Data-Mining Approaches Reveal Hidden Families of Proteases in the Genome of Malaria Parasite

Yimin Wu, Xiangyun Wang, Xia Liu, et al.

Genome Res. 2003 13: 601-616
Access the most recent version at doi:10.1101/gr.913403

Letter

Data-Mining Approaches Reveal Hidden Families of Proteases in the Genome of Malaria Parasite

Yimin Wu,[1,4] Xiangyun Wang,[2] Xia Liu,[1] and Yufeng Wang[3,5]

[1]*Department of Protistology, American Type Culture Collection, Manassas, Virginia 20110, USA;* [2]*EST Informatics, Astrazeneca Pharmaceuticals, Wilmington, Delaware 19810, USA;* [3]*Department of Bioinformatics, American Type Culture Collection, Manassas, Virginia 20110, USA*

The search for novel antimalarial drug targets is urgent due to the growing resistance of *Plasmodium falciparum* parasites to available drugs. Proteases are attractive antimalarial targets because of their indispensable roles in parasite infection and development, especially in the processes of host erythrocyte rupture/invasion and hemoglobin degradation. However, to date, only a small number of proteases have been identified and characterized in *Plasmodium* species. Using an extensive sequence similarity search, we have identified 92 putative proteases in the *P. falciparum* genome. A set of putative proteases including calpain, metacaspase, and signal peptidase I have been implicated to be central mediators for essential parasitic activity and distantly related to the vertebrate host. Moreover, of the 92, at least 88 have been demonstrated to code for gene products at the transcriptional levels, based upon the microarray and RT-PCR results, and the publicly available microarray and proteomics data. The present study represents an initial effort to identify a set of expressed, active, and essential proteases as targets for inhibitor-based drug design.

[Supplemental material is available online at www.genome.org.]

Malaria remains one of the most dangerous infectious diseases in the world. It kills 1–2 million people each year, and is responsible for enormous economic burdens in endemic regions. The development of new antimalarial drugs is urgently needed due to the continuing high mortality and morbidity caused by malaria and the increasing prevalence of drug-resistance in the pathogenic parasite *Plasmodium falciparum*.

Malarial proteases have long been considered potential targets for chemotherapy due to their crucial roles in the parasite life cycle, and the feasibility of designing specific inhibitors (for reviews, see McKerrow et al. 1993; Rosenthal 1998; Blackman 2000; Rosenthal 2002). Efforts to identify functional proteases targeted by inhibition assays are ongoing. Subtilase-1 and Subtilase-2, two homologous serine proteases, are demonstrated to be involved in schizont rupture and merozoite invasion (Blackman et al. 1998; Barale et al. 1999; Hackett et al. 1999). Cysteine proteases have also been implicated in the rupture/invasion process (Salmon et al. 2001). A cluster of Serine Repeat Antigens (SERAs) exhibit limited sequence similarity to cysteine proteases, though their proteolytic activity remains undocumented (Delplace et al. 1988; Miller et al. 2002). A zinc-metallo-aminopeptidase has also been demonstrated to possess enzymatic activity (Florent et al. 1998). Meanwhile, three classes of proteases have been identified to be involved in hemoglobin degradation: (1) Four aspartic proteases (plasmepsin I, II, IV, and HAP) (see Banerjee et al. 2002 for a review); (2) three cysteine proteases (falcipain-1, -2, and -3) (see Rosenthal 2002 for a review); and (3) one metalloprotease (falcilysin; Eggleson et al. 1999). The successful crystallization of plasmepsin II and the expression of recombinant plasmepsin I/II and falcipain-2 represented a significant advance towards a functional understanding and a rational design of inhibitors of these enzymes (Silva et al. 1996; Bernstein et al. 1999; Tyas et al. 1999; Shenai et al. 2000; Dua et al. 2001).

The recent completion of the *P. falciparum* genome provides a basis on which to identify new proteases. The first pass annotation has predicted 25 proteases that belong to ten families of five catalytic classes (Table 1). Despite this initial progress, direct evidence from protease inhibition assays and independent comparisons with other genomes suggest that in addition to the limited number of characterized and predicted proteases, many important proteolytic enzymes remain uncharacterized (Olaya and Wasserman 1991; Southan 2001). The following six sets of experimental data suggest that unidentified proteases are responsible for additional critical hydrolytic activities: (1) A calpain-type protease, which appears to be involved in merozoite invasion of red blood cells (Olaya and Wasserman, 1991); (2) an entire group of threonine proteases in the proteasome complex (Gantt et al. 1998); (3) proteases that catalyze the primary processing of Merozoite Surface Protein (MSP-1; David et al. 1984), Apical Merozite Antigen-1 (AMA-1; Narum and Thomas 1994), and the precursor of SERA (Li et al. 2002); (4) the gp76 and gp68 GPI-anchored serine proteases that cleave host erythrocyte surface proteins in *P. falciparum* and *P. chabaudi*, respectively (Braun-Breton et al. 1988); (5) a 75-kD merozoite serine protease (Rosenthal et al. 1987); and (6) a neutral aminopeptidase essential in hemoglobin digestion (Curley et al 1994). Additional supportive evidence that the majority of malarial proteases are unexplored comes from a comparison with the number of proteases found in other organisms. According to the statistics in the protease database Merops (http://www.merops.ac.uk) as released on March 18, 2002, all the model organisms possess

[4]Present address: Malaria Vaccine Development Unit, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
[5]Corresponding author.
E-MAIL ywang@atcc.org; FAX (703) 365-2740.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.913403.

Table 1. Ninety-two (92) *P. falciparum* Protease Homologs Predicted From Comparative Genomic Analysis

Catalytic class	Protease family	Protease nomenclature	Gene ID	Protease homolog with highest BLAST score: Accession (species, protease name)	E-score	Pfam domain structure: Domain ID (name)	E-score	AP[a]
Aspartic	A1	**PM I**	PF14_0076	P39898 (*P. falciparum*, PM I)	0	PF00026 (eukaryotic asp protease)	1.5e-59	+
		PM II	PF14_0077	P46925 (*P. falciparum*, PM II)	0	PF00026 (eukaryotic asp protease)	8.9e-52	–
		HAP(PM III)	PF14_0078	CAB40630 (*P. falciparum*, HAP)	0	PF00026 (eukaryotic asp protease)	3.3e-30	+
		PM IV	PF14_0075	AAC15794 (*P. malariae*, PM)	0	PF00026 (eukaryotic asp protease)	5.1e-56	–
		PM V	*PF13_0133*	Q05744 (Chicken, Cathepsin D)	9e-08	PF00026 (eukaryotic asp protease)	0.00037	+
		PM VI	*PFC0495w*	CAC20153 (*Eimeria tenella*, eimepsin)	2e-99	PF00026 (eukaryotic asp protease)	1.1e-40	–
		PM VII	*PF10_0329*	CAC20153 (*Eimeria tenella*, eimepsin)	8e-28	PF00026 (eukaryotic asp protease)	1.3e-11	+
		PM VIII	PF14_0623	CAC20153 (*Eimeria tenella*, eimepsin)	1e-35	PF00026 (eukaryotic asp protease)	7.5e-11	–
		PM IX	*PF14_0281*	AAD56283 (*Pseudopleuronectes americanus*, pepsinogen A form IIa)	1e-33	PF00026 (eukaryotic asp protease)	4.5e-26	–
		PM X	PF08_0108	AAA31096 (Pig, pepsinogen A)	1e-42	PF00026 (eukaryotic asp protease)	2.5e-26	+
Cysteine	C1	**Falcipain-1**	PF14_0553	P25805 (*P. falciparum*, falcipain-1)	0	PF00112 (Papain family)	9.5e-92	–
		Falcipain-2	PF11_0165	AAF63497 (*P. falciparum*, falcipain 2)	0	PF00112 (Papain family)	3.0e-84	+
		Falcipain-3	PF11_0162	AAF86352 (*P. falciparum*, falcipain-3)	0	PF00112 (Papain family)	4.8e-79	–
		Papain	PF11_0161	AAF97809 (*P. falciparum*, Falcipain 2)	1e-134	PF00112 (Papain family)	8.8e-84	+
		DPP I	PFL2290w	AAD02704 (Dog, DPP I)	5e-30	PF00112 (Papain family)	9.9e-16	–
		DPP I	*PFD0230c*	AAD02704 (Dog, DPP I)	6e-06	PF00112 (Papain family)	2.7e-05	+
		Cathepsin C	PF11_0174	AAL48191 (Human, Cathepsin C)	5e-23	PF00112 (Papain family)	3.3e-14	+
		SERA	*PFB0360c*	H71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	2.1e-51	–
		SERA	*PFB0325c*	F71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	2.9e-10	–
		SERA	*PFB0330c*	G71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	2.9e-46	–
		SERA	*PFB0335c*	H71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	3.4e-16	–
		SERA	*PFB0340c*	B71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	1.5e-53	–
		SERA	*PFB0345c*	C71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	1.2e-47	–
		SERA	*PFB0350c*	D71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	8.8e-51	–
		SERA	*PFB0355c*	H71617 (*P. falciparum*, SERA)	0	PF00112 (Papain family)	1.5e-49	–
		Papain	*PFI0135c*	G71617 (*P. falciparum*, SERA)	1e-177	PF00112 (Papain family)	2.2e-34	–
	C2	***Calpain***	***MAL13P1.310***	NP_497964 (*C. elegans*, Calpain)	2e-35	PF00648 (Calpain family)	8.0e-13	–
	C12	UCH1	PF14_0577	BAB47136 (Rat, UCH 13)	5e-31	PF01088 (UCH 1)	1.3e-37	–
		UCH1	PF11_0177	NP_062508 (Mouse, UCH37)	6e-30	PF01088 (UCH 1)	1.5e-17	–
	C13	GPI8p transamidase	PF11_0298	CAB96076 (*P. falciparum*, GPI8p transamidase)	1e-179	PF01650 (Peptidase C13)	2.7e-13	+
	C14	***Metacaspase***	***PF13_0289***	CAD24804 (*T. brucei*, metacaspase)	5e-32	No hits	–	–
	C19	UCH2	PFA0220w	NP_566486 (*Arabidopsis*, UBP25)	1e-12	PF00442 (UCH2)	1.2e-09	–

(continued)

