Tropical Medicine and Health Vol. 39 No. 3, 2011, pp. 77-82
doi:10.2149/tmh.2011-12

Original article

Positive selection on the *Plasmodium falciparum clag2* gene encoding a component of the erythrocyte-binding rhoptry protein complex

Jean SF Alexandre[1,2], Morakot Kaewthamasorn[1], Kazuhide Yahata[1], Shusuke Nakazawa[1] and Osamu Kaneko[1*]
Received 23 May, 2011 Accepted 24 May, 2011 Published online 6 August, 2011

Abstract: A protein complex of high-molecular-mass proteins (*Pf*RhopH) of the human malaria parasite *Plasmodium falciparum* induces host protective immunity and therefore is a candidate for vaccine development. Clarification of the level of polymorphism and the evolutionary processes is important both for vaccine design and for a better understanding of the evolution of cell invasion in this parasite. In a previous study on 5 genes encoding RhopH1/Clag proteins, positive diversifying selection was detected in *clag8* and *clag9* but not in the paralogous *clag2*, *clag3.1* and *clag3.2*. In this study, to extend the analysis of *clag* polymorphism, we obtained sequences surrounding the most polymorphic regions of *clag2*, *clag8*, and *clag9* from parasites collected in Thailand. Using sequence data obtained newly in this study and reported previously, we classified *clag2* sequences into 5 groups based on the similarity of the deduced amino acid sequences and number of insertions/deletions. By the sliding window method, an excess of nonsynonymous substitutions over synonymous substitutions was detected in the group 1 and group 2 *clag2* and *clag8* sequences. Population-based analyses also detected a significant departure from the neutral expectation for group 1 *clag2* and *clag8*. Thus, two independent approaches suggest that *clag2* is subject to a positive diversifying selection. The previously suggested positive selection on *clag8* was also supported by population-based analyses. However, the positive selection on *clag9*, which was detected by comparing the 5 sequences, was not detected using the additional 34 sequences obtained in this study.
Key words: malaria, rhoptry, polymorphism

1. INTRODUCTION

Plasmodium falciparum, the causative agent of malignant tertian malaria is, for part of its life cycle, an obligate intra-erythrocytic parasite. This stage of the life cycle is characterized by repeating cycles of erythrocyte invasion, followed by growth and schizogonic multiplication within the cell, egress, and re-invasion. The erythrocyte invasive unit, the merozoite, expresses a panel of proteins on its surface, all of which are exposed to the host immune system for the brief time that the parasite is free in the blood plasma. These proteins, exemplified by merozoite surface protein 1 (MSP1), tend to be more polymorphic (showing a higher degree of amino acid diversity between parasite strains) than other non-exposed parasite proteins. A possible explanation for this is that such proteins have undergone "positive diversifying selection", mediated by host immune pressure [1]. Proteins secreted by the parasite at the point of cell invasion are also exposed, however briefly, to the host immune system, and may be targeted by immune pressures, and polymorphism selected within them, in much the same way as merozoite surface proteins. Apical membrane antigen (AMA1), released from the merozoite micronemes during invasion, for example, is believed to have undergone positive diversifying selection in this manner [2].

The RhopH complex is a high molecular mass

[1] Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global COE Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
[2] Centro Nacional de Control de Enfermedades Tropicales, Santo Domingo, República Dominicana
*Corresponding author:
Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global COE Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
Tel.: (+81) 95 819 7838
Fax: (+81) 95 819 7805
E-mail: okaneko@nagasaki-u.ac.jp
Sequence data from this article were deposited in the GenBank™/EMBL/DDBJ databases under accession numbers AB633214–AB633323, and AB634454.
Abbreviations: *Clag*, cytoadherence-linked asexual gene; DNA, deoxyribo–nucleotide acid; indels, insertions/deletions; nt, nucleotides; PCR, polymerase chain reaction

