Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9867–9871, November 1993
Genetics

Gene sequence tags from *Plasmodium falciparum* genomic DNA fragments prepared by the "genease" activity of mung bean nuclease

G. ROMAN REDDY*†, DEBOPAM CHAKRABARTI*‡, SHELDON M. SCHUSTER‡§, ROBERT J. FERL¶, ERNESTO C. ALMIRA‡, AND JOHN B. DAME*†

*Department of Infectious Diseases, College of Veterinary Medicine, ‡Interdisciplinary Center for Biotechnology Research (ICBR), §Department of Biochemistry and Molecular Biology, College of Medicine, ¶Department of Horticultural Sciences, College of Agriculture, University of Florida, Gainesville, FL 32611

Communicated by Louis H. Miller, June 30, 1993

ABSTRACT **A genes-first approach to genome sequencing is described which efficiently generates gene sequence tags from genomic DNA. Mung bean nuclease (EC 3.1.30.1) cleaves the genomic DNA of many organisms before and after genes and within some introns. Analysis of gene sequence tags prepared from mung bean nuclease-digested *Plasmodium falciparum* DNA demonstrates that this method has several advantages over the popular cDNA expressed sequence tag approach. To date, 673 sequence tags containing over 215 kb of sequence have been generated from 400 clones. Sixty clones (15%) have significant similarity to sequences in the protein and translated nucleic acid data bases. These represent 51 unique genes, of which only 5 encode previously known *P. falciparum* proteins. The identified proteins include those expressed in erythrocytic, exoerythrocytic, and gametocytic stages of the parasite. Thirty percent of clones identified appear to carry complete coding regions. The spacer DNA separating genes is rarely cloned. These gene sequence tags will form a useful data base from which to initiate projects to develop new therapeutics, vaccines, and strategies to control human malaria.**

Plasmodium falciparum is the most dangerous and rapidly proliferating of the four *Plasmodium* species parasitic in humans and kills about 1 million people worldwide annually (1). The paradigm shift now occurring in biology is toward having all the genes of an organism recorded in data bases, available to be used as the starting point for further investigations (2). Identification, sequencing, and mapping of the structural genes will provide a foundation data base from which to launch applied research programs for antimalarial drug and vaccine development. Recent reports on random sequencing of cDNA libraries from human brain (3, 4) and *Caenorhabditis elegans* (5, 6) have demonstrated that this is an efficient method for obtaining preliminary data on coding sequences. A *P. falciparum* cDNA library is limited, however, to the genes expressed in the life cycle stage used to prepare the mRNA, and the probability of obtaining a given cDNA sequence depends on the level of expression of the gene. Obtaining rare cDNA clones at random, including those that encode regulatory enzymes/proteins, is problematic. Utilizing the "genease" activity of mung bean nuclease (EC 3.1.30.1) to obtain the sequences of genes from genomic DNA (7–9) overcomes these problems.

Under modified reaction conditions mung bean nuclease cleaves *P. falciparum* genomic DNA precisely before and after genes and within some introns (7, 9). This method makes it possible to clone intact genes or gene fragments from virtually all structural genes of the parasite. Although the DNA of *Plasmodium* spp. is the best studied (7–11), this approach has also been used in numerous other protozoans, including *Trypanosoma* (12), *Giardia* (13), *Toxoplasma* (14), *Leishmania* (15), and *Babesia* (16, 17). It has been suggested that altered DNA structure near gene boundaries determines the recognition sites for this activity (9), which are likely to be found in the genomes of a much larger range of organisms, including higher eukaryotes. This distinctive genes-first approach to genome characterization has been evaluated here by examining a large number of gene fragments.‖

MATERIALS AND METHODS

Library Construction and Preparation of DNA for Sequencing. *P. falciparum* clone HB3 (American Type Culture Collection no. 50113) was continuously cultivated *in vitro* in human erythrocytes (18). Cultures at a parasitemia of 8–10% were lysed with 0.1% saponin, and genomic DNA was isolated by using an SDS/proteinase K method (19). *P. falciparum* genomic DNA samples were routinely analyzed by Southern blot hybridization (20) using the human *Alu* I repeat as a probe, and only the DNA preparations that contained <1% human DNA were employed in the study. The DNA was digested with mung bean nuclease (Promega) at 50°C in the presence of 30% (vol/vol) formamide essentially as described (9). The digestion was monitored by comparing the size of the circumsporozoite protein (CSP) gene fragment with the size previously reported (7–9). Reaction products which contained the 1.3-kb CSP fragment were selected for cloning. The DNA was blunt ended with T4 DNA polymerase (21), ligated into *Eco*RV-cut, calf-intestine-alkaline-phosphatase-treated pBluescript SK (+) (Stratagene), and used to transform *Escherichia coli* strain XL1-blue. Recombinant clones were selected at random, and plasmid DNA was isolated from 1.5-ml liquid cultures by a boiling preparation method essentially as in ref. 22.

