Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 7100–7104, July 1995
Genetics

Isolation of multiple sequences from the *Plasmodium falciparum* genome that encode conserved domains homologous to those in erythrocyte-binding proteins

(malaria/genetic variation)

DAVID S. PETERSON, LOUIS H. MILLER, AND THOMAS E. WELLEMS

Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Louis H. Miller, January 3, 1995

ABSTRACT **Open reading frames in the *Plasmodium falciparum* genome encode domains homologous to the adhesive domains of the *P. falciparum* EBA-175 erythrocyte-binding protein (*eba-175* gene product) and those of the *Plasmodium vivax* and *Plasmodium knowlesi* Duffy antigen-binding proteins. These domains are referred to as Duffy binding-like (*DBL*), after the receptor that determines *P. vivax* invasion of Duffy blood group-positive human erythrocytes. Using oligonucleotide primers derived from short regions of conserved sequence, we have developed a reverse transcription–PCR method that amplifies sequences encoding the *DBL* domains of expressed genes. Products of these reverse transcription–PCR amplifications include sequences of single-copy genes (including *eba-175*) and variably transcribed genes that cross-hybridize to multiple regions of the genome. Restriction patterns of the multicopy genes show a high degree of polymorphism among different parasite lines, whereas single-copy genes are generally conserved. Characterization of the single-copy genes has identified a gene (*ebl-1*) that is related to *eba-175* and is likely to be involved in erythrocyte invasion.**

Sequence analysis of a *Plasmodium falciparum* chromosome linked to chloroquine resistance has identified open reading frames that contain regions related to conserved adhesive domains of an erythrocyte-binding protein of *Plasmodium falciparum* (EBA-175) and the Duffy antigen-binding proteins (DABPs) of *Plasmodium vivax* and *Plasmodium knowlesi* (this work; ref. 1). These domains, termed "Duffy binding-like" (*DBL*) after the receptor that determines *P. vivax* invasion of Duffy blood group-positive human erythrocytes, occur within cysteine-rich domains that determine binding of the EBA-175 and DABP ligands to surface receptors on erythrocytes (2–4). The presence of *P. falciparum DBL* domains in many different molecules suggests that they have a broad role in diverse receptor–ligand interactions. Conserved sequences within the domains may therefore provide the basis to isolate and identify important determinants of cell–cell interactions in malaria.

Here we describe the cloning of transcribed *DBL* sequences by reverse transcription (RT)–PCR amplifications that employ oligonucleotide primers designed from relatively conserved motifs of *DBL* domains. Several novel *DBL*-containing sequences are identified, including transcripts that recognized multiple cross-hybridizing genes and a single-copy gene (*ebl-1*) that is homologous to *eba-175* and may be involved in invasion of erythrocytes by malaria parasites.*

MATERIALS AND METHODS

***P. falciparum* Clones and Lines.** *P. falciparum* clones 3D7, D10, LF4/1, MCamp/A1, SL/D6, HB3, 7G8, V1/S, T2/C6, KMW II, ItG2F6, FCR3/A2, and Dd2 have been previously tabulated (5). Line Dd2/NM1 was selected from clone Dd2 for invasion via a sialic acid-independent pathway (6). All parasites were maintained *in vitro* by standard methods (7).

DNA and RNA Isolation and Analysis. DNA was extracted as described (8). Endonuclease digestion, agarose gel electrophoresis, and filter hybridizations were performed by standard methods (9). All hybridizations were at 56°C (9). Blots were washed for 2 min at room temperature in 2× standard saline/phosphate/EDTA (SSPE) with 0.5% SDS, followed by two higher-stringency washes at 50°C in 0.3 × SSPE with 0.5% SDS. Parasite chromosomes were embedded in agarose blocks and separated by pulsed-field gel electrophoresis (10). RNA was isolated from cultured parasites by LiCl extraction of Catrimox 14-precipitated RNA (11). Agarose gel electrophoresis of total RNA and filter hybridizations were performed by standard methods (9).