erythrocyte-binding protein complex secreted from the merozoite rhoptry during invasion, and antibodies raised against it have been shown to confer anti-parasite protection to the host [3–6]. The precise role of this complex in erythrocyte invasion remains unclear. Components of this complex have been detected on the erythrocyte cytosol side of the parasitophorous vacuole membrane and the parasite-infected erythrocyte membrane, suggesting a role in the formation of the parasite-restructured membranous architecture in the infected erythrocyte [7, 8]. The RhopH complex is itself comprised of three distinct proteins: RhopH1, RhopH2 and RhopH3, each encoded by separate genes [9–11]. RhopH1 is encoded by a multigene family termed the cytoadherence-linked asexual gene (*rhoph1/clag*) family, that consists of at least five paralogous genes (*clag2*, *clag3.1*, *clag3.2*, *clag8*, and *clag9*), and each RhopH complex contains one of the *rhoph1/clag* gene products [9, 12, 13].

In a previous study, the degree of inter-allelic polymorphism for seven RhopH complex-related genes, 5 *rhoph1/clag*, *rhoph2* and *rhoph3*, was evaluated by comparing nucleotide sequences from 5 culture-adapted parasite lines. It was found that *clag2*, *clag3.1*, *clag3.2* and *clag8* were highly polymorphic and that amino acid substitutions and insertions/deletions (indels) were found mainly in a region encompassing amino acid positions 1000–1200 of these gene products. The signature of positive selection was detected by an excess of nonsynonymous substitutions over synonymous substitutions for *clag8* and *clag9*. Further evidence for the involvement of positive selection in driving *clag8* polymorphism was independently gathered in an analysis that compared the *Plasmodium reichenowi* orthologous sequence for *clag8* to the *P. falciparum* gene [14].

In this study, we extend our analysis of *clag* polymorphism further by increasing the number of sequences and employing a population-based approach utilizing sequences obtained from parasites collected in Thailand.

2. MATERIALS AND METHODS

2.1. Parasite culture and DNA extraction

All 39 *P. falciparum* lines examined in this study (MS802, MS803, MS804, MS805, MS806, MS807, MS808, MS809, MS810, MS811, MS812, MS814, MS815, MS816, MS817, MS818, MS819, MS820, MS821, MS822, MS824, MS825, MS826, MS827, MS828, MS829, MS830, MS831, MS833, MS834, MS835, MS837, MS838, MS840, MS842, MS843, MS844, MS946, and MS947) were collected in Mae Sot, Thailand from November 21, 1988 to January 16, 1989 and were maintained in vitro essentially as previously described [15–17]. Since MS814 and MS822 showed unclear chromatograms for *clag* sequences, these lines were cloned by limiting dilution, yielding clones MS814K and MS814R, and MS822B6 and MS822G8, respectively. The human erythrocytes and plasma used for culture were obtained from the Nagasaki Red Cross Blood Center. Parasites were harvested when parasitemia reached about 2%, and parasite DNA was extracted using DNAzol BD (Invitrogen).

2.2. Polymerase chain reaction (PCR) amplification and sequencing

DNA fragments were PCR-amplified twice, independently, with oligonucleotide primers: TATATGGAAAAA GTAGTAATATACAGG and TACTAGTATGTGGTTGAT ATTCTTTTG for *clag2* (resulting PCR product with the size of 702 bp); GTTTATGGAAAAAGTGGTAAAATAGG and CTCTTTAAGTTTTCTTCTGAATAGTTC for *clag8* (750 bp); and ATAAACTTGATAGAATATATGGTAAAGC and ATTGAATAATCTTTTAATGTACATGCAC for *clag9* (764 bp) in a 20 μL reaction mixture using a high-fidelity KOD Plus DNA polymerase (TOYOBO, Japan). The PCR conditions were as follows: 94°C for 2 min; 40 cycles of 92°C for 15 sec, 54°C for 20 sec, 68°C for 1 min 10 sec; final extension step of 68°C for 5 min. PCR products were subjected to 1.5% agarose gel electrophoresis, and when a single band product with no background was observed, PCR-amplified DNA fragments were directly sequenced following treatment of PCR mixture with ExoSAP-IT (GE Healthcare, UK). Two independent PCR products were sequenced using a panel of primers described previously [14]; one in the forward direction and the other in the reverse direction, to ensure the accuracy of the obtained sequences. Sequencing reactions were performed using the BigDye® Terminator v1.1 Kit (Applied BioSystems, UK) with an ABI3730 DNA analyzer (Applied BioSystems). Sequences were manually corrected using BioEdit 7.0.0 software [18]. Sequences of *clag2* (AB250822) and *clag8* (AB250849) for MS838 were obtained from the database.