Sequencing Reactions and Analysis. Cyclic sequencing reactions were performed on clones with inserts ≥0.3 kb by using a *Taq* dye primer cycle sequencing kit with T3 and T7 primers, as described by the supplier (Applied Biosystems). The samples were analyzed on an Applied Biosystems 373A DNA sequencer, and data were edited manually to remove vector sequences and also to remove 3′ end sequences of low reliability. Sequence tags translated into all six reading frames were compared against the protein data bases [translated GenBank (GP), daily GenPept update (GPU), Protein

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BLAST, Basic Local Alignment Search Tool; ORF, open reading frame.

††To whom reprint requests should be addressed.

‖The sequences described in this paper have been deposited in the GenBank data base (accession nos. T02634–T02808 and T09496–T09993).

Table 1. Genes putatively identified by producing high-scoring segment pairs with sequences in GenPept, GPUpdate, SwissProt, and PIR databases

Clone	Est. Expanse of Clones	Sequences Producing Highest-Scoring Segment Pairs	Accession Number	Homologous Region (a.a.)	Identity/ Similarity (%)	Highest Score	Random Probability
0026M	E	30S Ribosomal Protein S-18	PIR:JH0419	9 to 68	60/83	188	3.90e-21
0368M	C*	30s Ribosomal Protein S12@	PIR:R3EG12	24 to 121	61/84	322	3.60e-43
0524M	C*	30s Ribosomal Protein S12@	PIR:R3EG12	27 to 121	62/82	317	6.40e-42
0324M	C*	50s Ribosomal Protein L16	PIR:R5KT16	73 to 110	44/65	87	4.30e-05
0204M	C*	60S Ribosomal Protein L18A	PIR:R5RT18	7 to 107	54/77	304	3.00e-40
0482M	C*	60s Ribosomal Protein L27a	GP:TETRPL29A_1	58 to 149	45/72	187	1.90e-21
0390M	E	60s Ribosomal Protein L8	PIR:R5RTL8	7 to 110	59/80	283	2.20e-36
0088M	E	Actin II@@	GP:PFAACTII_1	150 to 217	89/89	320	4.00e-42
0284M	U	Adenylyl Cyclase Gene	GPU:DDIADCYA_1	734 to 794	34/63	85	5.30e-05
0100M	C*	Alternative Splicing Factor-1@	PIR:B40041	104 to 185	57/70	212	1.00e-24
0202M	C	Alternative Splicing Factor-1@	PIR:B40041	17 to 106	51/74	234	4.20e-28
0187M	E	ATP-Dependent RNA Helicase	PIR:S11485	455 to 520	50/75	194	3.50e-22
0470M	E	Beta Coat Protein	PIR:S13520	882 to 934	51/78	92	9.60e-06
0211M	U	Bkm-like Sex-Determining Region Hypothetical Protein CS319	PIR:C21124	21 to 77	38/59	92	4.60e-09
0487M	E	Ca+2 Dependent Protein Kinase@	GP:SOYCADPK_1	159 to 247	50/68	143	7.70e-14
0532M	E	Ca+2 Dependent Protein Kinase@	PIR:A40811	140 to 220	35/53	170	3.70e-18
0075M	C	Cathepsin D	PIR:KHPGD	333 to 407	59/74	82	2.50e-08
0405M	U	Clustered Aspargine Rich Protein	PIR:A23535	197 to 283	36/64	107	1.40e-14
0017M	C	Cyclophilin	PIR:S07585	2 to 104	67/78	327	7.40e-41
0362M	C*	DNAJ Protein (B. Subtilis)	GP:BACHSP_4	5 to 86	53/71	120	3.20e-10
0291M	E	DNAJ Protein (E. coli)	PIR:HHECDJ	37 to 84	41/60	100	1.80e-07
0422M	C*	DNAJ Protein Homologue HSJ1	GP:HUMHSJ1MR_2	3 to 34	53/68	99	2.70e-15
0431M	U	Duplicate Procyclin	PIR:S06171	48 to 73	61/84	91	7.20e-06