Oligonucleotide Primers and PCR. Primers specific for E31a used in a RT–PCR to test for expression of this sequence were E31aT2 (5′-AGA-CCT-CAA-TTT-CTA-AG-3′) and E31aRev1 (5′-AAT-CGC-GAG-CAT-CAT-CTG-3′).

Two primers were used to amplify additional sequences from genes encoding *DBL* domains. These were designed from conserved amino acids encoded in the *DBL* domain of the *eba-175* and E31a sequences. After adaptation to incorporate the most frequently used *P. falciparum* codons, forward primer UNIEBP5′ [5′-CC(A/G)-AG(G/A)-AG(G/A)-CAA-(G/A)AA-(C/T)TA-TG-3′], based upon the amino acid sequence PRRQKLC, and reverse primer UNIEBP3′ [5′-CCA-(A/T)C(T/G)-(T/G)A(A/G)-(A/G)AA-TTG-(A/T)GG-3′], based upon the amino acid sequence PQFLRW, were synthesized.

RT–PCR amplifications were performed as described (12). In brief, 0.5–1 mg of total RNA was treated with RQ1 DNase (Promega), phenol/chloroform extracted, and ethanol precipitated. The RNA was then annealed with random oligonucleotide primers and extended with Superscript reverse transcriptase (GIBCO/BRL). PCR cycling conditions were 94°C for 10 sec, 45°C for 15 sec, and 72°C for 45 sec, for 30 cycles. All PCRs were performed in an Idaho Technology air thermal cycler using buffer containing 2 mM Mg^{2+}.

PCR amplification products were separated by use of PCR Purity Plus gels and protocols (AT Biochem, Malvern, PA).

DNA Clones and Hybridization Probes. Clone pE31a was isolated from a genomic library prepared from the region of chromosome 7 linked to chloroquine resistance (13). Clone pS31H (GenBank accession no. L38454), containing an insert encompassing that of pE31a, was cloned from a size-selected *Hin*dIII restriction digest of Dd2 genomic DNA.

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: DABP, Duffy antigen-binding protein; RT, reverse transcription.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L38450, L38453, L38454, and L38455).

Clone pEBLe1 was cloned from a RT–PCR of Dd2 cDNA after amplification with primers UNIEBP5′ and UNIEBP3′. Clone pEBP1.2 (GenBank accession no. L38450), containing an insert encompassing that of pEBLe1, was isolated from a Dd2 cDNA library probed with pEBLe1. *DBL*-encoding sequences of *dbl-nm1-4* (GenBank accession no. L38455) and *dbl-nm1-5* (GenBank accession no. L38453) were amplified by RT–PCR from first-strand cDNA of line Dd2/NM using primers UNIEBP5′ and UNIEBP3′. Sequencing was performed on double-stranded DNA templates by standard protocols for the dideoxynucleotide method (Sequenase; United States Biochemical).

Sequences related to the E31a sequence were detected with the 3005-bp insert of clone pS31H. The *eba-175* gene was detected with a PCR-amplified probe consisting of the first 1825 bp of the coding sequence. *ebl-1* sequences were detected with the 2098-bp insert of clone pEBP1.2. All probes were comparable in organization, each containing a region encoding at least one *DBL* domain and various amounts of flanking sequence.

Homology Searches and Alignments. Homology searches were performed with BLAST and the Genetics Computer Group program FASTA (14, 15). Optimized alignments were produced with MACAW sequence alignment software (16). Alignments were prepared for publication with the computer programs BOXSHADE (Kay Hofmann, Bioinformatics Group, Institut Suisse de Recherchas Experimentales sur le Cancer, CH-1066 Epalinges s/Lausanne, Switzerland) and DNADRAW (17).

RESULTS

Multiple *P. falciparum* Sequences Encode *DBL* Domains. Positional cloning experiments directed to *P. falciparum* chromosome 7 identified an open reading frame (E31a) encoding a *DBL* domain that is homologous to the domains found in the *P. vivax* and *P. knowlesi* DABPs and the *P. falciparum* glycophorin-binding protein EBA-175. Fig. 1*a* shows the relative position of the E31a open reading frame on chromosome 7.