2.3. Statistical analyses of genetic diversity

Nucleotide diversity (π) and its standard error (SE) were computed by the Jukes and Cantor method using MEGA 4.0 software [19]. The mean number of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) and their standard errors were computed using the Nei and Gojobori method [20] with the Jukes and Cantor correction, implemented in MEGA 4.0. The statistical difference between d_S and d_N was tested using a one-tailed Z-test with 500 bootstrap pseudo-samples in MEGA 4.0. A value of d_N significantly higher than d_S at the 95% confidence level was taken as evidence for positive selection. The d_N/d_S ratio was evaluated using a sliding window method (90 bases with a step size of 3

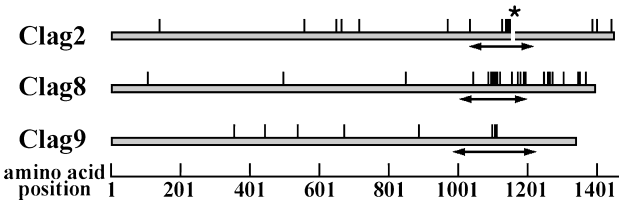

Fig. 1. Regions used for the analysis. Bars with arrows under each schematic for Clag protein indicate the regions used for the analysis in this study. Vertical bars and a gap with asterisk above each schematic indicate polymorphic sites among 5 *P. falciparum* laboratory isolates and an indel reported previously [14]. Amino acid position is according to the 3D7 line sequence.

bases) in DnaSP 4.0 [21]. Population genetics tests of neutrality were applied to the 3 *clag* gene sequences. Tajima's test was used to test for departure from neutrality as measured by the difference between π (observed average pairwise nucleotide diversity) and θ (expected nucleotide diversity under neutrality derived from the number of segregating sites, *S*). Under positive diversifying selection, rare alleles are selected and maintained at intermediate frequencies, elevating π above that expected under neutrality and making the value of the test statistic (*D*) positive [22]. Fu and Li's test was also used to evaluate positive diversifying selection by comparing estimates of θ based on the number of singletons and that derived from *S* (the *D** index) or π (the *F** index). Under positive diversifying selection, an excess of intermediate frequency polymorphisms and lower number of singletons make the value of *D** and *F** positive [23]. The regions analyzed in this study are shown in Fig. 1.

3. Results and Discussion

3.1. Clag2 sequence is classified into 5 groups with the signature of positive diversifying selection detected on groups 1 and 2.

In this study, we obtained 35 sequences of the most polymorphic region of *clag2*, 40 sequences for *clag8* and 34 sequences for *clag9*. Iriko *et al.* (2008) did not detect the signature of positive diversifying selection on *clag2* using a set of sequences which excluded the most polymorphic regions with indels, due to difficulty in obtaining a reliable nucleotide sequence alignment [14]. Thus, we classified *clag2* into distinct groups using the deduced amino acid sequences from 35 *P. falciparum clag2* alleles generated during this work, 31 previously reported alleles and one *clag2* ortholog (*prclag2*) from the chimpanzee malaria parasite *Plasmodium reichenowi*, based on amino acid similarity and numbers of indels, and attempted to detect the signature of positive selection for each group separately. We classified *Pf*Clag2 sequences into five groups (Fig. 2). Among 66 *Pf*Clag2 sequences, 44 sequences (including the 3D7 line sequence) were classified in group 1, 11 sequences with a double amino acid insertion in group 2, five sequences with a seven amino acid insertion in group 3, four sequences with an eight amino acid insertion in group 4, and two sequences showing a high degree of similarity to the *P. reichenowi* Clag2 ortholog sequence in group 5. The observations that 1) amino acid sequences are clearly distinct between groups, 2) the group 5 sequence is similar to the *P. reichenowi* Clag2 ortholog sequence, and 3) the older origin of *clag2* polymorphism than the time to the most recent common ancestor of the extant *P. falciparum* population was previously proposed based on the analysis using sequences excluding indels [14], raise the possibility that these 5 distinct Clag2 groups may have been generated before the time to the most recent common ancestor of the extant *P. falciparum* population. This may be clarified by analyzing the sequences of Clag2 orthologs from the malaria parasite species recently discovered in gorillas, and currently thought to be the closest relative species [25].