0132M	C	Dynamin	PIR:S11508	6 to 79	50/75	208	6.10e-25
0145M	E	Dynein Beta Chain#	PIR:S17653	2212 to 2274	38/68	120	2.70e-10
0545M	E	Dynein Beta Chain#	PIR:S17653	3400 to 3464	29/56	84	8.20e-05
0136M	C	Glucose 6-P Isomerase@@	PIR:A36567	1 to 99	90/91	456	1.40e-64
0419M	E	Glutamate Dehydrogenase	PIR:DENCED	824 to 882	61/79	182	4.60e-21
0065M	E	Glycerol 3-P Dehydrogenase@	PIR:A25189	199 to 242	63/81	143	4.00e-14
0147M	E	Glycerol 3-P Dehydrogenase@	PIR:A26687	197 to 242	45/65	97	1.60e-06
0321M	E	Glycerol 3-P Dehydrogenase@	PIR:A26687	199 to 242	61/79	132	2.70e-12
0421M	E	GTP Binding Protein RYH1	PIR:S12789	108 to 188	53/69	201	1.20e-23
0058M	E	GTP Binding Protein YPT3	PIR:S10026	62 to 168	61/86	316	3.90e-42
0406M	E	Initiation Factor 4A-1#	PIR:S00986	244 to 360	75/86	488	3.00e-69
0197M	E	Initiation Factor eIF-4A#	PIR:JS0039	154 to 188	54/80	113	4.50e-09
0225M	E	Iron Responsive Element Binding Protein	PIR:S18720	483 to 569	62/75	284	4.70e-37
0424M	E	Lactate Dehydrogenase	GPU:LISACTLDH_6	45 to 102	32/70	97	5.40e-07
0224M	E	Liver Stage Antigen @@	GPU:PFALSA1G_1	1660 to 1761	95/95	423	2.10e-57
0522M	E	Mitochondrial Phosphate Carrier Protein	GP:HUMMPCP_1	99 to 191	55/69	255	4.00e-32
0175M	E	Mitochondrial Pyridine Nucleotide Transhydrogenase Beta Subunit#	GP:ECOPNTAB_2	61 to 106	54/78	133	1.30e-12
0344M	E	Mitochondrial Pyridine Nucleotide Transhydrogenase Beta Subunit#	GPU:ECOPNT_2	153 to 271	49/68	270	4.50e-35
0314M	E	Nitrogen fixation-U Protein	GP:AVINIFREG_1	234 to 281	43/66	120	9.60e-11
0488M	E	Poly A Nuclease	GP:YSCPAN1A_1	1064 to 1114	43/54	110	2.10e-08
0031M	E	Proteasome C3 (Rat)	SP:PRC3$RAT	60 to 176	52/68	297	4.50e-38
0158M	E	Ras-Like Protein TC4	GP:HUMRASAC_1	39 to 110	70/83	299	1.50e-38
0160M	C*	Regulatory Protein - Yeast	PIR:S17012	12 to 99	29/59	129	1.20e-11
0188M	C	Rfbf Protein - S. typhimurium	PIR:S15304	1 to 63	39/61	103	8.00e-08
0229M	E	Ribonucleotide Reductase@	GP:HUMRR2SS_1	153 to 190	55/81	129	1.10e-11
0436M	E	Ribonucleotide Reductase@	GP:HUMRR2SS_1	153 to 209	40/61	113	3.60e-09
0347M	E	Ring Infected Surface Antigen##	PIR:A25526	755 to 841	33/59	88	3.00e-05
0337M	E	S. cerevisiae Gene (Vasa Protein)	GP:YSCSEQ_1	239 to 327	32/66	147	1.50e-14
0176M	E	S. pombe cdc21 Gene	GP:YSPCDC21_1	762 to 831	37/62	118	2.10e-10
0046M	E	Serine Hydroxy Methyl Transferase	GPU:NEUSERHMT_1	45 to 120	57/77	215	9.30e-26
0201M	C*	Serine Proteinase type 1 Precursor	PIR:A38738	342 to 424	37/57	131	4.30e-12
0104M	C*	Sexual Stage Specific Protein@@	PIR:S10313	60 to 145	76/79	216	3.30e-26
0334M	E	Staphylococcus xylosus BBM3XM	GP:STABBM3XM_1	295 to 363	17/79	116	7.80e-10
0167M	E	Thioredoxin	SP:THIO_RHORU	22 to 83	37/64	98	5.10e-07
0343M	U	Thrombospondin Precursor##	PIR:A37905	384 to 402	57/73	71	3.90e-07
0049M	E	Tubulin II (alpha)@@	GP:PFAATUBII_1	373 to 450	91/91	376	2.40e-50
0543M	C	Ubiquitin Carrier Protein	GPU:ALFUBIQUIT_1	42 to 141	59/77	308	1.10e-39

