONR/Zeo master plasmid were screened by colony PCR using plasmid-specific primers (M13 forward and reverse). The figure shows amplified DNA products from 12 BP reactions representing different genes in groups of four colonies (a, b, c, d) per clone analyzed on a 1% agarose gel. (*B*) The screening of recombinants clones in the VR1020-DV destination vector was done by PCR on DNA from 4 single colonies (a, b, c, d) as well as bulk cultures (*B*) from the same transformation. Clones shown were picked at random. (M) 1-kb DNA extension ladder.

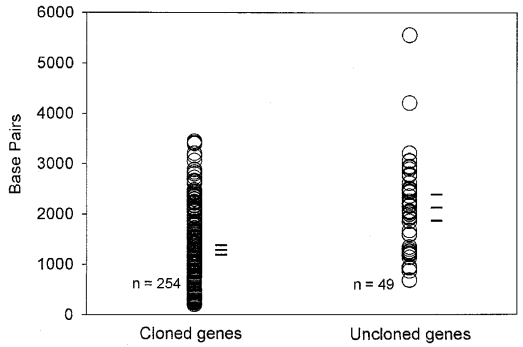

Figure 3 Size distribution of cloned and noncloned ORFs. A total of 303 amplified *P. falciparum* ORFs were screened for cloning into the Gateway entry plasmid by PCR and DNA sequencing. The horizontal lines to the right of the data points are the mean and the upper and lower 95% confidence limits.

throughput format. A single colony of each LR reaction from each of three destination vectors (VR1020-DV, pDEST17, and pMAL-2c) was screened by PCR. Initial studies showed that 100% of the selected colonies from the destination vector reaction were positive and with no observed rearrangements or deletions, and the identity of each clone was confirmed by DNA sequencing.

To increase the throughput of the cloning process we further optimized the subcloning of the ORFs by analyzing VR1020-DV clones derived from bulk cultures of transformed DH5α competent cells. This modification eliminated the need to plate each LR-transfected reaction and reduced handling errors and the time-consuming screening step. This high-throughput screening approach was also very efficient (93%; Fig. 2) and comparable to the traditional single-colony screening (100%) method.

Gene Expression

The VR1020 plasmid chosen as an expression vector for the development of DNA vaccines, expresses genes as fusion proteins with the tPA signal peptide at the N terminus. The conversion of the VR1020 vector to the Gateway system introduced the *attB1* recombination sequence between the tPA signal and the ORF sequences, encoding for 11 amino acids in-fusion with the ORF (Hedstrom et al. 1998). To verify that the Gateway-converted VR1020-DV DNA vaccine plasmid could perform as such, we cloned two well-characterized malaria genes, the *P. yoelii* (Py) circumsporozoite protein (CSP) and the *P. falciparum* (Pf) CSP and compared their in vitro and in vivo expression to the same genes cloned in the original VR1020 plasmid. COS-7 cells were transfected with PyCSP and PfCSP DNA constructs in both VR1020 and VR1020-DV plasmids as described (Hedstrom et al. 1998). The analysis of transient expression for both the PyCSP and PfCSP by the cells transfected with the VR1020-DV constructs as compared with the VR1020 constructs indicated that their expression in both plasmids was comparable. The immunogenicity of these DNA vaccines was compared by analyzing the antibody response in mice as measured by the capacity of antisera to recognize whole sporozoites (parasites) in an immunofluorescence antibody test (IFAT). IFAT titers from mice immunized with the PyCSP and PfCSP cloned in the original VR1020 DNA constructs were equivalent to the antibody titers of the mice immunized with the same genes cloned in the Gateway-converted VR1020-DV plasmid (Table 2). Based on these in vitro and in vivo results, we concluded that expression of these two proteins by the VR1020-DV destination vector was comparable to expression by the VR1020 plasmid.