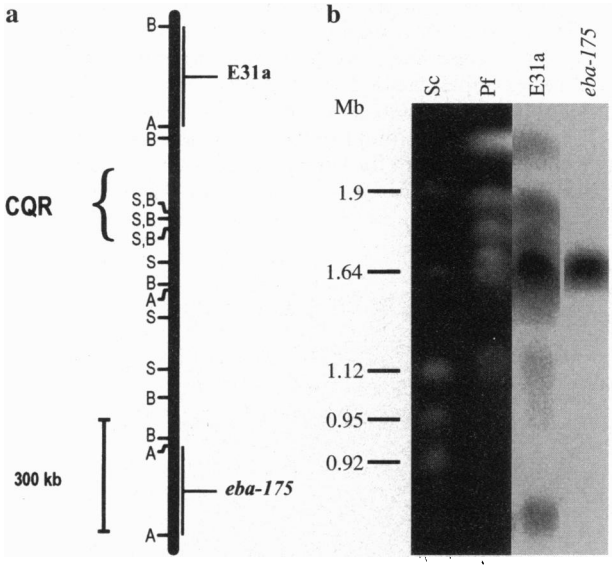

FIG. 1. Chromosome mapping of the E31a sequence. (*a*) E31a and *eba-175* sequences map to opposite ends of chromosome 7. Restriction enzymes: A, *Apa* I; B, *Bgl* I; S, *Sma* I. CQR: limits of the chromosome segment linked to chloroquine response in the *P. falciparum* HB3 × Dd2 cross. (*b*) Hybridization signals from *P. falciparum* (Pf) chromosomes separated by pulsed-field gel electrophoresis. The E31a probe detects cross-hybridizing sequences on multiple chromosomes whereas the *eba-175* probe hybridizes to chromosome 7 only. Yeast (*Saccharomyces cerevisiae,* Sc) chromosomes were run as size markers. Mb, megabases.

The homology between the *DBL* domains of E31a and the erythrocyte-binding proteins is due to the presence of short motifs of highly conserved amino acids. These well-conserved stretches (Fig. 2) are separated by nonhomologous sequences and by deletions and insertions that vary the size of the domain by >60 aa. The typical *DBL* domain contains 12 or more cysteine residues and has 7 conserved tryptophan residues. Additional well-conserved amino acids include 4 arginines, 3 aspartates, 9 positions with aliphatic residues (alanine, isoleucine, leucine, or valine), and 4 with aromatic amino acids (tryptophan, phenylalanine, or tyrosine).

Probes spanning the sequence that encodes the E31a *DBL* domain hybridized to multiple fragments within a single restriction digest and yielded bands that varied among parasite lines (Fig. 3). The numerous distinct bands from a selection of different parasite DNAs indicated a large number of diverse but related elements (Fig. 3*a*). These multiple bands varied among different *P. falciparum* clones, in contrast to the well-conserved, single-copy signal obtained with the *eba-175* probe (Fig. 3*b*).

Because of the numerous cross-hybridizing sequences, it seemed likely that many of these related sequences would be on different chromosomes of the parasite. Fig. 1*b* shows the chromosomal distribution of sequences detected by the E31a probe. Although the predominant hybridization signal maps to the chromosome 7/9 doublet of the Dd2 parasite, the presence of hybridization signals throughout the pulsed-field gel separation indicates that *DBL* domains are encoded by sequences in most, if not all, *P. falciparum* chromosomes. A control hybridization with the *eba-175* probe under identical conditions yielded a single band of hybridization from chromosome 7.

RNA Analysis of *DBL* Elements. Sequences from E31a (pS31H insert) were used to probe RNA blots for corresponding transcripts. No hybridization was detected. Because it was still possible that a message of low abundance was not being detected on the RNA blot, RT–PCR was used as a means of more sensitive detection. For this purpose, cDNA was generated by RT from random primers annealed to DNase-treated total RNA. E31a-specific oligonucleotides were then used to test for amplification from the cDNA. No amplification of the E31a sequence was obtained, whereas genomic DNA controls and amplification from cDNA by dihydrofolate reductase/thymidylate synthetase-specific primers yielded the expected bands (data not shown). A screen of a cDNA library with E31a-specific probes also failed to detect any clones hybridizing with the open reading frame. These results indicate that E31a either is a pseudogene or is expressed in parasite strains or stages not examined in this work.