Sequence tags from 400 clones (273 with tags from both strands and 127 with tags from 1 strand) were compared against the translated nucleic acid and protein databases using the NCBI network BLAST. The minimum length of sequence submitted to BLAST search was 200 bases. All the sequence pairs with scores of 70 and above were manually reviewed for putative identification of genes. The 60 clones (15% of total) with putative identification are given in the table. These represent 51 unique genes (13%) of which only 5 closely match previously reported *P. falciparum* sequences. @, Clones of same exons; #, Clones of different exons; @@, Exact match to *P. falciparum* genes; ##, Non-exact match to *P. falciparum genes*; C, complete protein coding sequence; C*, Sufficient length for a complete protein coding sequence; E, Exon; U, unknown.

Identification Resource (PIR) data base, and Swiss-Prot (SP)], by using the National Center for Biotechnology Information (NCBI) network Basic Local Alignment Search Tool (BLAST) server (23).

RESULTS

Random Sequencing of Mung Bean Nuclease Library Clones and BLAST Search Analysis for High-Scoring Segment Pairs. Four hundred and sixty clones (average insert size 1.3 kb) were selected at random from a library of mung bean nuclease digestion fragments for sequencing from one or both ends (Table 1). A total of 673 unique sequence tags were obtained that met criteria for sequence quality and length. These sequences, derived from 400 clones, have an average length of 320 nucleotides, and together make up more than 215 kb of DNA sequence from *P. falciparum*. These sequences were compared with those in the protein sequence data bases by using the BLAST network server (23). High-scoring segment pairs from 60 clones were considered biologically significant, and these established the putative identification of 51 *P. falciparum* genes (Table 1). (In Table 1, 3.90e-21 in the random probability column indicates a probability of 3.90×10^{-21}.) The other 9 identified clones were linked to at least one other clone by forming a high-scoring segment pair with one of eight different data base sequences as shown in Table 1. Three of these represent different exons from the same gene, and 6 are clones of the same exons. These 6 clones represent 10% of the identified sequences and define the project's frequency of redundancy, since it is assumed that a similar percentage of unidentified clones are likewise represented by multiple clones. Five of the 51 genes identified were *P. falciparum* sequences already in the data bases and were derived from erythrocytic, exoerythrocytic, and sexual stages of the parasite [α-tubulin, actin II, glucose-6-phosphate isomerase, sexual-stage-specific protein, and liver stage antigen (LSA-1)]. Two other clones had homology to *P. falciparum* genes (ring-infected surface antigen and thrombospondin precursor).

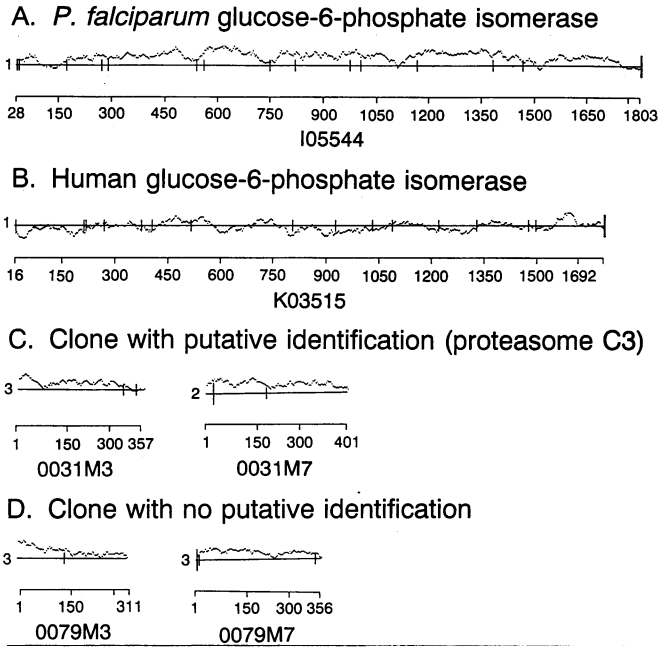

FIG. 1. PREDICT analysis. The output from PREDICT is presented for previously described mRNA coding sequences from *P. falciparum* and human glucose-6-phosphate isomerase, a clone with high-scoring segment pairs with proteasome C3, and an unidentified clone with a long ORF. In *A* and *B*, the frame that encoded the protein is used for analysis. In *C* and *D*, the frame that yielded the longest ORF is shown. The solid horizontal line indicates a score of zero. Scores for *P. falciparum* protein sequences almost never fall below zero, whereas the scores for human genes or the other two frames on the coding strand rarely exceed zero. Initiation codons and stop codons are marked as short and long vertical bars, respectively.