We attempted to detect the signature of positive diversifying selection on *clag2* alleles belonging to groups 1 (n = 24) and 2 (n = 8) and *clag8* (n = 41) by comparing d_N and d_S for sequences obtained in Thailand during the 1988–1989 period ("Thai" in Fig. 3), or 44, 11, or 69 sequences, respectively, after combining sequences reported previously ("Thai + others" in Fig. 3) [14]. No significant difference was detected between d_N and d_S, when the nucleotide sequences corresponding to nucleotide position (nt) 3106–3642 for *clag2* or nt 3022–3591 for *clag8* (numbered according to the 3D7 line sequence) were analyzed. However, when d_N and d_S were compared for the region where a high d_N/d_S ratio was detected by the sliding window plot method, a significant excess of d_N over d_S was detected for group 1 *clag2* (mid point nt position of 3444 for "Thai + others", d_N/d_S = 3.13, p<0.02; mid point nt position of 3447–3453 for "Thai + others", d_N/d_S = 3.13–3.27, p<0.05), group 2 *clag2* (mid point nt position of 3396–3459 for "Thai" only, d_N/d_S = 5.90–7.65, p<0.02; mid point nt position of 3396–3459 for "Thai + others", d_N/d_S = 8.02–10.02, p<0.02), and *clag8* (mid point nt position of 3258–3309 for Thai only, d_N/d_S = 3.49–53.96, p<0.02; mid point nt position of 3258–3309 for Thai + others, d_N/d_S = 3.21–33.27, p<0.02; mid point nt position of 3531–3546 for Thai + others, d_N/d_S = 6.99–7.68, p<0.05), suggesting that positive diversifying selection had operated on these regions. Positive selection was not detected for *clag9* using sequences from Thai only (n = 34) and Thai + others (n = 39). Thus, in