Table 1. *Continued*

Catalytic class	Protease family	Protease nomenclature	Gene ID	Protease homolog with highest BLAST score: Accession (species, protease name)	E-score	Pfam domain structure: Domain ID (name)	E-score	AP[a]
		UCH2	*PFD0165w*	O57429 (Chicken, UCH2)	5e-13	PF00442 (UCH2) PF00443 (UCH2)	1.9e-06 9.2e-07	–
		UCH2	PFD0680c	AAG42755 (*Arabidopsis*, UBP14)	1e-34	PR00442 (UCH2)	8.5e-11	–
		UCH2	PFE1355c	NP_566680 (*Arabidopsis*, UBP7)	3e-30	PF00442 (UCH2) PF00443 (UCH2)	7.5e-07 1.9e-14	–
		UCH2	PFE0835w	094966 (Human, UBP19)	3e-51	PF00442 (UCH2) PF00443 (UCH2)	1.8e-11 1.7e-20	–
		UCH2	MAL7P1.147	NP_568171 (*Arabidopsis*, UBP12)	1e-54	PF00442 (UCH2) PF00443 (UCH2)	3.7e-10 1.9e-16	–
		UCH2	PFI0225w	AAC68865 (Chicken, UBP66)	1e-18	PF00442 (UCH2)	9.9e-23	–
		UCH2	PF13_0096	NP_494298 (*C. elegans*, UBP)	5e-58	PF00442 (UCH2) PF00443 (UCH2)	4.3e-08 0.00048	–
		UCH2	PF14_0145	T41069 (Fission yeast, UCH)	1e-20	PF00442 (UCH2) PF00443 (UCH2)	4.8e-05 3.1e-08	–
	C48	Sumo1 protease	PFL1635w	XP_128236 (Mouse, SUMO protease)	7e-33	PF02902 (Ulp1 protease)	2.3e-38	–
		Ulp2 peptidase	MAL8P1.157	O13769 (Fission yeast, Ulp2)	1e-06	PF02902 (Ulp1 protease)	9.2e-07	–
	C56	DJ-1 peptidase	MAL6P1.153	BAB79527 (Chicken, DJ-1 protease)	6e-16	PF01965 (DJ-1/PfpI protease)	8.1e-18	–
Metallo	M1	**AMPN**	MAL13P1.56	O96935 (*P. falciparum*, M1 peptidase)	0	PF01433 (Peptidase family M1)	6.2e-74	–
	M3	Dcp	PF10-0058	NP_601500 (*Corynebacterium glutamicum*, Zn-dependent peptidases)	1e-04	PF01432 (Peptidase family M3)	6.4e-10	–
		Neurolysin	MAL13P1.184	Q02038 (Pig, neurolysin)	4e-09	PF01432 (Peptidase family M3)	1.8e-07	–
	M16	MPPa	PFE1155c	AAF00541 (*Toxoplama gondii*, MPPa)	1e-109	PF00675 (Insulinase, family M16)	6.4e-19	–
		MPPb	PFI1625c	AAK51086 (*Avicennia* marina, MPPb)	1e-108	PF00675 (Insulinase, family M16)	1.3e-55	–
		M16 peptidase	PF11_0189	NP_593544 (yeast metallopeptidase)	1e-29	PF00675 (Insulinase, family M16)	0.0076	–
		Insulysin	PF11_0226	P35559 (Rat, Insulysin)	3e-10	PF00675 (Insulinase, family M16)	1.0e-05	–
		Falcilysin	PF13_0322	AAF06062 (*P. falciparum*, falcilysin)	0	PF00675 (Insulinase, family M16)	0.19	–
		Pitrilysin	PF14_0382	T06521 (Pea pitrilysin)	5e-27	PF00675 (Insulinase, family M16)	0.0018	–
	M17	AMPL	PF14_0439	NP_194821 (*Arabidopsis* AMPL)	2e-83	PF00883 (Peptidase family M17)	9e-150	–
	M18	DNPE	PFI1570c	AAL16034 (*Coccidioides immitis*, DNPE)	7e-55	PF02127 (Peptidase family M18)	3.6e-31	–
	M22	GCP	PF10_0299	NP_194003 (*Arabidopsis* GCP)	4e-54	PF00814 (Glycoprotease family)	1.3e-53	–
	M24A	AMPM	PFE1360c	AAG33975 (*Arabidopsis*, AMPM)	3e-61	PF00557 (Peptidase family M24)	2.5e-48	–
		AMPM	MAL8P1.140	P53582 (Human, AMP1)	1e-26	PF00557 (Peptidase family M24)	9.1e-06	+
		AMPM	PF10_0150	P53582 (Human, AMP1)	1e-104	PF00557 (Peptidase family M24)	5.4e-67	–
		AMPM	PF14_0327	AAL76285 (*P. falciparum*, AMPM2)	0	PF00557 (Peptidase family M24)	8.7e-50	–
	M24B	AMPP	PF14_0517	CAC59823 (Tomato, AMPP)	3e-85	PF00557 (Peptidase family M24)	2.8e-07	–
	M41	Ftsh peptidase	PF11_0203	NP_006787 (Human, AFG3)	1e-163	PF01434 (Peptidase family M41) PF00004 (AAA)	2.6e-80 4.0e-80	–
		Ftsh peptidase	PFL1925w	NP_422020 (*Caulobacter crescentus*, cell division protein FtsH)	1e-122	PF01434 (Peptidase family M41) PF00004 (AAA)	3.2e-93 5.7e-90	–

(continued)

Table 1. *Continued*

Catalytic class	Protease family	Protease nomenclature	Gene ID	Protease homolog with highest BLAST score: Accession (species, protease name)	E-score	Pfam domain structure: Domain ID (name)	E-score	AP[a]
		Ftsh peptidase	PF14_0616	NP_568787 (*Arabidopsis,* FtsH)	1e-141	PF01434 (Peptidase family M41) PF00004 (AAA	2.7e-69 9.9e-85	–
Serine	S1	DegP protease	*MAL8P1.126*	NP_568577 (Arabidopsis, DegP protease)	4e-52	PF00089 (Trypsin)	1.8e-07	–
		Neurotypsin-like	PF14_0067	BAA23986 (Mouse, neurotrypsin)	7e-17	PF01477 (PLAT/LH2 domain)	1.4e-09	–
	S8	**Subtilase-1**	*PFE0370c*	CAA05627 (*P. falciparum,* subtilase-1)	0	PF00082 (Subtilase family)	2.6e-15	–
		Subtilase-2	*PF11_0381*	CAB43592 (*P. falciparum,* Subtilase-2)	0	PF00082 (Subtilase family)	1.6e-35	–
		Subtilase-like	*PFE0355c*	CAA05627 (*P. falciparum,* subtilase-1)	2e-22	PF00082 (Subtilase family)	5.6e-19	–
	S9	ACPH	PFC0950c	P13676 (Rat, ACPH)	1e-17	PF00561 (α/β hydrolase fold)	0.021	–
	S14	Clp	*PFC0310c*	P54416 (*Synechocystis sp* Clp1)	6e-42	PF00574 (Clp protease)	1.8e-65	+
		Clp	*PF14_0348*	NP_567521 (*Arabidopsis* clp)	2e-29	PF00574 (Clp protease)	3.0e-37	+
		ClpB	PF08_0063	AAA88777 (*P. berghei,* ClpB)	0	PF00574 (Clp protease)	1.0-16	–
		ClpB	*PF14_0063*	NP_439019 (Haemophilus influenzae ClpB)	3e-95	PF00004 (AAA)	7.1e-06	+
		ClpC	PF11_0175	T07807 (Soybean clp)	1e-154	PF00574 (Clp protease)	0.0026	–
	S16	Lon	*PF14_0147*	AAA61616 (Human, Lon)	4e-53	PF00004 (AAA)	1.4e-17	–
	S26A	***SP1***	***PF13_0118***	T40251 (Fission yeast, IMP)	2e-10	PF00461 (Signal peptidase)	0.055	–
	S26B	signalase	MAL13P1.167	AAD19813 (*Drosophila,* signalase SPC21)	3e-46	PF00461 (Signal peptidase)	3.7e-19	–
	S54	Rhomboid	PFE0340c	NP_523536 (*Drosophila,* rhomboid-5)	6e-07	PF01694 (Rhomboid family)	5.5e-27	–
		Rhomboid	MAL8P1.16	NP_654179 (*Bacillus anthracis* rhomboid)	2e-07	PF01694 (Rhomboid family)	3.8e-27	–
Threonine	T1	Proteasome α1	PF14_0716	P92188 (*Trypanosoma cruzi,* α1)	8e-57	PF00227 (Proteasome)	7.7e-39	–
		Proteasome α2	MAL6P1.88	O9LSU2 (Rice, α2)	3e-73	PF00227 (Proteasome)	1.4e-49	–
		Proteasome α3	PFC0745c	O24362 (Spinach, α3)	3e-44	PF00227 (Proteasome)	2.0e-21	–
		Proteasome α4	PF13_0282	O81148 (*Arabidopsis,* α4)	1e-64	PF00227 (Proteasome)	7.4e-47	–
		Proteasome α5	PF07_0112	Q95083 (*Drosophila,* α5)	4e-70	PF00227 (Proteasome)	2.1e-47	–
		Proteasome α6	MAL8P1.128	Q9LSU3 (Rice, α6)	4e-52	PF00227 (Proteasome)	2.8e-30	–
		Proteasome α7	MAL13P1.270	O24616 (*Arabidopsis,* α7)	3e-66	PF00227 (Proteasome)	8.5e-43	–
		Proteasome β1	PFE0915c	P42742 (*Arabidopsis,* β1)	3e-44	PF00227 (Proteasome)	1.2e-39	–
		Proteasome β2	MAL8P1.142	Q9LST6 (Rice, β2)	1e-35	PF00227 (Proteasome)	2.9e-17	–
		Proteasome β3	PFA0400c	P25451 (Yeast, β3)	8e-35	PF00227 (Proteasome)	1.5e-18	–
		Proteasome β4	PF14_0676	XP_079788 (*Drosophila,* β4)	4e-36	PF00227 (Proteasome)	1.4e-22	–
		Proteasome β6	PFI1545c	O43063 (Fission yeast, β6)	4e-26	PF00227 (Proteasome)	1.1e-10	–
		Proteasome β7	PF13_0156	Q99436 (Human, β7)	1e-74	PF00227 (Proteasome)	5.5e-35	–

The cut-off criteria of E-score ≤ 1e-04 was employed to define protease homologs. The nomenclature of the protease family is: A1 (pepsin), C1 (papain), C2 (calpain), C12 (ubiquitin carboxyl-terminal hydrolase, family 1, UCH1), C13 (hemoglobinase), C14 (caspase), C19 (ubiquitin C-terminal hydrolase family 2, UCH2), C48 (Ubiquitin-like protease, Ulp), C56 (DJ-1 peptidase), M1 (alanyl aminopeptidase, AMPN), M3 (thimet oligopeptidase), M16 (pitrilysin and mitochondrial processing peptidase, MPP), M17 (leucyl aminopeptidase, AMPL), M18 (aspartyl aminopeptidase, DNPE), M22 (O-sialoglycoprotein endopeptidase, GCP), M24A (methionyl aminopeptidase, AMPM), M24B (X-Pro dipeptidase, AMPP), M41 (FtsH endopeptidase), S1 (trypsin), S8 (subtilsin), S9 (acylaminoacyl-peptidase, ACPH), S14 (clp), S16 (Lon protease, La), S26A (prokarytotic signal peptidase I, SP1), S26B (signalase), S54 (rhomboid), and T1 (threonine endopeptidase).
Abbreviations for proteases include: PM, plasmepsin; DPPI, dipeptidyl-peptidase I; UBP, ubiquitin-specific protease; IMP, mitochondrial inner membrane peptidase; SPC21, microsomal signal peptide 21 kDa subunit.
Previously characterized proteases with proteolytic activity are highlighted in bold. The 23 proteases predicted by the official annotation published in PlasmoDB are highlighted in italic.
Potential candidate proteases Calpain, Metacaspase, and Signal peptidase I (SP1) are highlighted in bold italic.
[a]± indicate the gene is predicted to contain/not contain an apicoplast transit peptide.

a large number of predicted and characterized proteases (human, 493; mouse, 431; *Drosophila melanogaster*, 529; *Caenorhabditis elegans*, 360; *Arabidopsis thaliana*, 568; Baker's yeast, 112; *Escherichia coli,* 127; *Bacillus subtilis*, 119). An average of 2.21% of the gene products belong to the protease superfamily in 77 completed genomes. Hence, given the observation that the number of predicted proteases appears to be positively correlated with organismal complexity, one might envisage that a considerable number of malaria proteases have yet to be identified in the ~23 Mbp *Plasmodium falciparum* genome that encodes for approximately 5300 gene products.

Here, we report a complete survey of protease homologs in the predicted and annotated *P. falciparum* genome (Gardner et al. 2002). Our initial comparative sequence search identified 92 putative malaria proteases, including potentially an interesting calpain, a metacaspase, and a signal peptidase I. Their expressions have been evaluated by microarray and RT-PCR assays. This study helps to develop an integrated view of a number of novel malarial proteases within an organismal, evolutionary, and functional context, and offers an intriguing opportunity to further target expressed and active proteases for chemotherapy.