We set out to generate antibody reagents to individual antigens and determine their expression in the malaria parasite using a high-throughput approach. For this purpose, we immunized groups of CD-1 outbred mice with individual Gateway VR1020-DV DNA vaccine constructs selected from Panel A. Because of the large number of sera to be analyzed, we performed the initial screen on group-pooled sera collected 2 wk after the second and third immunizations. The antibody screening was performed by the IFAT on three parasite stages, sporozoite and a mix of mature asexual erythrocytic and sexual erythrocytic stages. The screening of the pooled and individual mice sera collected after three doses resulted in 21 positives out of 95 genes (22%) examined (Fig. 4). A majority of these genes induced antibodies that reacted with erythrocytic stages (19/21). A list of the positive clones is provided as Supplemental Table 1. These results were consistent with the stage expression profiles determined by analysis of the *P. falciparum* transcriptome and proteome (Florens et al. 2002; Le Roch et al. 2003), where most (17/19) of the positive genes had high expression levels during the erythrocytic stages of development, both at RNA and protein profiles. All the positive genes for the gametocyte (three) and sporozoite (three) stages had high levels of RNA detected in these respective stages in the study of the transcriptome (Le Roch et al. 2003). The remaining 74 DNA constructs that failed to induce antibodies to parasites generally were not identified in the proteome analysis in the parasite stages evaluated; only 24 of these genes had peptides detected from any parasite stage examined. These results may reflect the fact that genes from Panel A were selected prior to publication of large-scale gene expression profiles. This correlation may indicate that the use of expression data as determined by proteome and transcriptome studies will be of significant value for predicting genes that will most likely induce an immune response and be useful in a reverse vaccinology approach. Figure 4 shows examples of the stage specificity and the cellular localization of novel proteins observed by IFAT. It is possible that some of these genes showing no expression in the sporozoite and erythrocytic stages could be inducing antibodies to other stages of the parasite life cycle not tested here, including those inside of the mosquito (oocysts and ookinetes) and in the vertebrate host hepatocyte.

Preliminary attempts to express genes from Panel A as recombinant proteins in GST- and MBP-fusion expression constructs had limited success; only seven out of 95 fusion proteins were expressed, and five of these were determined soluble. These finding are in contrast with protein expression efficiency shown for Gateway constructs encoding human ORFs using the same expression vectors (Braun et al. 2002). Subsequent to this effort, we have attempted to express recombinant proteins from 197 genes from Panel B, cloned into a GST-fusion expression plasmid. The screening indicated that 32 genes (16%) produced recombi-

Table 2. Immunogenicity of *P. falciparum* (Pf) Circumsporozoite Protein (PfCSP) and *P. yoelii* (Py) CSP DNA Constructs in VR1020 and VR1020-DV

	Dose 2	Dose 3
PfCSP–VR1020	40 (0–160)[a]	2560 (0–120)[a]
PfCSP–VR1020-DV	40 (0–160)[a]	1280 (0–2560)[a]
PyCSP–VR1020	2560 (0–10,240)[b]	5120 (0–20,480)[b]
PyCSP–VR1020-DV	1280 (0–5120)[b]	2560 (0–5120)[b]
VR1020	0[a,b]	0[a,b]

Groups of 10 CD-1 mice were immunized ID with 100 μg of each DNA construct. Sera were collected 2 wks after doses 2 and 3.
[a]Sera tested by IFAT on air-dried *P. falciparum* sporozoites.
[b]Sera tested by IFA on air-dried *P. yoelli* sporozoites.