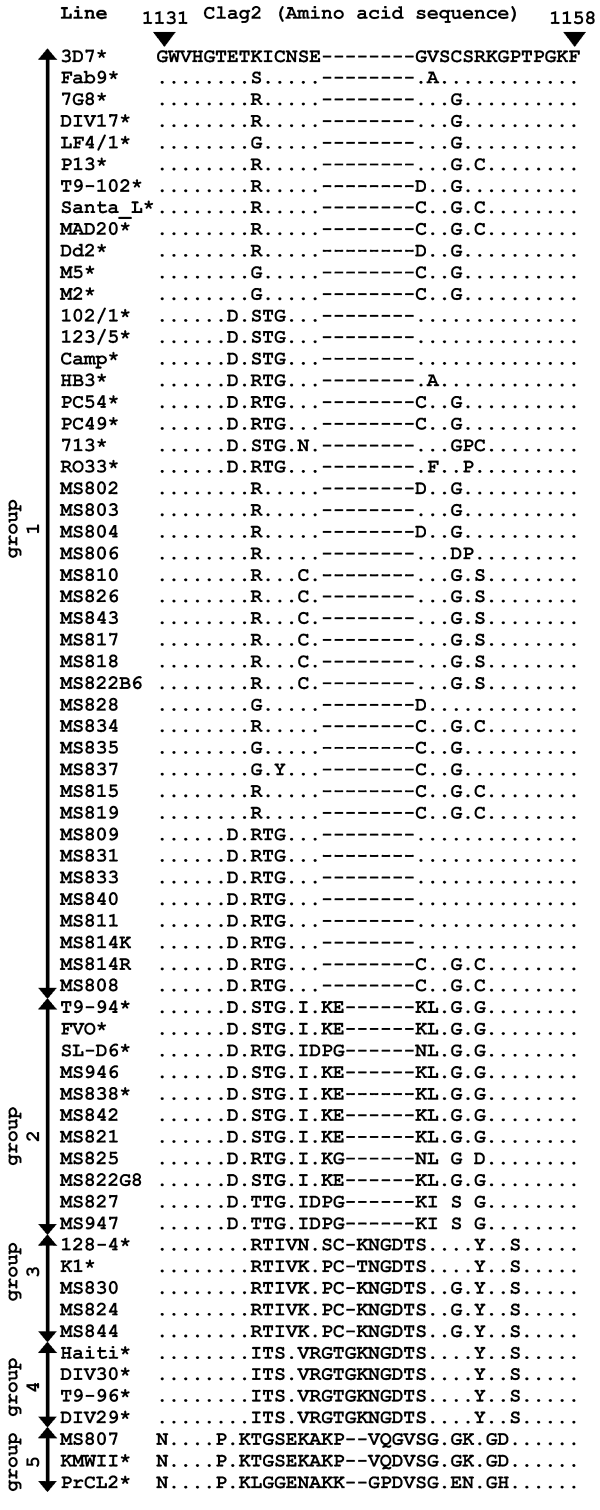

Fig. 2. Clag2 is classified into 5 groups based on the similarity in the variable region. Dots and bars indicate identical amino acid residues with the 3D7 line sequence and gaps. Asterisks indicate sequences reported previously [14]. PrCL2 indicate *Plasmodium reichenowi* Clag2. Amino acid positions shown above the sequences are according to the 3D7 line sequence.

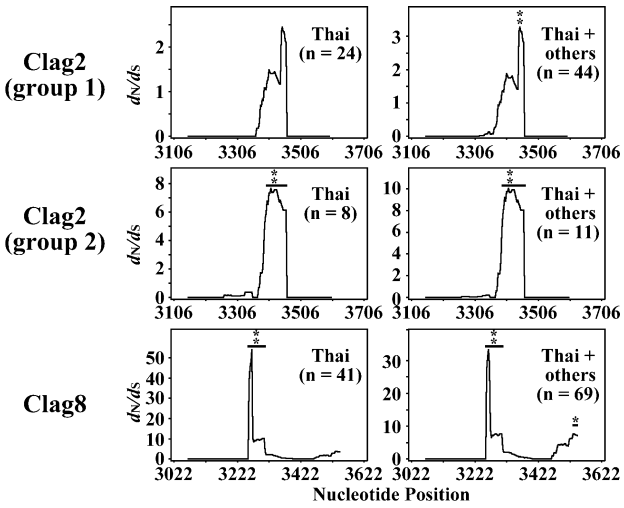

Fig. 3. Sliding window plot of d_N/d_S ratio for *Plasmodium falciparum clag2* (group 1 and 2) and *clag8* in Thai isolates and in a set combined with laboratory isolates. Nucleotide positions are according to the 3D7 line sequence. Window length is 90 bp, and step size is 3 bp. n indicates the number of samples analyzed. Asterisks indicate the region where a significant excess of non-synonymous substitutions (d_N) over synonymous substitutions (d_S) was observed (single for $p<0.05$, and double for $p<0.02$). The statistical difference between d_S and d_N was tested using a one-tailed Z-test with 500 bootstrap pseudosamples implemented in MEGA4.0. Others indicate sequence reported previously and shown in Fig. 2 [14].

addition to *clag8* and *clag9*, for which a positive selection was detected in a previous study [14], positive diversifying selection was detected here for at least two groups of *clag2* sequences (Fig. 3).