A PCR Method to Isolate Sequences Encoding *DBL* Domains. The identification of short conserved motifs in *DBL* domains that otherwise have extreme diversity led us to consider general ways of cloning the corresponding sequences from genomic DNA or RNA transcripts. To this end, we developed a PCR strategy using degenerate oligonucleotide primers designed from conserved amino acid sequences in the *DBL* domains. Sequences PRRQKLC and PQFLRW (underlined in Fig. 2) were judged most suitable for minimizing degeneracy while allowing amplification of expressed *DBL* sequences. After these considerations and adjustment for *P. falciparum* codon usage, primers UNIEBP5′ and UNIEBP3′ were synthesized.

Fig. 4*a* presents results of PCR amplifications in which the UNIEBP primers were applied to genomic DNAs of different *P. falciparum* lines. While some samples yielded similar patterns of amplified bands (e.g., Dd2 and MCamp; FCR3/A2 and K-1), no two separate isolates showed identical patterns, reflecting the diversity of the *DBL* domains in the parasite lines. A few bands of the same apparent size were present in

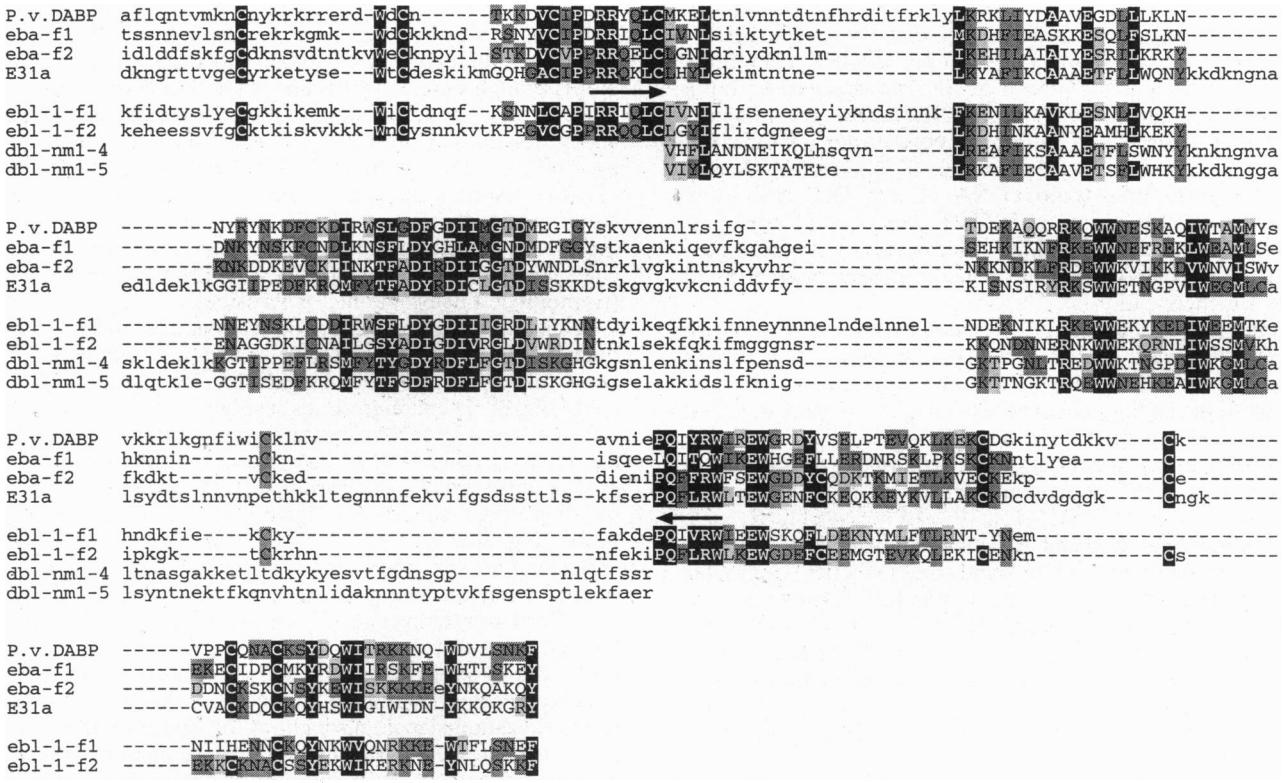

FIG. 2. Amino acid alignment of conserved *DBL* domains from *P. falciparum eba-175*, the *P. vivax* DABP gene, and five other domains of the *P. falciparum DBL* family. The *eba-175* f1 and f2 domains are those of Adams *et al* (2); corresponding regions are marked similarly for *ebl-1*. Frequently conserved residues are shaded; more strictly conserved residues are highlighted in white-on-black type. Locations corresponding to the UNIEBP PCR primers are underlined with arrows.

many isolates. These included a consistent 490-bp product that was determined to be the *eba-175* gene by its expected size and hybridization to a gene-specific probe (data not shown). We note that the number of discernible bands probably underestimates the number of amplifiable sequences because of overlapping products of the same size and possible preferential amplification of some sequences over others. Nevertheless, the parasite-specific patterns in the amplified bands may provide a means to quickly type isolates and could serve as a measure of parasite diversity in field samples.