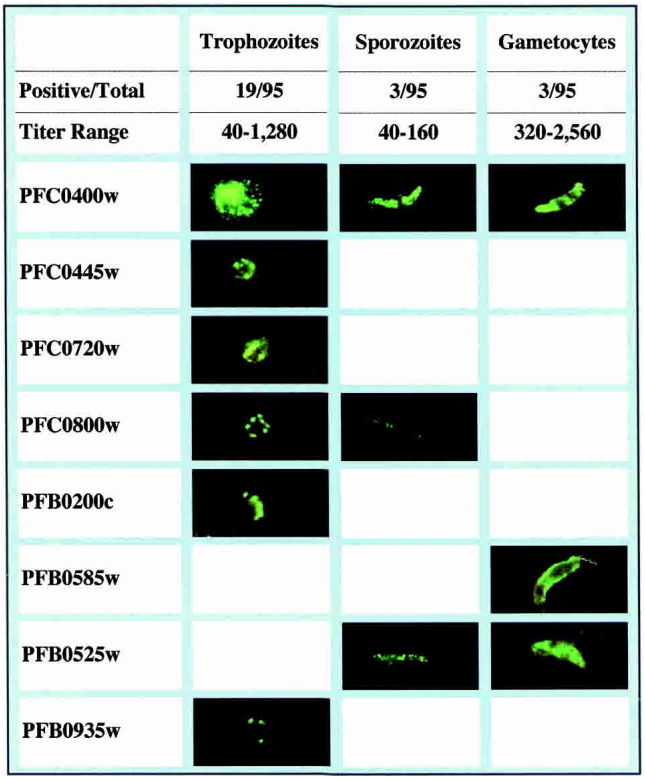

Figure 4 Immunogenicity, stage expression, and localization of *P. falciparum* proteins encoded by ORFs cloned into DNA vaccine plasmids. In all, 95 genes from Panel A were batch cloned into the VR1020-DV DNA vaccine vector, and used to immunize 5 CD-1 mice per gene. The murine sera were pooled and screened by IFAT against the sporozoite, asexual erythrocytic (trophozoites and schizonts), and sexual erythrocytic (gametocyte) stages of *P. falciparum*. The asexual erythrocytic stage was a mixture of early and late trophozoites and schizonts. Sporozoites were isolated from mosquitos' salivary glands and gametocytes were percoll-purified from in vitro cultures. Of the 95 constructs screened, the number of positive genes were shown. The titer range is the lowest and the highest end-point IFA dilution scored for any positive group. Numbers on the *left* indicate the accession number for ORFs.

nant proteins (J. Aguiar, K. Simmon, J. Ho, K. Widjaja, M. Von Rechenberg, T. Zarembinski, R. Hughes, and J. Peltier, unpubl.). The A+T content of the expressing genes was equivalent to those of the nonexpressing genes (73% vs. 76%). The major feature noted between the expressing and nonexpressing sets of clones is the size of the genes (averaging 767 bp vs. 1888 bp, respectively). Many malaria researchers attribute the low rate of recombinant protein expression to its toxic effect on *E. coli*. We set out to bypass the *E. coli* environment by attempting to express a subset of 39 genes from Panel B in a Gateway-compatible cell-free system (Kim et al. 1996). Using this cell-free system, we were able to express 17 out of 39 genes (44%) proteins. Although this represents a slight improvement over our *E. coli* expressed GST-fusion proteins (12 out of 39 genes), this apparent improvement did not hold true in the expression of larger transgene products. We are currently characterizing the proteins generated in both *E. coli* as well as by the cell-free system. In an attempt to express small to mid size (200 bp to 1.4 kb), *P. yoelii* single-exon collaborators have generated similar moderate expression rate levels (S. Sharma, J. Bray, J. Lew, M. Vedadi, A. Edwards, and C. Arrowsmith, unpubl.).

Although other alternative expression systems have been shown to improve the expression of malaria proteins such as baculovirus (Liang et al. 2000), *Pichia* (Zou et al. 2003), and the use of codon optimized transgenes (Zou et al. 2003), the viability of these approaches depends on the availability of a high-throughput format. Codon optimization has routinely improved protein expression, although the cost of synthesizing large numbers of codon-optimized genes limits their utility in high-throughput cloning efforts. Improved methods of expressing *P. falciparum* proteins will continue to be required despite the methods of producing large numbers of recombinant *P. falciparum* molecules.