3.2. Population-based analyses also detected positive selection on group 1 *clag2* and *clag8*

Positive diversifying selection was evaluated for group 1 *clag2* (n = 24), *clag8* (n = 41), and *clag9* (n = 34) using sequences obtained from Thailand within two months by a population-based approach. Using Tajima's test and Fu and Li's test for nt sequences obtained in this study (nt 3106–3642 of *clag2*, 3022–3591 of *clag8*, and nt 2947–3639 of *clag9*; nucleotide positions are according to the 3D7 line sequence), no significant departure from the neutral expectation was detected. However, the sliding window plot method detected a significantly high Tajima's *D* value for group 1 *clag2* (mid point nt position of 3492, $D = 2.26$, $p<0.05$) and *clag8* (mid point nt position of 3363, $D = 2.07$, $p<0.05$), Fu and Li's D^* value for *clag8* (mid point nt position of 3258–3264, $D^* = 1.48$–1.56, $p<0.05$; mid point

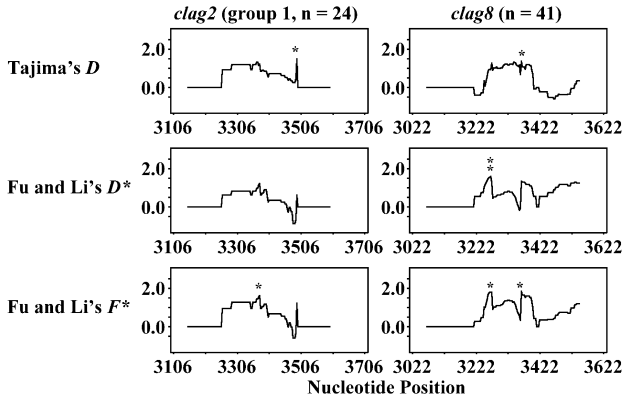

Fig. 4. Sliding window plot of Tajima's *D*, Fu & Li's *D** and *F** tests for *Plasmodium falciparum clag2* (group 1) and *clag8* genes in Thai isolates. Nucleotide positions are according to the 3D7 line sequence. Window length is 90 bp, and step size is 3 bp. n indicates the number of samples analyzed. Asterisks indicate the region where a significant departure from the neutrality was observed (single for $p<0.05$, and double for $p<0.02$).

nt position of 3267, *D** = 1.61, $p<0.02$), and Fu and Li's *F** value for group 1 *clag2* (mid point nt position of 3375, *F** = 1.64, $p<0.05$) and *clag8* (mid point nt position of 3261–3270, *F** = 1.75–1.81, $p<0.05$; mid point nt position of 3363, *F** = 1.86, $p<0.05$) (Fig. 4). Thus, in addition to the detection of positive selection on group 1 *clag2* and *clag8* detected by an excess of d_N over d_S, the results of the population-based method also support the finding that both group 1 *clag2* and *clag8* are subject to a positive diversifying selection.

In summary, two independent tests, one by comparing d_N and d_S and the other based on the population, suggest that the region of *clag2* is under positive diversifying selection. Previously suggested positive selection on *clag8* was also supported by population-based analyses. Such observations are consistent with the action of balancing selection maintaining allelic variation in the population. Contrary to a previous report [14], positive selection on *clag9* was not detected in this study using an additional 34 sequences. As only 5 *clag9* sequences were used to compare d_N and d_S in this previous study, a low sample number might have generated a false positive result. Further study is required to determine if *clag9* is under positive selection.

ACKNOWLEDGEMENTS

We thank R. Culleton for critical reading. We are grateful to I. Sekine, head of the Nagasaki Red Cross Blood Center for human erythrocyte and plasma. This work was supported in part by Grants-in-Aids for Scientific Research 19590428 (to OK) and the Global COE Program, Nagasaki University (to OK) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The nucleotide sequence data reported in this paper are available in the GenBank™/EMBL/DDBJ databases under the accession numbers: AB633214–AB633323, and AB634454. Part of this work was performed during Master of Tropical Medicine course in Nagasaki University supported by the Japanese International Cooperation Agency (JICA). J.A. is a recipient of MEXT PhD scholarship, Japan.