To identify *DBL*-encoding sequences in RNA transcripts, we used the UNIEBP primers to amplify first-strand cDNAs generated from DNase-treated RNA preparations. Amplified products from Dd2, 3D7, HB3, and MCamp cDNAs had diverse sizes ranging from 400 bp to nearly 1 kb (Fig. 4*b*). These included a band at 480–500 bp that was determined to be *eba-175* from its expected size and cross-hybridization to an *eba-175*-specific probe. Other bands were from amplification of different transcripts encoding *DBL* domains. Dd2/NM1 RNA, for example, yielded bands above the *eba-175* product that included two related sequences (*dbl-nm1-4*, and *dbl-nm1-5*). These bands were found to be isolate-specific and to have features consistent with the *var* genes described by Su *et al.* (1). Probes that detect *dbl-nm1-4* and *dbl-nm1-5* hybridized to

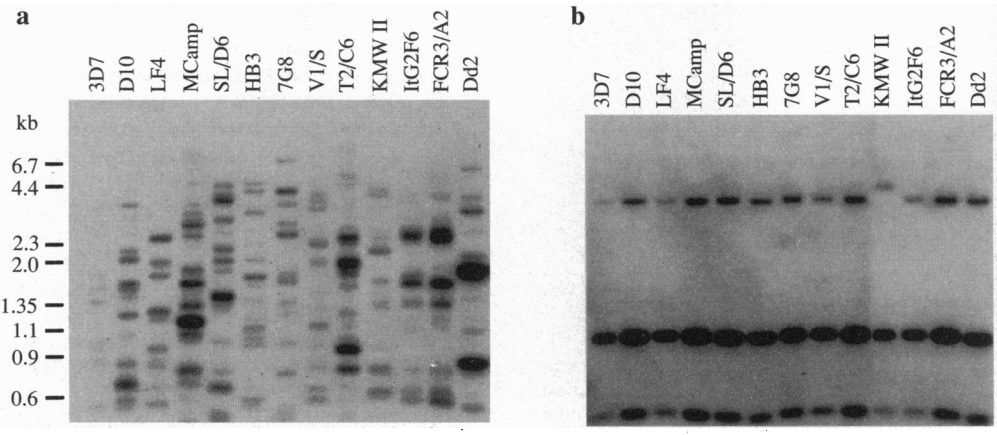

FIG. 3. Southern hybridization patterns demonstrating E31a diversity in 13 *P. falciparum* lines. Signals shown in *a* and *b* were from hybridizations with E31a and *eba-175* sequences, respectively. Genomic DNAs were digested with *Hin*fI and fractionated in a 1.0% agarose gel. Between hybridizations, the blot was stripped with boiling 0.5% SDS and removal of probe was verified by autoradiography. Three *Hin*f1 fragments are predicted from the sequence of the E31a probe, one of 1680 bp, and two others that are greater than 568 and 757 bp. Additional bands detected in the blot derive from cross-hybridizing sequences. The *eba-175* probe detects bands of 331 and 924 bp and one of 3500 bp in the homologous gene because of *Hin*fI sites in that part of the gene corresponding to the probe.