RESULTS AND DISCUSSION

Ninety-Two Putative Proteases Are Predicted by Comparative Genomic Analysis

To gain further insight into the proteolytic machinery of the malaria parasite, the protein sequences in the annotated *P. falciparum* genome were subjected to an exhaustive search against the Merops protease database, which has a catalog and a structure-based classification of proteases. We adopted a relatively stringent threshold of E ≤ 1e-04 for BLASTP to ensure the high coverage with low false-positives. Redundant hits and partial sequences were excluded, resulting in a total of 92 protease homologs (Table 1). As highlighted in the Protease nomenclature column in Table 1, all twelve previously characterized proteases with proteolytic activity are included. In addition, as highlighted in the Gene ID column, 23 out of 25 proteases predicted by first-pass annotation published in PlasmoDB are included, among which subtilases 1 and 2 have been demonstrated to possess proteolytic activity; PFI0660c is not included because the E-score (0.39) of its closest homolog (*Bacillus anthracis* CAAX amino terminal protease, accession number NP_655263) is far below the cut-off 1e-04; PF11_0314 is not included because it is more likely to possess ATP hydrolytic and regulatory function than protoelytic function based on sequence homology.

The domain/motif organization of predicted proteases was revealed by the InterPro Search. For each putative malaria protease, the known protease sequence or protease domain of the highest similarity was used as a reference for annotation; the catalytic type and protease family were predicted in accordance with the classification in the protease database Merops (http://www.merops.co.uk/merops/merops.htm), and the enzyme was named in accordance with the SWISS-PROT enzyme nomenclature (http://www.expasy.ch/cgi-bin/lists?peptidas.txt) and literatures.

New Catalytic Types and Families

Proteases are classified into five major clans (Aspartic, Cysteine, Metallo, Serine, and Threonine) based on their catalytic mechanisms. They can be further grouped into distinct families and subfamilies by intrinsic evolutionary relationships (Rawlings and Barrett 1993). Using the comparative database search, we detected a total of 59 new protease homologs, in addition to 12 characterized proteases with proteolytic activity and 21 predicted by official annotation (Table 1). Moreover, a spectrum of conserved core characteristic domains/motifs for specific catalytic classes has been detected in most of the predicted proteases, indicating their potential activity.

The 92 putative proteases belong to 26 families of five clans, compared to the previously reported 12 proteases that belong to six families of four clans (Rosenthal 2002). The distribution (11% aspartic, 36% cysteine, 22% metallo, 17% serine, and 14% threonine) resembles those in other model organisms, supporting the fundamental premise that a prototype protease system is conserved throughout evolution (Rawlings and Barrett 1993; Southan 2001). Our speculation that a large number of potential proteases remain unexplored in the *P. falciparum* genome appears justified. Undoubtedly, some of the uncharacterized proteases will perform crucial functions in the parasite life cycle as discussed below.

Examples of Potentially Important Proteases

Calpain

Calpain is a group of intracellular cysteine proteases that mediate a wide variety of physiological and pathophysiological processes, including signal transduction, cell motility, apoptosis, and cell cycle regulation (Sorimachi et al. 1997; Glading et al. 2002). In *P. falciparum*, a calpain, yet unidentified, was believed to be essential in merozoite invasion, based on the observation that Calpain inhibitors I and II strongly blocked invasion (Olaya and Wasserman 1991).

We have identified a putative calpain (MAL13P1.310) in the *P. falciparum* genome, which exhibits high sequence similarity to *C. elegans* calpain-7 (E=2e-35). Moreover, its ortholog (accession no. EAA19663) has been identified in the newly released genome of the model rodent malaria parasite *Plasmodium yoelii yoelii*. It possesses a catalytic domain (985–1453) detected by the Hidden Markov Model in the pfam search, with E = 8.0e-13 (Fig. 1). The most intriguing aspect of this domain is the presence of three active sites (Cys1035, His1371, and Asn1391) that constitute a cleft crucial for catalytic activity (Arthur et al. 1995). A multiple alignment of the catalytic regions was produced for the putative plasmodial calpain and the representative human calpains. In addition to the invariable Cys-His-Asn triad, a high degree of identity is also observed in its vicinity, reflecting stringent functional and mechanistic conservation (Fig. 1). Indeed, the experimental demonstration that a single catalytic subunit in rat and chicken calpains possesses a full bona fide proteolytic activity (Yoshizawa et al. 1995) reinforces the potential processing capacity of the putative plasmodial calpain.

Our further phylogenetic analysis of the putative *P. falciparum* calpain revealed its striking origin, which might have attributed to an alternative Ca^{2+}-independent regulatory mechanism. Figure 2 shows the evolutionary tree inferred by the neighbor-joining (NJ) method using Poisson corrected distance (Saitou and Nei 1987). Evolutionary trees based on Parsimony (PAUP4.0) and Maximum Likelihood (PHYLIP) also yielded topologies and clade structures congruent with NJ (data not shown). Apparently, two putative plasmodial calpains belong to a novel monophyletic group of animal calpain-7 proteases, with 61% bootstrap support. They share the common domain architecture in the calpain-7 clade: lacking any significant similarity to the C-terminal EF-hand Ca^{2+}-

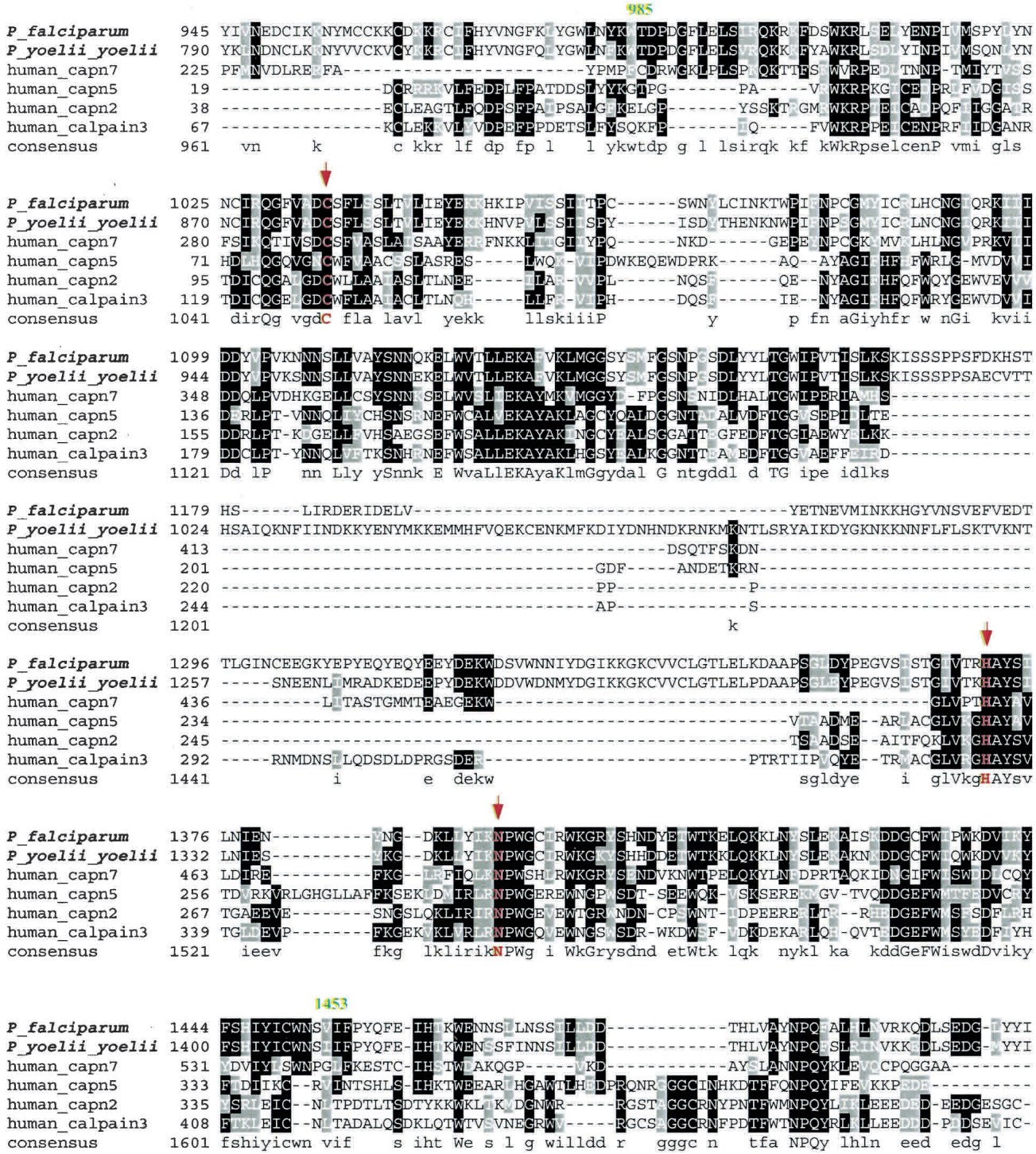

Figure 1 Multiple alignment of the catalytic domains of the putative *P. falciparum* calpain (MAL13P1.310) and the representative human calpains using T-coffee program followed by manual correction. The catalytic domain region is predicted to be from amino acid residue 985 to 1453 by pfam HMM algorithm. The three conserved amino acids, C(1035), H(1371), N(1391), that are part of the active sites are highlighted with arrowheads. Graphic presentation of the alignment and the consensus sequence were obtained by the program BOXSHADE 3.21. Conserved residues are shaded with black and gray. The accession numbers of calpain protein sequences used for alignment refer to Figure 2.

binding domain present in most of the essential Ca^{2+}-dependent mammalian calpain subtypes (calpains -1, -2, -3, -9, -11, and Mu/M-type) (Franz et al. 1999). Provided that fungi cysteine protease PalB, the nearest neighbor of calpain-7, contains a PBH domain resembling the Ca^{2+}-binding domain (Denison et al. 1995), one could speculate that the loss