In conclusion, we demonstrated that the Gateway cloning system provides a highly efficient method to produce recombinant plasmids containing genes from *P. falciparum*. Noteworthy is the fact that most of the *P. falciparum* DNAs, which are known to be recombination-prone, were cloned without deletions in the ORF. Using a limited number of ORFs and the DNA vaccine strategy, we have been able to further explore the functional profile and identify the stage-specific expression of novel antigens with antisera produced in mice. These data establish the Gateway system as a platform technology, which provides for the rapid and flexible design and construction of recombinant plasmids needed for functional genomics studies and for vaccine and drug development efforts. Using this system, one will be able to perform high-throughput functional assays using a variety of expression vectors, that is, protein expression vectors to generate reagents for protein microarray experiments and animal immunization, conduct large-scale transfection experiments to assay protein localization, determine immunogenicity of candidate antigens, develop expression vectors for Y2H systems to screen for host receptors and parasite protein interactions, and develop large numbers of recombinant virus constructs that can be used in immune assays for screening protective antigens (Doolan et al. 2003a) and may allow for the functional identification of the large number of unknown and hypothetical proteins in the *P. falciparum* genome.

METHODS

Selection and PCR Amplification of ORFs

Different ORF selection criteria were applied to evaluate the high-throughput feasibility of the Gateway technology across multiple *P. falciparum* genes. The first panel of genes (Panel A) was comprised of 111 single-exon ORFs from early Chromosomes 2 and 3 annotations and ranged in size from 0.3 to 3 kb. The genes in Panel B were selected based on a set of credentialed data for their potential as vaccine targets (Supplemental Table 2). A separate functional genomics approach led to the creation of each of the six categories. The first category, AffyCluster1 ($n = 38$), is genes that were found to be up-regulated in a transcriptional cluster only in sporozoite stages (Le Roch et al. 2003). The High MudPIT SPZ category ($n = 7$) contains sporozoite-expressed genes identified by mass spectrometry found to have high peptide sequence coverage (Florens et al. 2002). The Bioinformatics category ($n = 30$) is comprised of genes with predicted 5′-signal sequences and 3′-transmembrane domain(s) also possessing evidence of protein expression in sporozoite stages. The PyLCM category ($n = 65$) contains *P. falciparum* orthologs of *P. yoelii* liver-stage genes identified from an EST sequencing project (J.B. Sacci, J.M.C. Ribeiro, F. Huang, U. Alam, J. Russell, P.L. Blair, A. Witney, D.J. Carucci, A.F. Azad, J.C. Aguiar, et al., unpubl.) not identified by mass spectrometry. The Immunosense category ($n = 6$) is comprised of genes with encoded proteins originally identified in the sporozoite proteome and also recognized by volunteers immunized with the irradiated sporozoites (Doolan et al. 2003b). The PRBC Surface group ($n = 13$) contains genes whose proteins are candidates for being expressed on the surface of infected erythrocytes (Florens et al. 2004). In addition, this panel was predomi-

nantly selected to explore ORFs with unknown function (annotated as "hypothetical") for downstream determination of potential function. Panel B was comprised of either single-exon genes or the major exon from genes with predicted multiexon gene structures.

The ORFs from Panel B varied in size from 0.2 to 5.5 kb. An entire list of cloned genes is provided in Supplemental Table 2. A customized computer script was used with the automated computational program Primer 3 (http://www.broad.mit.edu/genome_software/other/primer3.html) to specifically design gene-specific oligonucleotides used for the amplification of Gateway-compatible in-frame full-length ORFs. An additional 12 bases of overlapping sequence at the 5′-end to the universal outer adapter *attB* primers were added to use a "nested" PCR with a universal outer set of primers. The overlapping sequence for the forward and reverse primers were designed so that the final amplified products contained *attB* sequences at both 5′- and 3′-ends (forward primer overlapping sequence 5′-GCAGGCTCCACC-3′ and reverse primer overlapping sequence 5′-AGAAAGCTGGGT-3′). The sequence for the *attB1* adapter primer was 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACC-3′ and for the *attB2* adapter primer was 5′-GGGGACCACTTTGTACAAGAAAGCTGGGT-3′. The nested PCR strategy briefly described below was used to reduce the overall length of hundred of primers, thereby significantly reducing the cost incurred with primer synthesis. PCR amplification as well as the PCR isolation and cloning were done in the 96-well format. Full-length genes were amplified from *P. falciparum* (3D7) genomic DNA (100 ng) using a high fidelity Platinum *Pfx* polymerase (Invitrogen) and each set of inner primers for 10 cycles. A 10-µL aliquot of this amplified mix was added to a second set of adapter primers and amplified for an additional 30 cycles. Before proceeding with the BP reactions, PCR products were run on an agarose gel to confirm the size of the amplicons and then purified using one of two protocols. The first method used to purify PCR products from ORF Panel A included the excision and gel purification of the single DNA specie by spinning through a UniFilter hydrophobic 96-well plate (Whatman). The second protocol used for ORFs from Panel B consisted of a modified 96-well purification method of precipitating DNA in a 30% PEG solution. This adaptation increased the experimental throughput, limited sample handling, and reduced time without affecting DNA yield or quality.