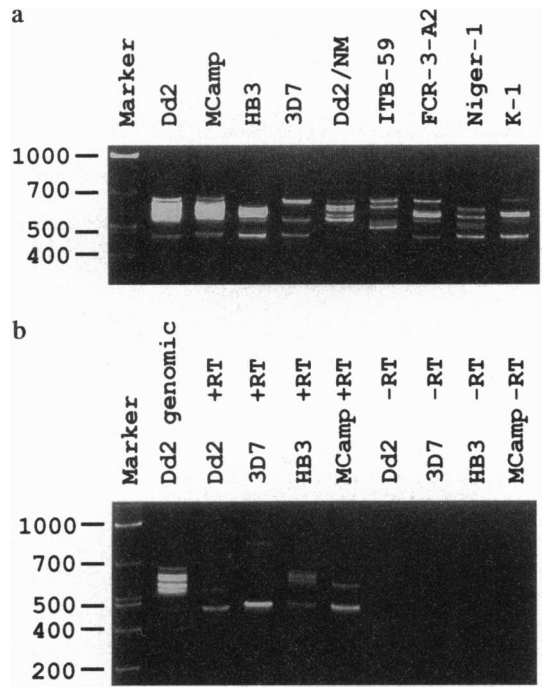

FIG. 4. Diversity in *DBL* amplification products from the genomic DNA and cDNA of various *P. falciparum* lines. (*a*) PCR amplification of genomic DNA with UNIEBP primers. A unique pattern of amplified fragments is evident in each lane. (*b*) PCR amplification of first-strand cDNA (+RT) with UNIEBP primers. The amplifications from cDNA of Dd2, 3D7, HB3, and MCamp parasites produced 0.5-kb bands from the *eba-175* gene as well as diverse products from other genes. Control amplifications of mock RT reactions performed without the addition of reverse transcriptase (−RT) are shown at the right. Marker lanes show relative size standards in base pairs. An amplification of Dd2 genomic DNA is shown (second lane from the left).

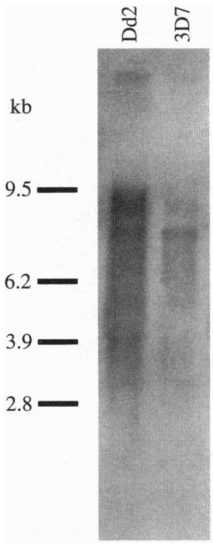

FIG. 5. Signals from *ebl-1* transcripts in two different *P. falciparum* lines. Samples (15 μg) of total RNA from the Dd2 and 3D7 lines were fractionated by electrophoresis, blotted to Nytran membrane (Schleicher & Schuell), and hybridized against an *ebl-1*-specific probe (pEBP1.2).

multiple chromosomes (data not shown) and aligned more closely with E31a than with EBA-175 or the DABPs (Fig. 2).

The RT–PCR amplifications also yielded a consistent band that encoded a *DBL* domain distinct from *eba-175*. A cDNA clone corresponding to this product was isolated by screening a λgt10 Dd2 cDNA library with a radiolabeled *ebl-1* probe. Sequence from this and additional overlapping cDNA clones confirmed the conserved motifs of the *DBL* domain. The alignment of the predicted amino acid sequences showed that the *DBL* domains of *ebl-1* were more similar to *eba-175* than to the multicopy genes (Fig. 2). There was, however, extensive divergence from *eba-175* and other known genes outside of the amplified region.

In contrast to the multicopy hybridization patterns of *dbl-nm1-4* and *dbl-nm1-5*, the *ebl-1* sequence, like that of *eba-175*, was found to have hybridization patterns consistent with a conserved single-copy gene. Probes specific for *ebl-1* hybridized only to chromosome 13, and restriction analysis with the enzymes *Cla* I, *Eco*RI, *Hin*dIII, *Hin*fI, *Nsi* I, *Rsa* I, and *Spe* I all yielded bands expected from a single-copy sequence. RNA blots probed with *ebl-1*-specific sequences showed several bands of hybridization, however, corresponding to 8- to 9.5-kb transcripts in mRNA from the Dd2 and 3D7 parasites (Fig. 5). The transcripts of different size may result from alternative start and termination points or from incompletely processed species containing introns.