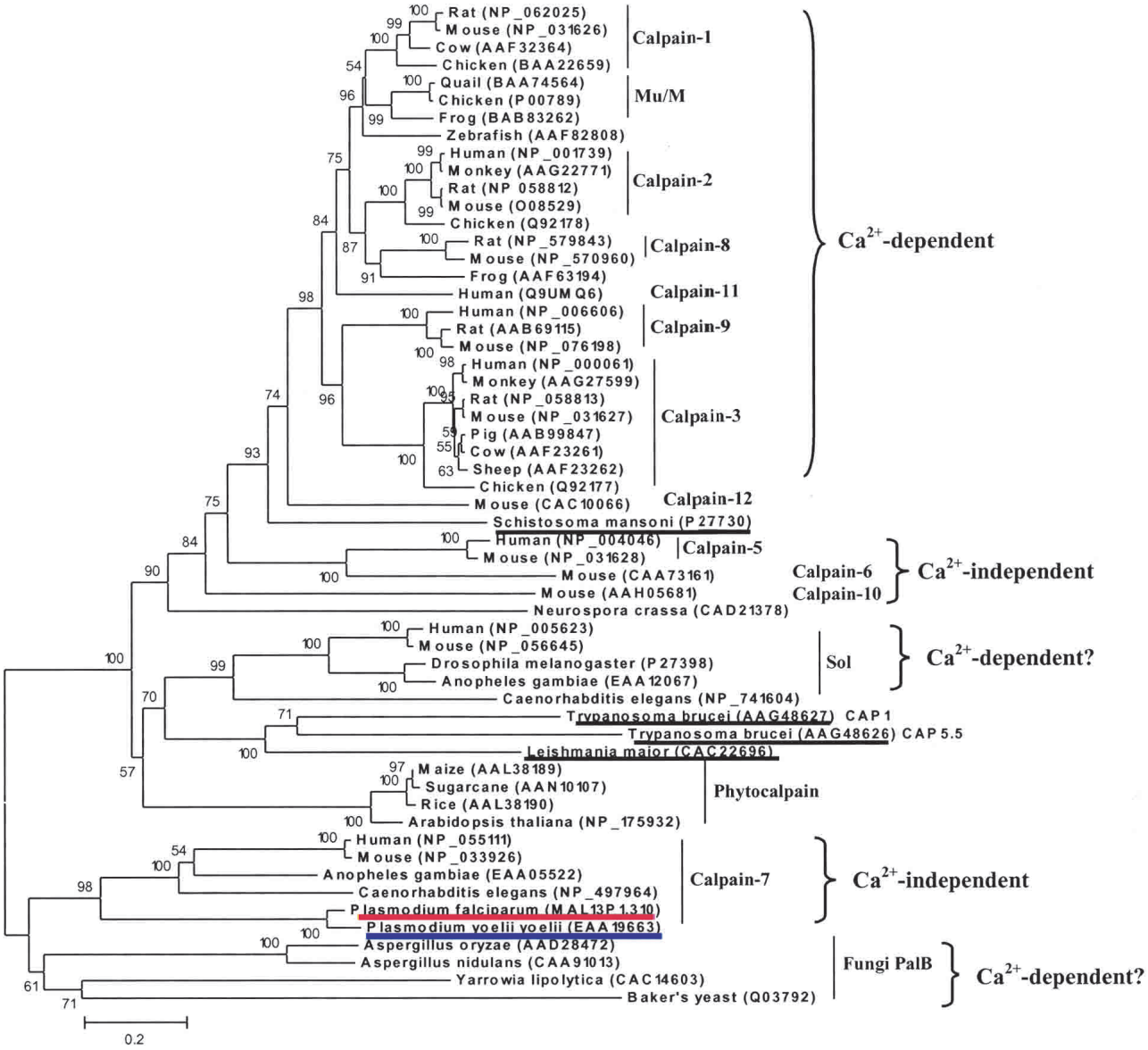

Figure 2 The phylogenetic tree of the calpains, inferred by the neighbor-joining method based on the amino acid sequences with Poisson corrected distance. The option of complete deletion of gaps was used for tree construction. 1000 bootstrap replicates were used to infer the reliability of branching points. Bootstrap values of >50% are presented. The scale bar indicates the number of amino acid substitutions per site. The parasite sequences are underlined. The putative *P. falciparum* calpain and *P. yoelii yoelii* ortholog are highlighted in red and blue, respectively. The accession number for each sequence is included in the parentheses after the species name.

of Ca^{2+} dependency in calpain-7 subtype had been derived from evolutionary events such as domain shuffling, which might be associated with the divergence of mRNA splicing sites (Craik et al. 1983). Such events appear to have occurred close to or prior to the origin of the animal kingdom (Fig. 2).

The identification of plasmodial calpain has also implicated the existence of calpain-mediated pathways. Its potential cognate targets include host cytoskeletal proteins such as spectrin, integrin, and ezrin. Moreover, the recent discovery of a typical endogenous substrate of calpain, Protein Kinase C (PFL1110c; PFI1685w) in *P. falciparum*, has provided the support of a parasite-controlled signaling cascade (Doerig et al. 2002).

It is conceivable that the putative protease-active and Ca^{2+}-independent plasmodial calpain may serve as a good antimalarial target for two reasons. First, it may be the central component of crucial signal transduction pathways that affect parasite biology and host-parasite interactions. Second, because it is evolutionarily divergent from the essential subtypes of host calpains, its specific inhibitor may have minimal effects on the host.

Metacaspase

Metacaspase (PF13_0289) is another interesting hypothetical protease. In vertebrates, a cascade of caspases (cysteine aspartyl proteases) is the major modulator of apoptosis (programmed cell death) (Thornberry and Lazbnik 1998; Aravind et al. 1999). Two families of ancient caspase-like proteins

(paracaspases and metacaspases) have been found in metazoans, fungi, and protozoa. As shown in the phylogenetic tree (Fig. 3), the putative plasmodial metacaspase occupies a distinct clade constituting paracaspases and metacaspases, which are likely to be the primordial form of 14 subfamilies of vertebrate caspases (bootstrap value = 100%). Interestingly, human paracaspase is capable of interacting with the oncogene Bcl10 and triggering NF-kB activation, indicative of the prone-to-apoptosis property of the ancestral caspase (Uren et al. 2000). Moreover, yeast metacaspase has been demonstrated as an effective executor for apoptosis, suggesting the root of apoptosis dates back to unicellular organisms (Madeo et al. 2002). The multiple alignment clearly reveals that the putative plasmodial metacaspase retains the typical caspase fold, which is centered with the His (404)-Cys (460) catalytic dyad conserved in all representative proteolytically active caspases (Fig. 4). Conversely, considerable sequence diversity is observed in the vicinity of this active site cleft. In particular, yeast metacaspase and the plasmodial homolog exhibit distinct sequence profile to other vertebrate caspases and human paracaspase. Previously, Uren et al. (2000) have postulated that ancient (paracaspases and metacaspases) and vertebrate subtypes differ in substrate-specificity. We have demonstrated that the experimentally confirmed differential substrate-specificity in major vertebrate subtypes is largely determined by the chemical property and configuration of residues situated in the caspase fold (Wang and Gu 2001). Thus, the observed distinct configuration of residues in the active site proximity could account for parasite-specific substrate-preference.

In *Plasmodium*, the physiological process of apoptosis has never been reported, nor the critical components identified. Nevertheless, the detection of the metacaspase homolog will allow us to investigate the role, if any, of apoptosis and/or analogous signal transduction pathway in the parasite. In addition, since metacaspases have only been found in protozo-

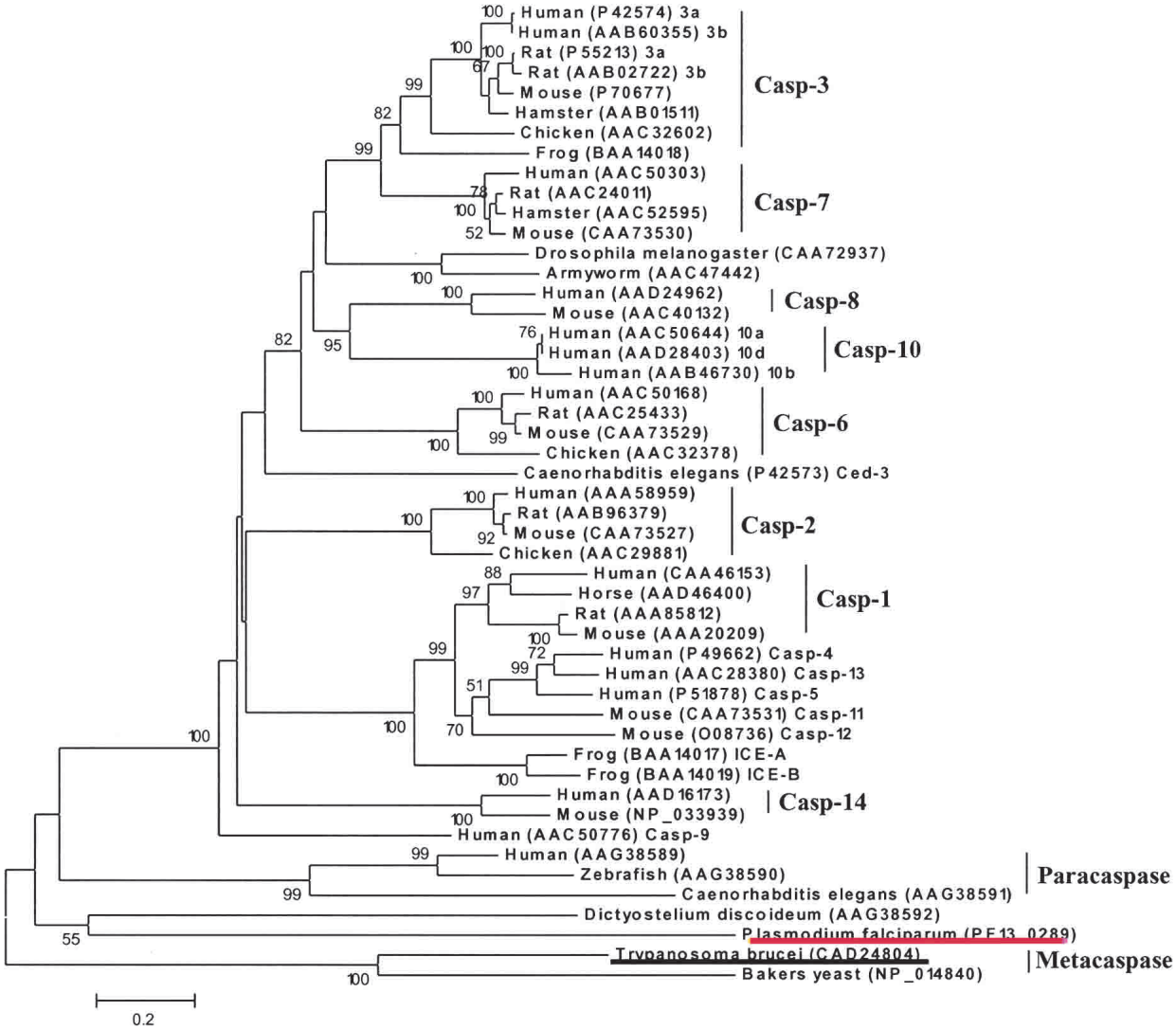

Figure 3 The phylogenetic tree of the caspases, inferred by the neighbor-joining method based on the amino acid sequences with Poisson corrected distance. The option of complete deletion of gaps was used for tree construction. 1000 bootstrap replicates were used to infer the reliability of branching points. Bootstrap values of >50% are presented. The scale bar indicates the number of amino acid substitutions per site. The protozoan sequences are underlined.

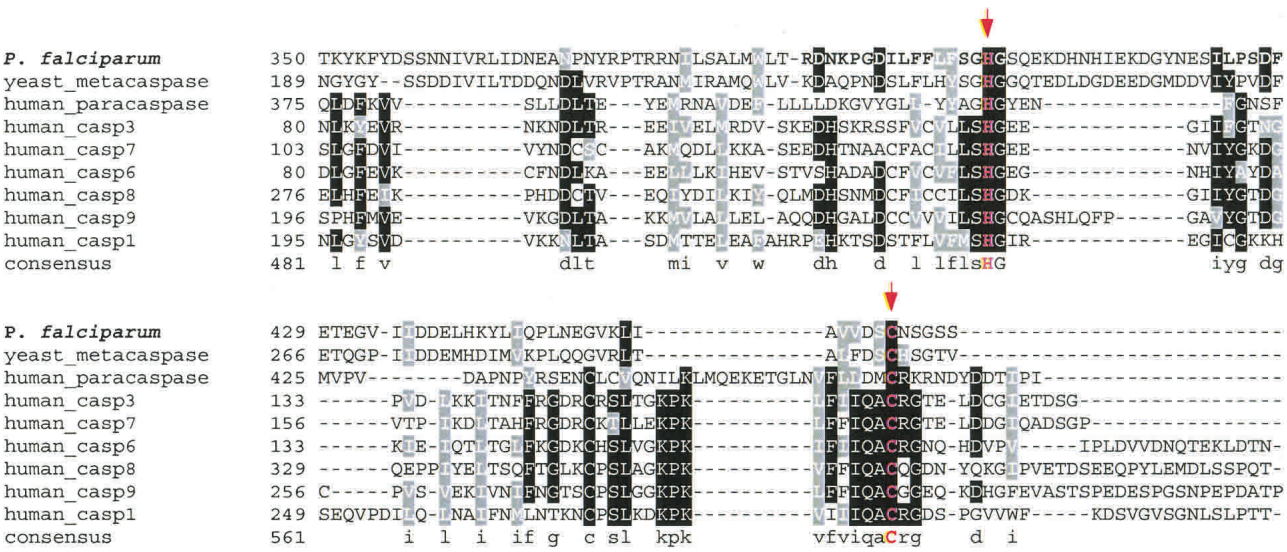

Figure 4 Multiple alignment of the catalytic regions of the putative *P. falciparum* metacaspase (PF13_0289) and the representative proteolytically active caspases using Clustal X1.8 followed by manual correction. The catalytic dyad H (404) and C (460), which are part of the active sites, is highlighted with arrowheads. Graphic presentation of the alignment and the consensus sequence were obtained by the program BOXSHADE 3.21. Conserved residues are shaded with black and gray. The accession numbers of caspase protein sequences used for alignment refer to Figure 3.

ans, yeasts, and possibly in plants, and are phylogenetically distinct from other caspase subtypes (Fig. 3), the putative plasmodial metacaspase may serve as a potential chemotherapeutic target.