REFERENCES

1. Conway DJ. Natural selection on polymorphic malaria antigens and the search for a vaccine. Parasitol Today 1997; 13: 26–29.
2. Crewther PE, Matthew ML, Flegg RH, Anders RF. Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. Infect Immun 1996; 64: 3310–3317.
3. Siddiqui WA, Tam LQ, Kramer KJ, *et al*. Merozoite surface coat precursor protein completely protects Aotus monkeys against *Plasmodium falciparum* malaria. Proc Natl Acad Sci USA 1987; 84: 3014–3018.
4. Cooper JA, Ingram LT, Bushell GR, *et al*. The 140/130/105 kilodalton protein complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. Mol Biochem Parasitol 1988; 29: 251–260.
5. Doury JC, Bonnefoy S, Roger N, *et al*. Analysis of the high molecular weight rhoptry complex of *Plasmodium falciparum* using monoclonal antibodies. Parasitology 1994; 18: 269–280.
6. Rungruang T, Kaneko O, Murakami Y, *et al*. Erythrocyte surface glycosyl-phosphatidyl inositol anchored receptor for the malaria parasite. Mol Biochem Parasitol 2005; 140: 13–21.
7. Hiller NL, Akompong T, Morrow JS, Holder AA, Haldar K. Identification of a stomatin orthologue in vacuoles induced in human erythrocytes by malaria parasites. A role for microbial raft proteins in apicomplexan vacuole biogenesis. J Biol Chem 2003; 278: 48413–48421.
8. Goel S, Valiyaveettil M, Achur RN, *et al*. Dual stage synthesis and crucial role of cytoadherence-linked asexual gene 9 in the surface expression of malaria parasite *var* proteins. Proc Natl Acad Sci USA 2010; 107: 16643–16648.
9. Kaneko O, Tsuboi T, Ling IT, *et al*. The high molecular mass rhoptry protein, RhopH1, is encoded by members of the *clag* multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2001; 118: 223–231.
10. Ling IT, Kaneko O, Narum DL, *et al*. Characterization of the *rhoph2* gene of *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2003; 127: 47–57.
11. Shirano M, Tsuboi T, Kaneko O, Tachibana M, Adams JH, Torii M. Conserved regions of the *Plasmodium yoelii*

rhoptry protein RhopH3 revealed by comparison with the *P. falciparum* homologue. Mol Biochem Parasitol 2001; 112: 297–299.
12. Holt DC, Gardiner DL, Thomas EA, *et al*. The cytoadherence linked asexual gene family of Plasmodium falciparum: are there roles other than cytoadherence? Int J Parasitol 1999; 29: 939–944.
13. Kaneko O, Yim Lim BY, Iriko H, *et al*. Apical expression of three RhopH1/Clag proteins as components of the *Plasmodium falciparum* RhopH complex. Mol Biochem Parasitol 2005; 143: 20–28.
14. Iriko H, Kaneko O, Otsuki H, *et al*. Diversity and evolution of the *rhoph1*/*clag* multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol 2008; 158: 11–21.
15. Trager W, Jensen JB. Human malaria parasites in continuous culture. Science 1976; 193: 673–675.
16. Jongwutiwes S, Tanabe K, Kanbara H. Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of *Plasmodium falciparum* from field isolates. Mol Biochem Parasitol 1993; 59: 95–100.
17. Nakazawa S, Culleton R, Maeno Y. In vivo and in vitro gametocyte production of *Plasmodium falciparum* isolates from Northern Thailand. Int J Parasitol 2011; 41: 317–323.
18. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999; 41: 95–98.
19. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24: 1596–1599.
20. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986; 3: 418–426.
21. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003; 19: 2496–2497.
22. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 1993; 135: 599–607.
23. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993; 133: 693–709.
24. Liu W, Li Y, Learn GH, *et al*. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 2010; 467: 420–425.