Analysis of Genes Putatively Identified by High-Scoring Segment Pairs. The putatively identified clones were compared to protein and nucleotide sequences in the data bases to judge whether the clones represent complete genes (Table 1). Seven clones appeared complete, since they showed similarity extending from the initiation codon through to the termination codon. Eleven clones probably represent complete gene fragments, since they showed similarity beginning from either the initiation codon or the termination codon, and they were large enough to encode the complete protein. Either the sequence from the other end for these clones is not available or homology was not observed. Together, these 18 clones represent 30% of the genes identified by similarity. Thirty-seven of the remaining clones (62%) have one or more exons but appear incomplete. Exons are defined here on the basis of the presence of consensus intron/exon junctions at or near the homologous regions specific for *P. falciparum* (24). Due to insufficient information, it was not possible to define for the other 5 clones whether they represent exon(s) or full-length gene fragments.

PREDICT Data Analysis for Identifying the *P. falciparum*-Specific Coding Sequences. The PREDICT computer program, based on patterns of codon usage and amino acid composition of *P. falciparum* genes, identifies open reading frames (ORFs) specific for *P. falciparum* genes and displays a running average of a prediction score in each of the three reading frames (25). The output from this program for glucose-6-phosphate isomerase (G6PI) mRNA from *P. falciparum* and human (GenBank accession nos. J05544 and K03515, respectively) and for sequence tags for clones 0031M and 0079M is in Fig. 1. *P. falciparum* sequences yielded scores above the solid horizontal lines, suggesting that they contain *P. falciparum*-specific ORFs. This is contrasted with human G6PI mRNA, where the score remains at or below the horizontal line for most of the sequence. Two hundred and twenty-one clones (60 clones with putative identification and 161 clones with no putative identification) were analyzed on PREDICT to determine the percentage of clones that contain *P. falciparum*-specific long ORFs near the ends (Fig. 2). Sequence tags where there was no ORF 60 amino acids or longer above the horizontal line in any frame were considered as

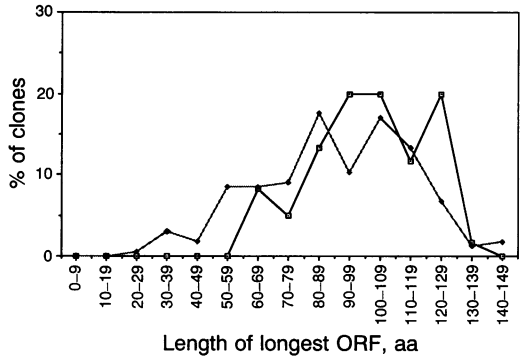

FIG. 2. Average length of longest *P. falciparum*-specific ORF among sequence tags. Sequence tags from 60 clones with putative identification (□) and 161 clones with no putative identification (◆) were analyzed by using PREDICT as in Fig. 1. The length of the longest ORF for each clone was determined. These data are presented as the percentage of clones in each group distributed in 10-amino-acid intervals along the range from 0 to 140 amino acids. The average length of the longest ORF for clones with a putative identification was 100 amino acids and that for clones with no putative identification was 89 amino acids.