Signal Peptidase 1 (SP1)

Signal peptidases (SP) play indispensable roles in protein trafficking and sorting by removing signal peptides from precursors of secretary proteins. This serine protease family consists of two subtypes, SP1 and signalase, based on their distinct structural, functional, and evolutionary features. To date, SPs have been found in bacteria, archaea, fungi, plants, and animals; however, SP has never been reported previously in protists, despite the fact that the dynamic parasite life cycle reflects a need of specific peptidase(s) to process proteins that are translocated across host and parasite membranes. Using the comparative genomic search, we first identified two homologs of signal peptidase, PF13_0118 (SP1) and MAL13P1.167 (signalase) in *P. falciparum*.

Between two subtypes, SP1 has generated extensive research interest because it represents a novel antibiotic target for its distinct prokaryotic origin and essential functions (Paetzel et al. 2000). We have also identified an ortholog of *P. falciparum* SP1 in the rodent parasite *P. yoelii yoelii* genome. Our evolutionary analysis revealed that the two putative plasmodial SP1 have three clusters of homologs: (1) Bacteria SP1; (2) an *Arabidopsis* chloroplast thylakoidal processing peptidase; and (3) mitochondrial inner membrane peptidases (Imp) found in eukaryotes, which appear to be the nearest neighbor to plasmodial SP1 (Fig. 5). Given the proposed prokaryotic origin of the chloroplast and mitochondrion, malarial SP1 is likely to have evolved via the prokaryotic-specific lineage. Moreover, the potential of its catalytic activity can be inferred from the comparative sequence analysis. The putative SP1 contains the catalytic dyad (Ser175, Lys274) that is invariable across representative SP1 proteins with confirmed signal peptidase activity (Fig. 6). Most notably, this Ser/Lys catalytic dyad mechanism is unique in SP1, compared with the typical Ser/His/Asp triad system in other serine proteases. It seems plausible that the putative plasmodial SP1 has a fundamental role yet to be determined, and represents a promising target given its distant relatedness to the host.

Important Protease–Mediated Pathways Implicated in P. falciparum

Our findings suggested at least five new protease-mediated activities: (1) an ATP-dependent ubiquitin-proteasome-mediated cell-cycle control and stress-response system (Verma et al. 2002). Although the mechanism by which proteasomes function in *P. falciparum* is poorly understood, their importance was suggested by the observed irreversible inhibition on the growth and development of the hepatic and erythrocytic stages of three different *Plasmodium* species by Lactacystin, a specific threonine protease inhibitor (Gantt et al. 1998). The identification of the clade of threonine proteases α and β, and a series of ubiquitinyl hydrolases (UCH1 and UCH2) brings new insight into this universally conserved proteasome machinery (Table 1). (2) A lysosomal proteolysis. This selective pathway to degrade cytosolic proteins may involve a number of cathepsins with versatile functions, which are assisted by cytosolic and lysosomal molecular chaperones and receptor proteins in the lysosomal membrane. (3) A calpain-activated signal transduction cascade, which may work in conjunction with upstream modulator and downstream effectors of host or parasite origin. (4) A caspase-mediated apoptosis or apoptosis-like signal transduction pathway. Although yeast metacaspase has been confirmed to induce apoptosis, the classical apoptosis regulators appear to be missing in the yeast genome. Thus, it is desirable yet challenging to identify the key components in this pathway, which may be conserved across organisms, or be parasite-specific. (5) A signal peptidase-initiated precursor protein processing pathway.

Evolutionary Implications

Studying the origin and the evolutionary mechanisms behind plasmodial proteases will contribute to the selection of target

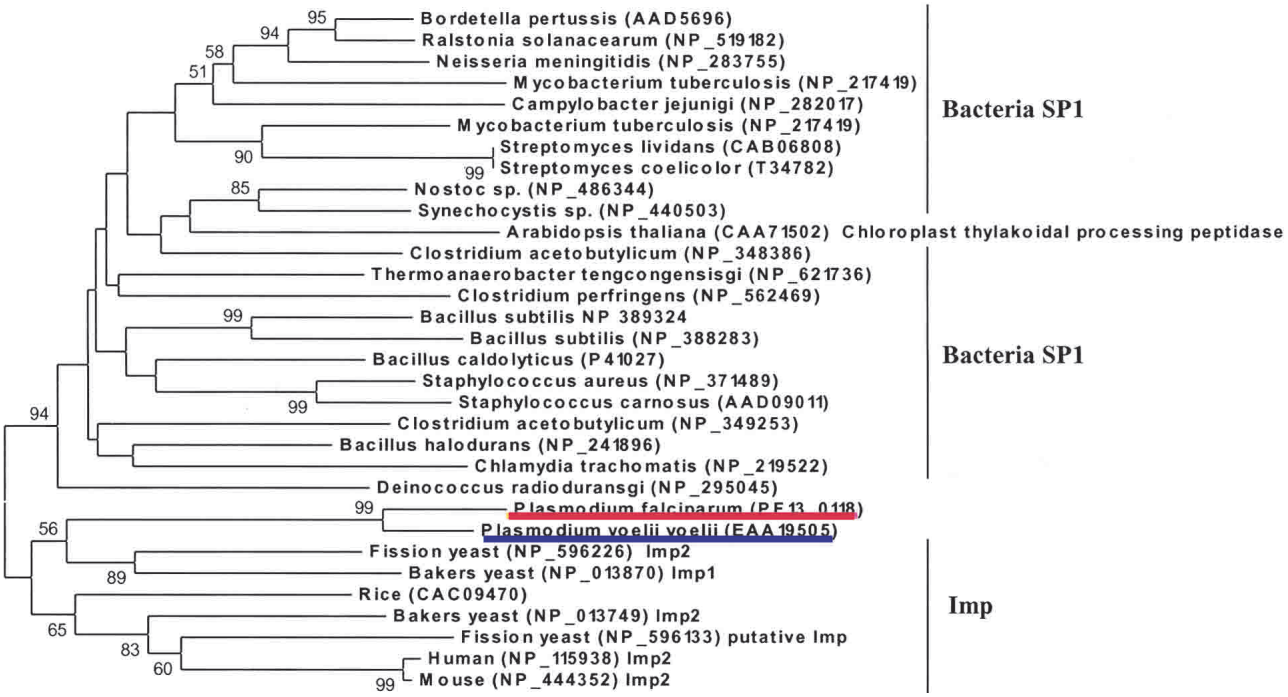

Figure 5 The phylogenetic tree of the Signal peptidases I (SP1), inferred by the neighbor-joining method based on the amino acid sequences with Poisson corrected distance. The option of complete deletion of gaps was used for tree construction. 1000 bootstrap replicates were used to infer the reliability of branching points. Bootstrap values of >50% are presented. The scale bar indicates the number of amino acid substitutions per site. The putative *P. falciparum* calpain and *P. yoelii yoelii* ortholog are highlighted in red and blue, respectively. The SP1 homolog in *Arabidopsis* is termed Chloroplast thylakoidal processing peptidase. Imp is the abbreviation for mitochondrial inner membrane peptidase.

proteases to be studied in detail, for which specific inhibitors with no or minimal effect on the host can be designed. A complex evolutionary scenario including gene duplication, domain shuffling, and lateral gene transfer has been implicated in the preliminary analysis of the predicted proteolytic machinery in *P. falciparum*. Gene duplication is believed to play important roles in the evolution of multigene families by providing raw material for the novel functionality under differential evolutionary constraints (Ohno 1970; Li 1983; Friedman and Hughes 2001; Gu et al. 2002; McLysaght et al. 2002). In *P. falciparum*, well-characterized falcipains (-1, -2, -3), plesmepsins (I-IV), and subtilases (-1, -2) exemplify the multigene families that arise from gene duplications (Coombs 2001; Rosenthal 2002). We have identified a series of putative proteases that may comprise multigene families (Table 1). Some reflect tandem gene duplications in adjacent chromosome loci. For example, eight SERA homologs aggregate as a cluster in chromosome 2 contig 11953 (Miller et al. 2002). In contrast, some potential paralogs are located in remote chromosome regions. For instance, the UCH2 family with the consensus domain is sparsely distributed over seven chromosomes. This suggests that ancient gene duplications and subsequent functional divergence may result in an extensive repertoire of the present multigene families. In addition to gene duplication, domain shuffling coupled with the splice-site variation, intron loss, and horizontal gene transfer are proposed to be important modes in the evolution of aspartic proteases in the parasite genus *Apicomplexa*, including *P. falciparum* (Jean et al. 2001). The proteases encoded by or destined to parasite organelles are of particular interest because organelles represent microenvironments in which proteases may evolve at different rates and thus achieve novel functions (Fast et al. 2001). The first target organelle is the apicoplast, the apicomplexan-specific plastid derived by secondary reduction of a red alga endosymbiont. Since the plastid-encoded gene is of prokaryotic origin, its inhibitor may have only a minor, if any, effect on the vertebrate host and therefore may represent a promising antimalarial target. Our preliminary analysis shows that the putative clpC gene "PF11_0175" matches one apicoplast-encoded gene (Wilson et al. 1996). Moreover, 14 predicted proteases may contain an apicoplast transit peptide, among 511 genes identified in the entire parasite genome by pattern-recognition program PATS (Predict Apicoplast-Targeted Sequences) (Zuegge et al. 2001). From the population genetics perspective, we would anticipate detecting a certain level of polymorphism among putative proteases, due to the ancient origin of *P. falciparum* as revealed by chromosome-wide SNP analysis (Verra and Hughes 1999; Mu et al. 2002; Wootton et al. 2002). However, the alternative Malaria's eve hypothesis of a severe recent population bottleneck may still be valid (Rich et al. 1998; Volkman et al. 2001). More detailed analysis of the genomics and proteomics of plasmodial proteases will help resolve these fundamental questions about *P. falciparum* evolution.