DISCUSSION

Two general classes of *P. falciparum* transcripts have been isolated that encode *DBL* domains homologous to those in erythrocyte-binding proteins. One class, represented by the E31a, *dbl-nm1-4*, and *dbl-nm1-5* sequences, comprises cross-hybridizing members whose transcription varies among parasite lines. The genes that produce these transcripts vary in number and yield diverse restriction patterns among parasite lines. Already, certain of these transcripts are known to be from the *var* family of genes that modulate cytoadherence and antigenic variation of *P. falciparum*-infected erythrocytes (1, 18, 19). It remains possible that additional multiple-copy gene families will be identified as additional transcripts of this *DBL* class are characterized and sequenced.

A second class of *DBL*-encoding transcripts includes those representing single-copy genes such as *P. falciparum eba-175* and *ebl-1*. Both of these genes show restriction patterns that are well conserved among different parasite isolates. This conservation of gene structure and the sequence relationships between the *ebl-1* and *eba-175* domains suggest that *ebl-1* may encode an erythrocyte-binding molecule having receptor properties distinct from those of EBA-175. Indeed, studies have demonstrated the existence of multiple erythrocyte invasion pathways that are dependent upon the recognition of different surface ligands by *P. falciparum* merozoites (20). *P. falciparum* receptor molecules in some of these different pathways may contain alternative *DBL* sequences that determine binding specificity in invasion processes.

1. Su, X., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A. & Wellems, T. E. (1995) *Cell*, in press.
2. Adams, J. H., Sim, B. K., Dolan, S. A., Fang, X., Kaslow, D. C. & Miller, L. H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7085–7089.
3. Sim, B. K., Chitnis, C. E., Wasniowska, K., Hadley, T. J. & Miller, L. H. (1994) *Science* **264,** 1941–1944.
4. Chitnis, C. E. & Miller, L. H. (1994) *J. Exp. Med.* **180,** 497–506.
5. Dolan, S. A., Herrfeldt, J. A. & Wellems, T. E. (1993) *Mol. Biochem. Parasitol.* **61,** 137–142.
6. Dolan, S. A., Miller, L. H. & Wellems, T. E. (1990) *J. Clin. Invest.* **86,** 618–624.
7. Trager, W. & Jensen, J. B. (1976) *Science* **193,** 673–675.
8. Peterson, D. S., Milhous, W. K. & Wellems, T. E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3018–3022.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, New York), 2nd Ed.
10. Dolan, S. A., Adam, R. D. & Wellems, T. E. (1993) *Methods Mol. Biol.* **21,** 319–332.

11. Dahle, C. & Macfarlane, D. (1993) *BioTechniques* **15,** 1102–1105.
12. Kawasaki, E. S. (1990) in *PCR Protocols, A Guide to Methods and Applications*, eds. Innis, M. A., Gelfand, D. H., Sninsku, J. J. & White, T. J. (Academic, San Diego), pp. 21–27.
13. Walker-Jonah, A., Dolan, S. A., Gwadz, R. W., Panton, L. J. & Wellems, T. E. (1992) *Mol. Biochem. Parasitol.* **51,** 313–320.
14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
15. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
16. Schuler, G. D., Altschul, S. F. & Lipman, D. J. (1991) *Proteins* **9,** 180–190.
17. Shapiro, M. B. & Senapathy, P. (1986) *Nucleic Acids Res.* **14,** 65–73.
18. Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F. & Howard, R. J. (1995) *Cell*, in press.
19. Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I. & Miller, L. H. (1995) *Cell*, in press.
20. Dolan, S. A., Proctor, J. L., Alling, D. W., Okubo, Y., Wellems, T. E. & Miller, L. H. (1994) *Mol. Biochem. Parasitol.* **64,** 55–63.