Recombinational Cloning (BP Reaction)

The cloning to obtain master clones was performed using the BP reaction kit (Invitrogen) following the protocol recommended by the manufacturer. Two master plasmids were used for the cloning: pDONR207 and pDONR/Zeo. The latter encodes the zeocin-resistance gene, which allows the subsequent use of destination plasmids with a variety of antibiotic resistance markers and the use of M13 universal primers. Chemical competent DH5α cells were transformed with 2-µL aliquots of each BP reaction and plated on gentamicin or zeocin LB plates. For the purpose of assessing the cloning efficiency, four colonies per capture reaction (BP reaction) were selected for screening by PCR and by DNA sequencing using plasmid-specific primers.

Plasmid DNA was prepared from each colony in 96-well format using either REAL plasmid kit (QIAGEN) or Montage plasmid miniprep kit (Millipore). PCR screening was performed in a 20-µL reaction with 1 µL of plasmid DNA using the Platinum SuperMix kit (Invitrogen) and analyzed by 1% agarose gels. Positive master clones were sequenced using M13 or pDONR207-specific forward (5′-TCGCGTTAACGCTAGCATGGATCTC-3′) and reverse (5′-GTAACATCAGAGATTTTGAGACAC-3′) primers and the BigDye terminator method on a 3100 Genetic Analyzer (ABI). To sequence across the hairpin recombination segments of the clones, an optimized sequence cycling protocol was developed. Optimized sequencing cycling for Gateway clones was as follows: denaturing for 5 min at 95°C, with 30 cycles of 10 sec at 96°C, 5 sec at 50°C, and 4 min at 60°C. Duplicate glycerol stocks were then generated from each positive Master clone. Apart from the genes selected, we also included two well-characterized malaria genes: the *P. falciparum* Circumsporozoite Protein (PfCSP) gene and the *P. yoelii* Circumsporozoite Protein (PyCSP) gene.

Recombinational Cloning (LR Reaction)

The transfer of ORFs from master clones to expression vectors was performed using the LR reaction kit (Invitrogen). Two protocols were used; one recommended by the manufacturer that screens single-colony for recombinants is described here; the other method described later was adapted to truly high throughput by screening clones in bulk. Briefly, 1 µL of each Master clone was incubated with 200 ng of Topoisomerase I-treated VR1020-DV plasmid in a 20-µL reaction containing the LR reaction reagents. After proteinase K treatment, a 2-µL aliquot of each reaction was used to transform DH5α competent cells, and part was plated on LB plates containing appropriate antibiotic. The screening was performed with one colony per LR reaction. Plasmid DNA was prepared as described above and screened by PCR and DNA sequencing using plasmid-specific primers.

Large-Scale Recombinant Production

In parallel to the single-colony approach described above, we also tested an alternative screening for these LR reactions. All steps for this procedure were performed in a 96-well platform. Following the LR reaction, bacteria growth media (LB broth) and kanamycin was added to a fraction of the transformed DH5α cells and cultured overnight. DNA was prepared from these bulk cultures and screened by PCR as described above. PCR amplicons were analyzed in an agarose gel running in parallel with PCRs from single colony screening.

DNA Vaccine Construction

To generate DNA vaccine constructs we used the VR1020-DV plasmid as the destination vector. The original DNA vaccine VR1020 plasmid (Vical Inc.; Luke et al. 1997) was modified to be compatible with the Gateway cloning system. The expression of encoded genes in the VR1020-DV plasmid is driven by a CMV immediate/early gene promoter in fusion with the tPA (tissue plasminogen activator) leader sequence. The transfer of ORFs from the master clone to VR1020-DV was performed via the LR reaction.