Eighty-Three Putative Proteases Are Actively Transcribed in the Intraerythrocytic Stage, and Sixty-Seven Are Actively Translated in the Life Cycle

We are bearing in mind that genome analysis based solely on sequence similarity clearly predicts many unknown putative

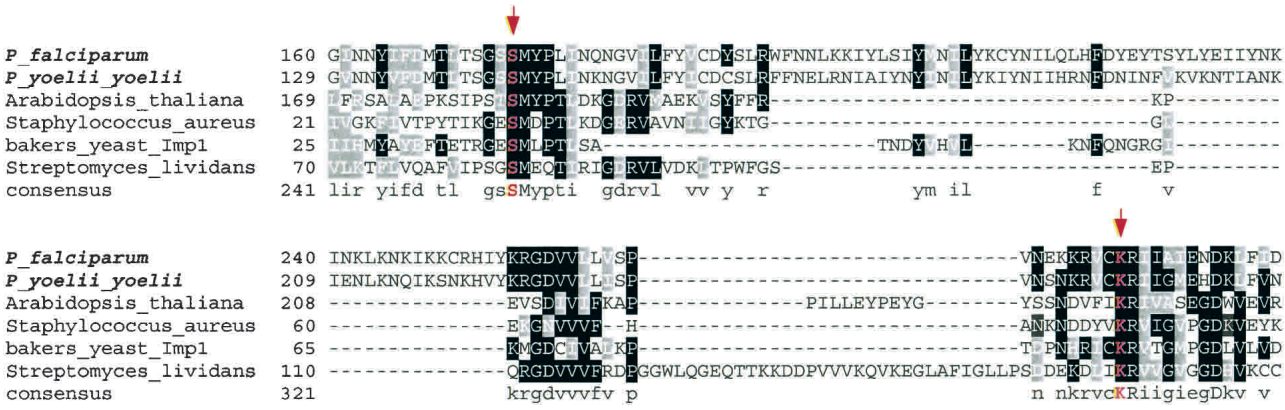

Figure 6 Multiple alignment of the catalytic regions of the putative *P. falciparum* SP1 (PF13_0118) and the representative proteolytically active signal peptidase I using T-coffee program followed by manual correction. The catalytic dyad Ser (S175) and Lys (K274), which are part of the active sites, is highlighted with arrowheads. Graphic presentation of the alignment and the consensus sequence were obtained by the program BOXSHADE 3.21. Conserved residues are shaded with black and gray. The accession numbers of SP1 protein sequences used for alignment refer to Figure 5.

malaria proteases, however, these are only predictions. Which of the 59 newly predicted proteases, in addition to the 12 characterized proteases and 21 proteases annotated previously, are true protein-encoding genes expressed in the parasite life cycle? This important question was first addressed by analyzing an en masse gene expression profile using microarray chips, and then followed by RT-PCR confirmation.

Microarray

We focused on the parasite expression profiles of the asexual erythrocyte stage not only because this stage is responsible for malaria clinical manifestations, but also because of the accessibility of the research materials. In order to obtain all genes transcribed throughout the erythrocyte stage of the parasite, we extracted and pooled mRNAs from *P. falicarpum* 3D7 culture samples collected at four 12-h intervals. Figure 7 shows the temporal development of parasites that includes rings, trophozoites, schizonts, and merozoites, indicating that an asynchrony was successfully achieved. Probes were labeled with fluorescent dyes using mRNAs purified from the asynchronous culture as a template, and then hybridized to the microchip arrayed with 6239 Malaria Genome Array Oligomers (Operon Technologies).

Results, summarized in Table 2, clearly demonstrated that 75 predicted proteases have signal intensities higher than those of negative controls. Being aware that the cut-off value for signal intensity is controversial, and that using the average intensity of the negative controls may be somewhat arbitrary, we selected the gametocyte-specific proteins (CS protein TRAP-related protein, Pfs25, Pfs48/45, Pfg377, and a gametocyte-specific var) and three large gene families in which the majorities are silent due to clonal expression switching (var, rifin, and stevor) as internal references (Hayward et al. 2000; Ben Mamoun et al. 2001). As anticipated, all gametocyte-specific genes, 39 of 45 var genes, 99 of 118 rifin genes, and 12 of 14 stevor genes displayed signal intensities below the level of the negative controls (data not shown). These data further support our conclusion that 75 predicted proteases are actively transcribed during the erythrocytic stage. Interestingly, the putative multigene families such as SERA and UCH2 exhibit variable expression levels across paralogous members, reflecting a certain level of functional divergence after gene duplication events.

We also analyzed two microarray datasets published in the PlasmoDB. The first dataset includes the expression profile of two erythrocyte stages (Trophozoites and Schizonts) using the Oligo Microarray (Hayward et al. 2000). The result of 66 predicted proteases transcribed in at least one stage supported our finding that the majority of the predicted proteases were actively transcribed during the erythrocyte stage (Table 2). The second dataset represents the first proof-of-concept experiment of using cDNA microarray to explore the expression profile of five erythrocytic forms and stages (Ben Mamoun et al. 2001). Among 944 elements or gene fragments (317 genes of identifiable homology) included in the probe design, eight corresponded to predicted proteases. The positive signals of seven genes are consistent with our result from asynchronous culture. The stage-specific profile also confirmed the ubiquitous expression of the putative proteosome β6 (PFI1545c), which does not have corresponding 70-mer in the Oligo Microarray.

Reverse Transcription Polymerase Chain Reaction

Among the 17 remaining predicted proteases that are not detected using microarray hybridization, seven showed signal

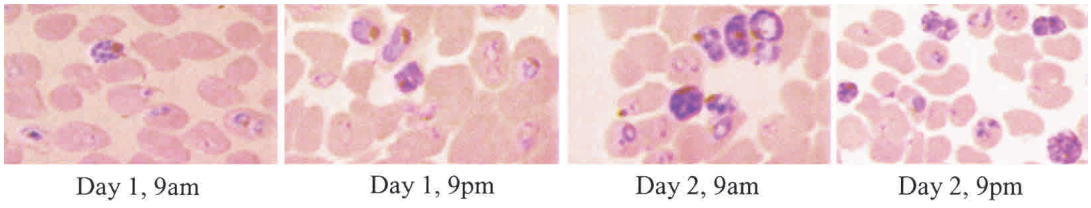

Figure 7 Four *P. falciparum* 3D7 culture samples were collected at 12-h intervals, and pooled to achieve a total asynchrony.

Table 2. Expression Profiles of Putative Plasmodial Proteases

		Oligo Microarray[b]		Proteomics[c]			
Gene ID	Oligo Microarray (asyn)[a]	T	S	T	M	G	S
PF14_0076	21,915	7595	6935	+	+	+	
PF14_0077	6544	6084	3034	+	+	+	
PF14_0078	2981	16,185	7152	+	+	+	
PF14_0075	20,151	20,979	21,090	+	+	+	
PF13_0133	1521	5636	4810				
*PFC0495w**	454	–	–				+
PF10_0329	76	–	–				
PF14_0623	215	–	–				
PF14_0281	**No oligo**						
PF08_0108	5561	2478	–				+
PF14_0553	3187	4947	4954				+
PF11_0165	8073	17,895	4771	+			
PF11_0162	9302	4559	5676	+		+	
PF11_0161	2436	7318	2025	+			
*PFL2290w**	467	–	–	+		+	
PFD0230c	8342	2473	25,913		+		
PF11_0174	19,855	36,201	19,570	+	+		
PFB0360c	376	–	–				
*PFB0325c**	179	–	–				+
PFB0330c	2415	2238	5476	+	+		
PFB0335c	3428	1695	4803				
PFB0340c	28,613	13,254	59,511	+	+		
PFB0345c	2273	3115	10,053				+
PFB0350c	4572	–	6320	+		+	
PFB0355c	1401	–	–				
PFI0135c	3655	2747	3516				+
MAL13P1.310	676	1635	53,119				
PF14_0577	1008	–	2060	+			
PF11_0177	781	2058	3604				+
PF11_0298	700	–	2514			+	
PF13_0289	643	–	1550				
PFA0220w	3831	5734	10,100				+
PFD0165w	2208	–	–				+
PFD0680c	894	1696	2387		+		
PFE1355c	2840	2251	2414	+	+	+	
PFE0835w	2422	–	1831		+		
MAL7P1.147	6182	3575	3628	+	+	+	+
PFI0225w	3780	1998	2117				+
PF13_0096	2105	3554	3789				
PF14_0145	1826	2428	–				+
PFL1635w	**No oligo**				+	+	
MAL8P1.157	2311	2226	1946				+
MAL6P1.153	**No oligo**			+	+		+
MAL13P1.56	2788	3250	1588	+	+	+	+
PF10_0058	1765	3672	4309				
MAL13P1.184	324	–	–				
PFE1155c	3316	3099	2134	+	+	+	
PFI1625c	2018	4583	4239		+	+	
PF11_0189	4006	4829	3862	+			+
PF11_0226	1002	–	–	+		+	+
PF13_0322	4279	8553	8017	+	+	+	+
PF14_0382	790	–	–				
PF14_0439	4828	10,517	4692	+	+		+
PFI1570c	2424	2650	–	+	+	+	+
PF10_0299	1326	2972	1991		+		+
PFE1360c	857	1919	1579	+			
MAL8P1.140	3379	2033	–				+
PF10_0150	2880	2103	–	+	+	+	
PF14_0327	4130	–	–	+		+	+
PF14_0517	7889	15,722	6630	+	+	+	+
PF11_0203	2440	1695	–				
PFL1925w	**No oligo**			+			
PF14_0616	5347	1855	–				
MAL8P1.126	1030	–	–				
PF14_0067	735	–	–			+	
PFE0370c	3384	–	5371		+		

(continued)

intensity below the negative controls. One possibility is that some of them are expressed in stages other than the asexual erythrocytic ones. This could be further investigated by using RNAs extracted from the intraerythrocytic and extraerythrocytic stages. The remaining ten predicted proteases were not included in the oligomer set printed on the array slides because only ~90% of the *P. falciparum* genome data was available when the oligomers were designed. To examine whether these predicted proteases were also expressed in the erythrocyte stage, we designed specific primers and performed RT-PCR using the RNA extracted from the asynchronous culture (Fig. 7) as templates. Data shown in Figure 8A clearly suggests that all ten predicted genes were actively transcribed.

As mentioned above, *P. falciparum* calpain, metacaspase, and signal peptidase1 are of particular interest due to the potential biological roles they may play. The microarray analysis suggested that the predicted genes for these proteases were actively transcribed (Table 2). We also performed RT-PCR to further confirm their expression (Fig. 8B).

The microarray and RT-PCR data only indicated the active transcription of 85 predicted proteases. In order to examine expression at the level of translation, we analyzed the proteomics data published in PlasmoDB (Florens et al. 2002). It appeared that 67 out of 92 predicted proteases are translated at some point during the life cycle (Table 2). Some proteases are ubiquitous, whereas others show stage-specific expression. It was notable that the three predicted proteases that did not have detectable transcription from the microarray assay did show positive translation in specific stages including intraerythrocytic stages.

Combining the complementary results from microarray, RT-PCR, and proteomics analysis, we found that of the 92 putative proteases identified by scanning the genome, 88 were transcribed and 67 were translated at some stage in the life cycle. The remaining four may be expressed at extraerythrocytic stages or may be pseudogenes, a result due to the frameshift in the open reading frame (Triglia et al. 2001).