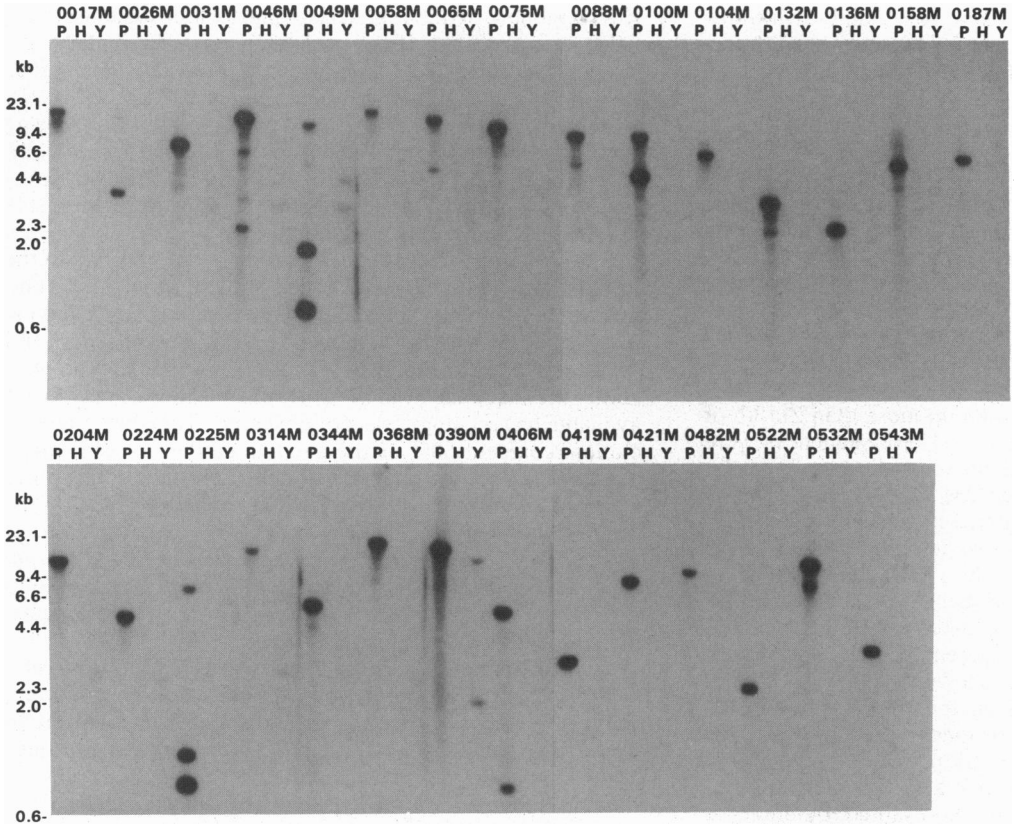

FIG. 3. Hybridization of putatively identified gene fragments to blots of genomic DNA from *P. falciparum*, human, and yeast. Genomic DNA from *P. falciparum* (1 μg), human leukocytes (7 μg), and *Saccharomyces cerevisae* clone YPH 274 (1 μg) was digested to completion with *Eco*RI, blotted onto nylon membrane, and hybridized to probes prepared from selected clones. Each filter hybridized is indicated at the top with the name of the clone used as probe, and the lanes of *P. falciparum*, human, and yeast DNA are marked as P, H, and Y, respectively.

lacking a long ORF. All 60 putatively identified clones had long ORFs (average ORF = 100 amino acids) in the same reading frame that yielded the highest-scoring segment pair. Eighty-six percent of the clones with no putative identification contained ORFs ≥60 amino acids (average ORF = 89 amino acids).

DNA Blot Analysis of Clones to Verify Derivation from *P. falciparum* Genome. Probes prepared from 32 putatively identified and 3 unidentified mung bean nuclease-prepared clones were hybridized to Southern blots of genomic DNA. Clones used include those for which the BLAST scores were ≥150 and 0065M, 0075M, and 0314M. All probes hybridized strongly with *P. falciparum* genomic DNA, but none hybridized to human genomic DNA (Fig. 3). Only 0049M and 0390M cross-hybridized weakly with yeast genomic DNA. Hybridizations for clones 0025M, 0030M, 0127M, 0202M, 0229M, and 0337M were performed at different times (data not shown).

DISCUSSION

Genes expressed in the erythrocytic, exoerythrocytic, and sexual stages of parasite development were identified among the clones reported here. These results suggest that this approach using genomic DNA will offer genes expressed in all stages of the life cycle. Access to all of the genes of the parasite is essential for establishing a gene sequence data base from which to approach the biology of the parasite.

The presence of introns did not significantly hinder the genease approach. Under the conditions of mung bean nuclease digestion selected, nearly one-third of the clones contain intact genes. The majority of the other two-thirds appear to be clones of one or more exons, suggesting that these genes contain one or more introns that are sensitive to mung bean nuclease digestion. Previously, in 9 of 10 genes analyzed, the major reaction product consisted of a single DNA fragment which contained the entire extent of the coding sequence, including introns where present (9). It thus appears that conditions of the digestion may be modified to release genes as intact coding sequences or exons. Conditions selected here may have resulted in a greater proportion of the introns being cleaved. However, they proved useful for identifying clones based on sequence similarity, since only ≈14% of the clones lacked *P. falciparum*-specific long ORFs located near the ends. Further, a proportion of these are likely to be gene fragments with long noncoding sequences at the ends or fragments containing exons encoding fewer than 60 amino acids. These results suggest that either little of the noncoding DNA of the parasite genome remains after mung bean nuclease digestion or it is not cloned. This is further supported by the fact that even though only clones containing ≥0.3-kb inserts were selected for sequencing, <20% of the overall library was rejected for small insert size (data not shown).