Recombinant Protein Expression

In an attempt to express the proteins of cloned genes in *E. coli*, we subcloned the ORFs into two recombinant protein expression vectors: pDEST17 and pMAL-2c. The PDEST17 vector (Invitrogen) is designed for expressing fusion proteins containing six histidines at the N terminus. Sequences transferred into pDEST17 from a master clone are expressed using T7 RNA polymerase under control of a salt-inducible promoter. In the pMAL-2c vector, the gene of interest is fused to the *malE*gene, which encodes maltose-binding protein (MBP), and expressed using the strong tac promoter and translation initiation signals of MBP. Also pMAL-2c contains the normal *malE*gene signal sequence, which directs the fusion protein through the cytoplasmic membrane. The pMAL-2c vector was converted to be compatible with the Gateway cloning system (Braun et al. 2002). The same protocol described for the LR reaction and screening used to transfer genes to the VR1020-DV vector was used for these protein expression vectors.

Mouse Immunization and Antibody Testing

Cloning grade DNA was purified (QBIOgene) for mouse immunization. Groups of five outbred CD-1 mice (Charles River Laboratories) were immunized three times with a 100-µg dose of each DNA construct at 3-wk intervals. The immunizations were done by the intradermal route (ID) at the base of the tail. Sera were collected 2 wk after the third dose, and antibody reactivity was assessed by standard immunofluorescence assay (IFAT) against air-dried sporozoites and erythrocytic stages of *P. falciparum* as described elsewhere (Aguiar et al. 2001).

ACKNOWLEDGMENTS

We are grateful to Trevor Jones and Annabelle Wright from WRAIR-NMRC for guidance with statistics and graphics. Many thanks to Kristi Strang, Allison Halleck, Steven Lloyd, and Pascal Braun from the Harvard Institute of Proteomics (HIP) for their assistance with cloning and supplying the pMAL-2c destination plasmid. We also thank Steven Max, Dolores Ciufo, Darby Cortijo, Graziella Piras, and Gary Temple from Invitrogen for their guidance with the Gateway system. The work was supported by funds allocated to the Naval Medical Research Center by the US Army Medical Research Material Command (61102A.S13.C.A0078 and 6000.RAD1.C.A0305) and by the Office of Naval Research (Navy ONR-Core 61153N.M458.S18.AO242). The views of the authors expressed herein are their own and do not purport to reflect those of the US Navy or the US Department of Defense. The experiments reported herein were conducted according to the principles set forth in the *Guide for the Care and Use of Laboratory Animals*, Institute of Laboratory Animal Resources, National Research Council, National Academy of Sciences Press, 1996.

REFERENCES

Aguiar, J.C., Hedstrom, R.C., Rogers, W.O., Charoenvit, Y., Sacci, J.B., Lanar, D.E., Majam, V.F., Stout, R.R., and Hoffman, S.L. 2001. Enhancement of the immune response in rabbits to a malaria DNA vaccine by immunization with a needle-free jet device. *Vaccine* **20:** 275–280.

Bacarese-Hamilton, T., Bistoni, F., and Crisanti, A. 2002. Protein microarrays: From serodiagnosis to whole proteome scale analysis of the immune response against pathogenic microorganisms. *Biotechniques* **Suppl:** 24–29.

Ben Mamoun, C., Gluzman, I.Y., Hott, C., MacMillan, S.K., Amarakone, A.S., Anderson, D.L., Carlton, J.M., Dame, J.B., Chakrabarti, D., Martin, R.K., et al. 2001. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Mol. Microbiol.* **39:** 26–36.

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.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. 2003. The Transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1:** E5.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., and LaBaer, J. 2002. Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci.* **99:** 2654–2659.

Carlton, J. 2003. The *Plasmodium vivax* genome sequencing project. *Trends Parasitol.* **19:** 227–231.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., et al. 2002. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419:** 512–519.