Table 2. *Continued*

Gene ID	Oligo Microarray (asyn)[a]	Oligo Microarray[b] T	Oligo Microarray[b] S	Proteomics[c] T	Proteomics[c] M	Proteomics[c] G	Proteomics[c] S
PF11_0381	1071	2229	10,369			+	
PFE0355c	840	–	4153				
PFC0950c	**No oligo**						+
PFC0310c	3426	2308	3281				
PF14_0348	640	2518	2312				
PF08_0063	3901	3233	1715	+	+	+	+
PF14_0063	4125	2130	3154				
PF11_0175	8264	2856	4534	+	+	+	+
PF14_0147	1023	–	–				
PF13_0118	3050	2405	1676				
MAL13P1.167	2758	2514	4077				
PFE0340c	6507	–	10,279				
MAL8P1.16	**No oligo**						
PF14_0716	8791	9053	6476	+	+	+	
MAL6P1.88	**No oligo**				+		
PFC0745c	5944	9387	7742	+	+		
PF13_0282	7652	7248	5541		+	+	
PF07_0112	4279	8258	5730	+	+	+	+
MAL8P1.128	**No oligo**			+	+	+	
MAL13P1.270	1487	6809	5159	+	+	+	+
PFE0915c	3396	5939	5761	+	+		
MAL8P1.142	5571	5814	5579	+	+	+	
PFA0400c	**No oligo**			+	+	+	+
PF14_0676	6984	–	–		+		
PFI1545c	**No oligo**				+		
PF13_0156	5240	9539	13,684	+	+	+	

[a]Expression profile of asynchronous culture using Oligo Microarrays. The microarray slide was printed with 6239 70-mers mapped to 4407 predicted open reading frames. Probes were labeled with fluorescent dyes using mRNAs purified from an asynchronous culture as a template (http://derisilab.ucsf.edu/). Briefly, messenger RNAs were purified using oligo T cellulose and reverse transcription was conducted to incorporate aminoallyl dUTP into the cDNAs. The Cy3 and Cy5 NHS esters were then coupled to amine groups of the cDNA, and dye-labeled probes were hybridized with the microarray slides under standard condition (3X SSC, 50% formamide, 0.1% SDS, 10 mg/ml salmon sperm DNA, 68°C). The slide was scanned with a GenePix 4000B (Axon Instrument) at default PMT settings, 100% power. The array data were analyzed initially with GenePixPro software (Axon Instrument), then with global normalization. The expression level is indicated by the mean signal intensity of all corresponding oligomers in triplicates on the microarray slides (MRA-452) obtained from Malaria Research and Reference Resource Center (http://www.malaria.mr4.org/). Ten predicted proteases without corresponding oligomers are highlighted in bold. Two sets of negative controls were included in the DeRisi design: (1) 20 oligomers from yeast intergenic region with the mean intensity 529; (2) 33 *P. falciparum* genes cloned in the plasmid, including 16 ribosomal proteins, 17 tRNA genes, LSU, Clp, and tufA. Their mean intensity was 598. The percentiles of expression level over all the spots are 297 (30%), 394 (35%), 512 (40%), 646 (45%), and 795 (50%), respectively. The genes that showed signal intensity below the mean of negative controls are highlighted in *italic.* An asterisk (*) indicates the gene was reported to be expressed in the parasitic life cycle from proteomics data (Florens et al. 2002).

[b]Expression profile in the erythrocytic stage using cDNA microarray chip containing 944 elements (317 genes of identiable homology; Ben Mamoun et al. 2001). The average intensities were extracted from PlasmoDB. R, ring. A minus sign (−) indicates signal not detected or below the cut-off (35% percentile over all the spots on the array).

[c]Expression profile in the parasite life cycle using MUDPIT proteomics technology (Florens et al. 2002), extracted from PlasmoDB. A plus sign (+) indicates at least one peptide of the protein was detected by Mass Spectrum. T, trophozoites; M, merozoites; G, gametocytes; S, sporozoites.

Conclusions

The exhaustive homology search and comparative sequence analysis have resulted in the delineation of 92 putative proteases, including 59 that had not been previously recognized in the *P. falciparum* genome. This set includes potentially important proteases such as calpain, metacaspase, and signal peptidase, and indicates protease-mediated activity that may be vital for parasite life cycle. Furthermore, 88 are demonstrated to be actively transcribed proteins by the microarray, RT-PCR data, and proteomics. This study is an initial attempt at the systematic identification of novel malaria proteases that have essential functions and assessment of their evolutionary relationship to the vertebrate host. By combining in silico genomics-based predictions with experimental confirmation, there is an increased likelihood of identifying new therapeutic targets.

METHODS

Genome Sequences, Homology Search, and Comparative Sequence Analysis

A total of 5865 nonredundant query sequences of characterized and predicted proteases from 1066 organisms were obtained from the Merops database (http://www.merops.ac.uk, release 5.8 of March 19, 2002), which has a catalog and a structure-based classification of proteases. The BLASTP searches with default setting were targeted to the predicted and annotated *Plasmodium falciparum* genome that was published in the PlasmoDB (http://plasmodb.org/; Kissinger et al. 2002). A cut-off criteria of E-value <1e-04 was adopted to define protease homologs. Partial sequences (<80% of full-length) and redundant sequences were excluded. Conserved domains/motifs in *P. falciparum* sequences were identified by searching InterPro release 5.1, which integrates Pfam 7.3, PRINTS 33.0, PROSITE 17.5, ProDom 2001.3, SMART 3.1, TIGRFAMs 1.2, and the current SWISS-PROT + TrEMBL data.

Multiple alignments were obtained by the program T-coffee (Notredame et al. 2000), followed by manual editing according to the structure information. Graphic presentation of the alignment and consensus sequences were deduced by the program BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html). Phylogenetic trees were inferred by the neighbor-joining method (Saitou and Nei 1987) using MEGA2.0 (http://www.megasoftware.net/). Unweighted Maximum Parsimony (as implemented in PAUP 4.0) and Maximum Likelihood (as implemented in PHYLIP) were used to examine whether the inferred phylogeny is sensitive to any tree-making method. The bootstrap resampling with 1000 pseudoreplicates was carried out to assess support for each individual branch. Bootstrap values of <50% were collapsed and treated as unresolved polytomies.

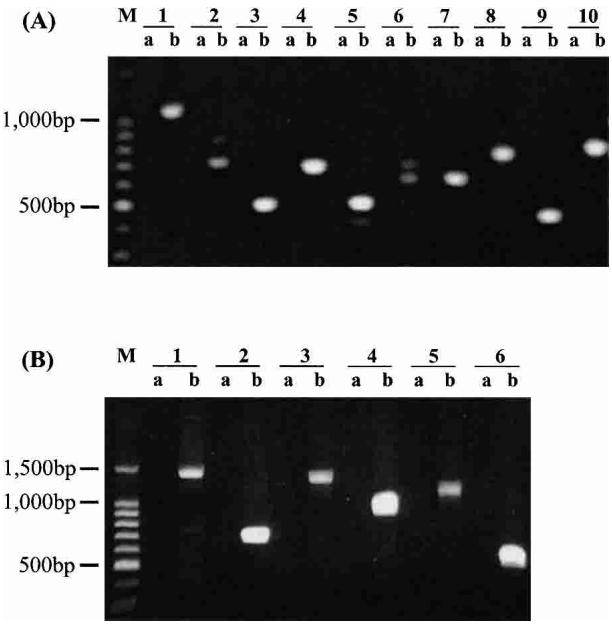

Figure 8 (*A*) Expression of ten putative proteases without corresponding oligomers in the microarray set. RT-PCR was conducted to examine the transcriptional expression of the ten putative proteases using specific primers based on the prediction. Lane *a* in each sample represents the negative control in which RT-PCR was conducted without reverse transcriptase. Lane *1b*: PF14_0281; Lane *2b*: PFL1635w; Lane *3b*: MAL6P1.153; Lane *4b*: PFL1925w; Lane *5b*: PFC0950C; Lane *6b*: MAL8P1.16; Lane *7b*: MAL6P1.88; Lane *8b*: MAL8P1.128; Lane *9b*: PFA0400c; Lane *10b*: PFI1545c.*M* indicates 1 kb DNA ladder. (*B*) Expression of putative calpain, caspase, and SP1 genes using two pairs of primers. RT-PCR was conducted to confirm the transcriptional expression of putative calpain, metacaspase, and SP-1 genes, using 2 pairs of specific primers for each gene. Lanes "a" represent negative controls in which RT-PCR was conducted without reverse transcriptase.*M* indicates 1 kb DNA ladder. Lanes *1b* and *2b*: MAL13P1.310; Lanes *3b* and *4b*: PF13_0289; Lanes *5b* and *6b*: PF13_0118. See Suppl. Table 1 for the predicted size of RT-PCR products.

Microarray Expression Analysis Using Asynchronous Erythrocytic *P. falciparum* Culture

An en masse gene expression profile was obtained using microarray chips arrayed with 6239 Malaria Genome Array Oligomers (Operon Technologies), designed by Dr. Joe DeRisi of the University of California at San Francisco (http://derisilab.ucsf.edu/). These 6239 70-mers mapped to 4407 predicted open reading frames which covered >90% of the available *P. falciparum* genome sequences. In order to obtain all genes transcribed throughout the erythrocyte stage of the parasite, we extracted and pooled mRNAs from *P. falciparum* 3D7 culture samples (Trager and Jensen 1976) collected at four 12-h intervals to achieve an asynchrony (shown in Fig. 3). Probes were labeled with fluorescent dyes using mRNAs purified from the asynchronous culture as a template. Messager RNAs were purified using oligo T cellulose, and reverse transcription was conducted to incorporate aminoallyl dUTP into the cDNAs. The Cy3 and Cy5 NHS esters were then coupled to amine groups of the cDNA, and dye-labeled probes were hybridized with the microarray slides under standard conditions (3×SSC, 50% formamide, 0.1% SDS, 10 mg/mL salmon sperm DNA, 68°C). The slide was scanned with a GenePix 4000B (Axon Instrument) at default PMT settings, 100% power. The array data were analyzed initially with GenePixPro software (Axon Instrument), then with global normalization. The expression level is indicated by the mean signal intensity of all corresponding oligomers in triplicates on the microarray slides (MRA-452) obtained from Malaria Research and Reference Resource Center (http://www.malaria.mr4.org/). Two sets of negative controls were included in the DeRisi design: (1) 20 oligomers from yeast intergenic region with the mean intensity 529, (2) 33 *P. falciparum* genes cloned into a plasmid, including 16 ribosomal proteins, 17 tRNA genes, LSU, *Clp*, and *tufA*. Their mean intensity was 598.

Reverse Transcription Polymerase Chain Reaction

RT-PCR was performed using the same mRNA described above as template. Reverse transcription was conducted using SuperScript II (Invitrogen). The PCR cycle: 95°C 1 min; (95°C 1 min, 54°C 30 sec, 52°C 30 sec, 65°C 1 min) x 35, 65°C 10 min, hold at 4°C. The primer sequences used to amplify 10 target genes without corresponding oligomers in the array set, and putative calpain, metacaspase, and signal peptidase I are included in the Supplemental Table 1.

ACKNOWLEDGMENTS

We thank Lois Blaine, David Emerson, and Thomas Nerad for their critical comments during the manuscript preparation, Truc Nguyen for computational support. This study is supported by an ATCC start-up fund to Y.W., and an NIH-grant (1R21AI49300) to Y.W. We thank the scientists and funding agencies comprising the International Malaria Genome Project for making sequence data from the genome of *P. falciparum* (3D7) public prior to publication of the completed sequence. The Sanger Centre (UK) provided sequence data for chromosomes 1, 3–9, and 13, with financial support from the Wellcome Trust. A consortium composed of The Institute for Genome Research, along with the Naval Medical Research Center (USA), sequenced chromosomes 2, 10, 11 & 14, with support from NIAID/NIH, the Burroughs Wellcome Fund, and the Department of Defense. The Stanford Genome Technology Center (USA) sequenced chromosome 12, with support from the Burroughs Wellcome Fund. The *Plasmodium* Genome Database is a collaborative effort of investigators at the University of Pennsylvania (USA) and Monash University (Melbourne, Australia), supported by the Burroughs Wellcome Fund.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aravind, L., Dixit, V.M., and Koonin, E.V. 1999. The domains of death: Evolution of the apoptosis machinery. *Trends Biochem. Sci.* **24:** 47–53.

Arthur, J.S., Gauthier, S., and Elce, J.S. 1995. Active site residues in m-calpain: Identification by site-directed mutagenesis. *FEBS Lett.* **368:** 397–400.

Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D.E. 2002. Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl. Acad. Sci.* **99:** 990–995.

Barale, J.C., Blisnick, T., Fujioka, H., Alzari, P.M., Aikawa, M., Braun-Breton, C., and Langsley, G. 1999. *Plasmodium falciparum* subtilisin-like protease 2, a merozoite candidate for the merozoite surface protein 1–42 maturase. *Proc. Natl. Acad. Sci.* **96:** 6445–6450.

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. Coordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Mol. Microbiol.* **39:** 26–36.