Recent reports identifying a major contamination of randomly sequenced human cDNA clones with nonhuman DNA (26, 27) raise a concern about the integrity of sequences generated from random clones. The results of 221 clones analyzed with the PREDICT program strongly suggest that all the clones containing long ORFs are derived from the *P. falciparum* genome. This conclusion was further verified by DNA blot analysis of 32 putatively identified and 3 unidentified clones. All clones hybridized strongly to *P. falciparum* DNA and failed to hybridize with human DNA. Only clones 0049M and 0390M, which encode tubulin and 60S ribosomal protein L8, respectively, weakly cross-hybridized to yeast DNA. This may be understood, since these genes have regions highly conserved among diverse species. Microbial contamination of the *in vitro* cultures was not observed, but a small amount of human DNA (<1%) was present in the DNA used for cloning. Since all of the 35 clones examined were from the *P. falciparum* genome, we are 95% confident that ≥90% of the clones are *P. falciparum* sequences (28). BLAST searches have not identified any clones as human, either gene coding regions or repetitive sequences. However, it is expected that any further work on clones of interest will

include first confirming that the sequence is from *P. falciparum*.

The malaria parasite has $\approx 1.5 \times 10^7$ bp of unique DNA (29); thus the genome may encode ≈7500 genes, assuming an average of 2 kb per gene. This set of 400 clones thus represents ≈5% of the genes of the parasite. In the UNDP/World Bank/WHO-TDR Malaria Sequence Database, sequences of fewer than 150 different genes were reported from *P. falciparum*. Our data add over 200 kb of genomic sequence from ≈360 unique clones and allow an initial, tentative identification of 46 previously undescribed *P. falciparum* genes. The majority of the unidentified clones are likely to be genes, the identity of which may be determined in the future.

Given the fundamental differences between the cDNA sequencing approach and the present genes-first genomic DNA approach, it is remarkable that the proportion of clones identified here (15%) by similarity to gene sequences in the data bases is comparable to that of human brain cDNA (17%) (3, 4). The proportion is less than half that of *C. elegans* cDNAs (31–42%), but a significant percentage of the identified *C. elegans* clones are redundant (5, 6). The redundancy among the mung bean clones is 10%, which is twice the rate for human brain cDNA (5%) (3) but less than half that for *C. elegans* cDNA (24%) (6) with a similar number of clones examined. These statistics are consistent with the expectation that a genomic gene library from a parasitic protozoan would contain a low percentage of redundant clones and a high proportion of genes which are not yet described in the data base as compared to a cDNA library. The relative abundance of various gene fragments in the library of mung bean digestion fragments will be a function of the gene copy number, the genome size, the mung bean nuclease reaction, and the effects of cloning methods used. Preliminary results comparing *P. falciparum* expressed sequence tags from erythrocytic stage mRNA with the mung bean nuclease sequence tags suggest that the frequency of redundancy in cDNA is higher and a different set of genes is identified (unpublished results).

Numerous avenues of research into the biochemistry of the parasite have been opened by using the genes identified to date. Clones encoding several ribosomal proteins, heat shock proteins, GTP-binding proteins, proteases, protein kinases, microtubular proteins, metabolic enzymes, iron-responsive element, alternative splicing factor, and cyclophilin are now available to initiate biochemical analyses. In addition to these better-understood gene identifications, finding a gene closely related to nitrogen fixation U protein raises questions about the significance of such a protein in a malaria parasite. This may be further evidence of the green ancestry of malarial parasites (30).

Recombinant expression of certain of these genes will provide proteins in quantities needed for targeted drug development. Hemoglobin proteolysis in the digestive vacuoles of malaria parasites is a biochemical pathway unique to *Plasmodium* spp. (31) and thus an excellent target for developing new therapeutics (32, 33). This ordered process requires initiation by an aspartic protease (31, 32). Of the three proteases putatively identified (Table 1), one (0075M) is a full-length aspartic protease (cathepsin D-like protein). Similarly, the gene for cyclophilin is putatively identified. This protein is thought to be the enzymatic target for cyclosporin A (34), a drug known to inhibit the growth of malaria parasites (35, 36). Transcripts for these two gene products were detected in the erythrocytic stages of this parasite by two independent methods (unpublished results). Gene sequence tags generated in this study can also be used to construct a fine-resolution chromosomal map.

We thank M. W. Weig for *in vitro* cultivation of the asexual stage parasites. We are grateful to A. Saul for providing the PREDICT computer program. We are indebted to T. P. Yang, T. C. Rowe, P. J. Laipis, H. S. Nick, A. F. Cockburn, N. Kamel, and D. R. Allred for their helpful discussions in the formative stages of the project. We thank R. C. Littell for assistance with the statistical analyses. We thank T. F. McCutchan for helpful discussions on this manuscript. This study was supported by a program grant from the Division of Sponsored Research at the University of Florida, Gainesville. This paper is University of Florida Agricultural Experiment Stations Journal Series no. R-03376.