Doolan, D.L., Aguiar, J.C., Weiss, W.R., Sette, A., Felgner, P.L., Regis, D.P., Quinones-Casas, P., Yates III, J.R., Blair, P.L., Richie, T.L., et al. 2003a. Utilization of genomic sequence information to develop malaria vaccines. *J. Exp. Biol.* **206:** 3789–3802.

Doolan, D.L., Southwood, S., Freilich, D.A., Sidney, J., Graber, N.L., Shatney, L., Bebris, L., Florens, L., Dobano, C., Witney, A.A., et al. 2003b. Identification of *Plasmodium falciparum* antigens by antigenic analysis of genomic and proteomic data. *Proc. Natl. Acad. Sci.* **100:** 9952–9957.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., et al. 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419:** 520–526.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., and Wu, Y. 2004. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.* **135:** 1–11.

Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., et al. 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282:** 1126–1132.

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.

Haddad, D., Bilcikova, E., Witney, A.A., Carlton, J.M., White, C.E., Blair, P.L., Chattopadhyay, R., Russell, J., Abot, E., Charoenvit, Y., et al. 2004. Novel antigen identification method for discovery of protective malaria antigens by rapid testing of DNA vaccines encoding exons from the parasite genome. *Infect. Immun.* **72:** 1594–1602.

Hartley, J.L., Temple, G.F., and Brasch, M.A. 2000. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10:** 1788–1795.

Hedstrom, R.C., Doolan, D.L., Wang, R., Kumar, A., Sacci Jr., J.B., Gardner, M.J., Aguiar, J.C., Charoenvit, Y., Sedegah, M., Tine, J.A., et al. 1998. In vitro expression and in vivo immunogenicity of *Plasmodium falciparum* pre-erythrocytic stage DNA vaccines. *Int. J. Mol. Med.* **2:** 29–38.

Hoffman, S.L., Rogers, W.O., Carucci, D.J., and Venter, J.C. 1998. From genomics to vaccines: Malaria as a model system. *Nat. Med.* **4:** 1351–1353.

Hoffman, S.L., Subramanian, G.M., Collins, F.H., and Venter, J.C. 2002. *Plasmodium*, human and *Anopheles* genomics and malaria. *Nature* **415:** 702–709.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H.K., et al. 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285:** 1502–1503.

Kim, D.M., Kigawa, T., Choi, C.Y., and Yokoyama, S. 1996. A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur. J. Biochem.* **239:** 881–886.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., et al. 2002. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419:** 537–542.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–1508.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303:** 540–543.

Liang, H., Narum, D.L., Fuhrmann, S.R., Luu, T., and Sim, B.K. 2000. A recombinant baculovirus-expressed *Plasmodium falciparum* receptor-binding domain of erythrocyte binding protein EBA-175 biologically mimics native protein. *Infect. Immun.* **68:** 3564–3568.

Luke, C.J., Carner, K., Liang, X., and Barbour, A.G. 1997. An OspA-based DNA vaccine protects mice against infection with *Borrelia burgdorferi*. *J. Infect. Dis.* **175:** 91–97.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capecchi, B., et al. 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287:** 1816–1820.

Rappuoli, R. and Covacci, A. 2003. Reverse vaccinology and genomics. *Science* **302:** 602.

Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Weimann, S. 2000. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1:** 287–292.

Triglia, T. and Kemp, D.J. 1991. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Mol. Biochem. Parasitol.* **44:** 207–212.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000. GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzym.* **328:** 575–592.

Waller, R.F., Keeling, P.J., Donald, R.G., Striepen, B., Handman, E., Lang Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S., and McFadden, G.I. 1998. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **95:** 12352–12357.

Zou, L., Miles, A.P., Wang, J., and Stowers, A.W. 2003. Expression of malaria transmission-blocking vaccine antigen Pfs25 in *Pichia pastoris* for use in human clinical trials. *Vaccine* **21:** 1650–1657.

WEB SITE REFERENCES

http://www.broad.mit.edu/genome_software/other/primer3.html; Primer 3.

http://www.malaria.mr4.org; Malaria Research and Reference Reagent Resource Center.

http://www.PlasmoDB.org; PlasmoDB.

Received June 22, 2004; accepted in revised form June 24, 2004.