Bernstein, N.K., Cherney, M.M., Loetscher, H., Ridley, R.G., and James, M.N. 1999. Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nat. Struct. Biol.* **6:** 32–37.

Blackman, M.J. 2000. Proteases involved in erythrocyte invasion by the malaria parasite: Function and potential as chemotherapeutic targets. *Curr. Drug Targets* **1:** 59–83.

Blackman, M.J., Fujioka, H., Stafford, W.H., Sajid, M., Clough, B., Fleck, S.L., Aikawa, M., Grainger, M., and Hackett, F. 1998. A subtilisin-like protein in secretory organelles of *Plasmodium falciparum* merozoites. *J. Biol. Chem.* **273:** 23398–23409.

Braun-Breton, C., Rosenberry, T.L., and da Silva, L.P. 1988. Induction of the proteolytic activity of a membrane protein in *Plasmodium falciparum* by phosphatidyl inositol-specific phospholipase C. *Nature* **332:** 457–459.

Coombs, G.H., Goldberg, D.E., Klemba, M., Berry, C., Kay, J., and Mottram, J.C. 2001. Aspartic proteases of *Plasmodium falciparum* and other parasitic protozoa as drug targets. *Trends Parasitol.* **17:** 532–537.

Craik, C.S., Rutter, W.J., and Fletterick, R. 1983. Splice junctions: Association with variation in protein structure. *Science* **220:** 1125–1129.

Curley, G.P., O'Donovan, S.M., McNally, J., Mullally, M., O'Hara, H., Troy, A., O'Callaghan, S.A., and Dalton, J.P. 1994. Aminopeptidases from *Plasmodium falciparum*, Plasmodium chabaudi chabaudi and Plasmodium berghei. *J. Eukaryot. Microbiol.* **41:** 119–123.

David, P.H., Hadley, T.J., Aikawa, M., and Miller, L.H. 1984. Processing of a major parasite surface glycoprotein during the ultimate stages of differentiation in Plasmodium knowlesi. *Mol. Biochem. Parasitol.* **11:** 267–282.

Delplace, P., Bhatia, A., Cagnard, M., Camus, D., Colombet, G., Debrabant, A., Dubremetz, J.F., Dubreuil, N., Prensier, G., Fortier, B., et al. 1988. Protein p126: A parasitophorous vacuole antigen associated with the release of *Plasmodium falciparum* merozoites. *Biol. Cell* **64:** 215–221.

Denison, S.H., Orejas, M., and Arst Jr., H.N. 1995. Signaling of ambient pH in Aspergillus involves a cysteine protease. *J. Biol. Chem.* **270:** 28519–28522.

Doerig, C., Meijer, L., and Mottram, J.C. 2002. Protein kinases as drug targets in parasitic protozoa. *Trends Parasitol.* **18:** 366–371.

Dua, M., Raphael, P., Sijwali, P.S., Rosenthal, P.J., and Hanspal, M. 2001. Recombinant falcipain-2 cleaves erythrocyte membrane ankyrin and protein 4.1. *Mol. Biochem. Parasitol.* **116:** 95–99

Eggleson, K.K., Duffin, K.L., and Goldberg, D.E. 1999. Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **274:** 32411–32417.

Fast, N.M., Kissinger, J.C., Roos, D.S., and Keeling, P.J. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18:** 418–426.

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.

Florent, I., Derhy, Z., Allary, M., Monsigny, M., Mayer, R., and Schrevel, J. 1998. A *Plasmodium falciparum* aminopeptidase gene belonging to the M1 family of zinc-metallopeptidases is expressed in erythrocytic stages. *Mol. Biochem. Parasitol.* **97:** 149–160.

Franz, T., Vingron, M., Boehm, T., and Dear, T.N. 1999. Capn7: A highly divergent vertebrate calpain with a novel C-terminal domain. *Mamm. Genome.* **10:** 318–321.

Friedman, R., and Hughes, A.L. 2001. Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11:** 1842–1847.

Gantt, S.M., Myung, J.M., Briones, M.R., Li, W.D., Corey, E.J., Omura, S., Nussenzweig, V., and Sinnis, P. 1998. Proteasome inhibitors block development of Plasmodium spp. *Antimicrob. Agents Chemother.* **42:** 2731–2738

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* **41:** 498– 511.

Glading, A., Lauffenburger, D.A., and Wells, A. 2002. Cutting to the chase: Calpain proteases in cell motility. *Trends Cell Biol.* **12:** 46–54.

Gu, X., Wang, Y., and Gu, J. 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet.* **31:** 205–209.

Hackett, F., Sajid, M., Withers-Martinez, C., Grainger, M., and Blackman, M.J. 1999. PfSUB-2: A second subtilisin-like protein in *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* **103:** 183–195.

Hayward, R.E., Derisi, J.L., Alfadhli, S., Kaslow, D.C., Brown, P.O., and Rathod, P.K. 2000. Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35:** 6–14.

Jean, L., Long, M., Young, J., Pery, P., and Tomley, F. 2001. Aspartyl proteinase genes from apicomplexan parasites: Evidence for evolution of the gene structure. *Trends Parasitol.* **17:** 491–498.

Kissinger, J.C., Brunk, B.P., Crabtree, J., Fraunholz, M.J., Gajria, B., Milgram, A.J., Pearson, D.S., Schug, J., Bahl, A., Diskin, S.J., et al. 2002. The Plasmodium genome database. *Nature* **419:** 490–492.

Li, J., Matsuoka, H., Mitamura, T., and Horii, T. 2002. Characterization of proteases involved in the processing of *Plasmodium falciparum* serine repeat antigen (SERA). *Mol. Biochem. Parasitol.* **120:** 177–186.

Li, W-H. 1983. Evolution of duplicate genes and pseudogenes. In *Evolution of genes and proteins* (eds. M. Nei and R.K. Keohn), pp. 14–37. RK Sinauer Associates, Sunderland, MA.

Madeo, F., Herker, E., Maldener, C., Wissing, S., Lachelt, S., Herlan, M., Fehr, M., Lauber, K., Sigrist, S.J., Wesselborg, S., et al. 2002. A caspase-related protease regulates apoptosis in yeast. *Mol. Cell.* **9:** 911–917.

McKerrow, J.H., Sun, E., Rosenthal, P.J., and Bouvier, J. 1993. The proteases and pathogenicity of parasitic protozoa. *Annu. Rev. Microbiol.* **47:** 821–853.

McLysaght, A., Hokamp, K., and Wolfe, K.H. 2002. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31:** 200–204.

Miller, S.K., Good, R., Drew, D.R., Delorenzi, M., Sanders, P.R., Hodder, A.N., Speed, T.P., Cowman, A.F., De Koning-Ward, T.F., and Crabb, B.S. 2002. A subset of *Plasmodium falciparum* SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J. Biol. Chem.* **277:** 47524–47532.

Mu, J., Duan, J., Makova, K.D., Joy, D.A., Huynh, C.Q., Branch, O.H., Li, W.H., and Su, X.Z. 2002. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* **418:** 323–326.

Narum, D.L. and Thomas, A.W. 1994. Differential localization of full-length and processed forms of PF83/AMA-1 an apical membrane antigen of *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* **67:** 59–68.

Notredame, C., Higgins, D.G., and Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302:** 205–217.

Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Berlin

Olaya, P. and Wasserman, M. 1991. Effect of calpain inhibitors on the invasion of human erythrocytes by the parasite *Plasmodium falciparum*. *Biochim. Biophys. Acta.* **1096:** 217–221.

Paetzel, M., Dalbey, R.E., and Strynadka, N.C. 2000. The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target. *Pharmacol. Ther.* **87:** 27–49.

Rawlings, N.D. and Barrett, A.J. 1993. Evolutionary families of peptidases. *Biochem. J.* **290:** 205–218.

Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. 1998. Malaria's Eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **95:** 4425–4430.

Rosenthal, P.J. 1998. Proteases of malaria parasites: New targets for chemotherapy. *Emerg. Infect. Dis.* **4:** 49–57.

Rosenthal, P.J. 2002. Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr. Opin. Hematol.* **9:** 140–145.

Rosenthal, P.J., Kim, K., McKerrow, J.H., and Leech, J.H. 1987. Identification of three stage-specific proteinases of *Plasmodium falciparum*. *J. Exp. Med.* **166:** 816–821.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.

Salmon, B.L., Oksman, A., and Goldberg, D.E. 2001. Malaria parasite exit from the host erythrocyte: A two-step process requiring extraerythrocytic proteolysis. *Proc. Natl. Acad. Sci.* **98:** 271–276.

Shenai, B.R., Sijwali, P.S., Singh, A., and Rosenthal, P.J. 2000. Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of *Plasmodium falciparum*. *J. Biol. Chem.* **275:** 29000–29010.

Silva, A.M., Lee, A.Y., Gulnik, S.V., Maier, P., Collins, J., Bhat, T.N., Collins, P.J., Cachau, R.E., Luker, K.E., Gluzman, I.Y., et al. 1996.

Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **93:** 10034–10039.
Sorimachi, H., Ishiura, S., and Suzuki, K. 1997. Structure and physiological function of calpains. *Biochem. J.* **328:** 721–732.
Southan, C. 2001. A genomic perspective on human proteases. *FEBS Lett.* **498:** 214–218.
Thornberry, N.A. and Lazbnik, L. 1998. Caspases: Enemies within. *Science* **281:** 1312–1316.
Trager, W. and Jensen, J.B. 1976. Human malaria parasites in continuous culture. *Science.* **193:** 673–675.
Triglia, T., Thompson, J.K., and Cowman, A.F. 2001. An EBA175 homolog which is transcribed but not translated in erythrocytic stages of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **116:** 55–63.
Tyas, L., Gluzman, I., Moon, R.P., Rupp, K., Westling, J., Ridley, R.G., Kay, J., Goldberg, D.E., and Berry, C. 1999. Naturally-occurring and recombinant forms of the aspartic proteinases plasmepsins I and II from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett.* **454:** 210–214.
Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. 2000. Identification of paracaspases and metacaspases: Two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol. Cell* **6:** 961–967.
Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates III, J.R., Koonin, E.V., and Deshaies, R.J. 2002. Role of Rpn11 Metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298:** 611–615.
Verra, F. and Hughes, A.L. 1999. Natural selection on apical membrane antigen-1 of *Plasmodium falciparum*. *Parassitologi.* **41:** 93–95.
Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., and Hartl, D.L. 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293:** 482–484.
Wang, Y. and Gu, X. 2001. Functional divergence in caspase gene family and altered functional constraints: Statistical analysis and prediction. *Genetics* **158:** 1311–1320.
Wilson, R.J., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., et al. 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261:** 155–172.
Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A., Mu, J., Baruch, D.I., Magill, A.J., and Su, X.Z. 2002. Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418:** 320–323.
Yoshizawa, T., Sorimachi, H., Tomioka, S., Ishiura, S., and Suzuki, K. 1995. A catalytic subunit of calpain possesses full proteolytic activity. *FEBS Lett.* **358:** 101–103.
Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., and Schneider, G. 2001. Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* **280:** 19–26.

WEB SITE REFERENCES

http://www.merops.ac.uk; a catalogue and structure-based classification of proteases.
http://www.expasy.ch/cgi-bin/lists?peptidas.txt; classification of peptidase (protease) families in SWISS-PROT.
http://plasmodb.org; official database of the malaria parasite genome project.
http://www.ch.embnet.org/software/BOX_form.html; software for printing and shading of multiple alignment files.
http://www.megasoftware.net/; software package for molecular evolutionary genetics analysis.
http://derisilab.ucsf.edu/; microarray resources provided by Dr. Joseph DeRisi at University of California, San Francisco.
http://www.malaria.mr4.org/; Malaria Research and Reference Reagent Resource Center.

Received October 16, 2002; accepted in revised form January 28, 2003.