1. World Health Organization (1990) *Practical Chemotherapy of Malaria* (WHO, Geneva).
2. Gilbert, W. (1991) *Nature (London)* **349**, 99.
3. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. & Venter, J. C. (1991) *Science* **252**, 1651–1656.
4. Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Fields, C. & Venter, C. J. (1992) *Nature (London)* **335**, 632–634.
5. Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J. & Sulston, J. (1992) *Nature Genet.* **1**, 114–123.
6. McCombie, W. R., Adams, M. D., Kelley, J. M., FitzGerald, M. G., Utterback, T. R., Khan, M., Dubnick, M., Kerlavage, A. R., Venter, J. C. & Fields, C. (1992) *Nature Genet.* **1**, 124–131.
7. McCutchan, T. F., Hansen, J. L., Dame, J. B. & Mullins, J. A. (1984) *Science* **225**, 625–628.
8. Dame, J. B., Williams, J. L., McCutchan, T. F., Weber, J. L., Wirtz, R. A., Hockmeyer, W. T., Maloy, W. L., Haynes, J. D., Schneider, I., Roberts, D., Sanders, G. S., Reddy, E. P., Diggs, C. L. & Miller, L. H. (1984) *Science* **225**, 593–599.
9. Vernick, K. D., Imberski, R. B. & McCutchan, T. F. (1988) *Nucleic Acids Res.* **16**, 6883–6896.
10. Szafranski, P. & Godson, G. N. (1990) *Gene* **88**, 141–147.
11. Bzik, D. J., Peck, J. Y. & Fox, B. A. (1993) *Mol. Biochem. Parasitol.* **56**, 185–188.
12. Brown, K. H., Brentano, S. T. & Donelson, J. E. (1986) *J. Biol. Chem.* **261**, 10352–10358.
13. Adam, R. D., Aggarwal, A., Lal, A. A., de la Cruz, V. F., McCutchan, T. F. & Nash, T. E. (1988) *J. Exp. Med.* **167**, 109–118.
14. Johnson, A. M., Illana, S., Dubey, J. P. & Dame, J. B. (1987) *Exp. Parasitol.* **63**, 272–278.
15. Muhich, M. L. & Simpson, L. (1986) *Nucleic Acids Res.* **14**, 5531–5556.
16. Tetzlaff, C. L., McMurray, D. N. & Rice-Ficht, A. C. (1990) *Mol. Biochem. Parasitol.* **40**, 183–192.
17. Tripp, C. A., Wagner, G. G. & Rice-Ficht, A. C. (1989) *Exp. Parasitol.* **69**, 211–225.
18. Trager, W. & Jensen, J. B. (1976) *Science* **193**, 673–675.
19. Dame, J. B. & McCutchan, T. F. (1983) *J. Biol. Chem.* **258**, 6984–6990.
20. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
21. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), p. 395.
22. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
24. Weber, J. L. (1988) *Exp. Parasitol.* **66**, 143–170.
25. Saul, A. & Battistutta, D. (1988) *Mol. Biochem. Parasitol.* **27**, 35–42.
26. Anderson, C. (1993) *Science* **259**, 1685.
27. Gersuk, V. H. & Rose, T. M. (1993) *Science* **260**, 605.
28. Diem, K. & Lentner, C., eds. (1970) *Documenta Geigy Scientific Tables* (Geigy Pharmaceuticals, Ardsley, NY), 7th Ed., pp. 85–103.
29. Dore, E., Birago, C., Frontali, C. & Battaglia, P. A. (1980) *Mol. Biochem. Parasitol.* **1**, 199–208.
30. Palmer, J. D. (1992) *Curr. Biol.* **2**, 318–320.
31. Goldberg, D. E., Slater, A. F. G., Cerami, A. & Henderson, G. B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2931–2935.
32. Goldberg, D. E., Slater, A. F. G., Beavis, R., Chait, B., Cerami, A. & Henderson, G. B. (1991) *J. Exp. Med.* **173**, 961–969.
33. Miller, L. H. (1992) *Science* **257**, 36–37.
34. Gasser, C. S., Gunning, D. A., Budelier, K. A. & Brown, S. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9519–9523.
35. Thommen-Scott, K. (1981) *Agents Actions* **11**, 770–773.
36. Nickell, S. P., Scheibel, L. W. & Cole, G. A. (1982) *Infect. Immun.* **37**, 1093–1100.