17

Clues to Evolution of the SERA Multigene Family in the Genus *Plasmodium*

Nobuko Arisue, Nirianne M. Q. Palacpac,
Kazuyuki Tanabe and Toshihiro Horii
Research Institute for Microbial Diseases, Osaka University
Japan

1. Introduction

Malaria, one of the most serious infectious diseases prevalent in the tropics, is caused by the genus *Plasmodium*. Despite considerable global efforts to control this parasitic disease at least 90% of deaths still occur in sub-Saharan Africa (WHO, 2010). The rising threat of drug-resistant parasites, together with key interventions dependent on the use of a limited class of insecticides, underscore the fragility of malaria control. A better understanding of the parasite biology is required to gain insights for cost-effective tools or strategies including malaria vaccines and new antimalarial drugs that can be instrumental for sustained control, if not elimination, of malaria.

Malaria parasites comprise a diverse group of over 200 *Plasmodium* species that infect mammals, birds and reptiles (Levine, 1988). Each *Plasmodium* species exhibits a restricted host range, such that primate parasite cannot infect rodent, bird or reptile hosts. To find genomic factors that determine host range transcends the interest of malaria researchers. It is essential for control and conservation of wildlife. Recently, genome projects on several *Plasmodium* species from different hosts have been completed (Gardner et al., 2002; Carlton et al., 2002, 2008; Pain et al., 2008), with gene information available in the public database (http://plasmodb.org/plasmo/). By comparing the genomes from different species we obtain basic information at the molecular level on how *Plasmodium* has evolved and allows us to infer the function of genes and noncoding regions in the genome. One of the prominent features of *Plasmodium* genomes is the presence of various unique multigene families. Multigene families are a group of related genes that are presumed to share a common ancestor and are derived from each other by duplication and subsequent divergence. One such example, the largest family identified so far in human, primate and rodent malaria, is the *Plasmodium* interspersed repeat, pir (Janssen et al., 2004). The pir gene family members are highly species-specific, suggesting evolution of lineage-specific immune evasion mechanisms. *P. falciparum* var gene family is by far the best documented multigene family of the most virulent human malaria parasite. Products of var genes appear on the surface of infected erythrocytes and are involved in antigenic variation to evade host immunity. Other species-specific gene families encode proteins involved in host cell invasion, e.g. rhoptry proteins and parasite surface antigens, merozoite surface protein-3 and -7. There are also examples of families with few gene members. In sharp contrast to several hundreds of tandem arrayed rRNA gene family members in other eukaryotes,

Plasmodium has 4-7 gene units physically separated in the genome (Nei & Rooney, 2005; Carlton et al., 2008). Thus, *Plasmodium* possesses unique multigene families with distinctive evolutionary conundrums.

For more than 10 years after the first description of a gene family member, the existence of the *Plasmodium* serine repeat antigen (SERA) multigene family has been overlooked. Serine repeat antigen family proteins share homology with the papain family of cysteine proteases (Kiefer et al., 1996; Gor et al., 1998; Bourgon et al., 2004; Arisue et al., 2007, 2011). Almost all SERA genes are clustered in a head-to-tail manner and the number of SERA genes in the clustered region varies among parasite species (Bourgon et al., 2004; McCoubrie et al., 2007; Arisue et al., 2007, 2011). This leads us to infer that gene duplication occurred repeatedly during evolution. Some SERA genes were confirmed to play essential role(s) in the parasite life cycle (Miller et al., 2002; Aly & Matuschewski, 2005; McCoubrie et al., 2007). In addition, a gene family member in *P. falciparum*, SERA5, is a vaccine candidate now on phase Ib clinical trial (Horii et al., 2010). Two observations promise SERA5 as a vaccine candidate: (1) SERA genes are not differently expressed like other antigen encoding gene families such as var and rifin that show antigenic variation to evade host immune response (Aoki et al., 2002; Miller et al., 2002; Palacpac et al., 2006; Schmidt- Christensen et al., 2008; Putrianti et al., 2010; Arisue et al., 2011); and (2) *P. falciparum* SERA5 is less polymorphic (Fox et al., 1997: Morimatsu et al., 1997; Liu et al., 2000) than other vaccine candidate genes such as merozoite surface protein 1 (McBride et al., 1985) and apical membrane protein 1 (Polley et al., 2003; Cortés et al., 2003). These characteristics are indeed appealing and show the unique biological features of *Plasmodium* SERA. Here, we summarize current reports and our recent findings to understand the evolution of the SERA gene family.

2. SERA gene repertories in *Plasmodium* species

We refer briefly to the research history of the SERA multigene family: (i) the identification of SERA5 and the proteolytic processing of the protein; (ii) the discovery of the gene family following chromosome 2 sequencing of the *P. falciparum* genome; (iii) currently known SERA genes from different species; and (iv) the resulting analyses of the multigene family in various malaria parasites.

2.1 Research history of the SERA multigene family

SERA was first found in *P. falciparum* as an abundant, exported, soluble late-trophozite to schizont stage protein (Perrin et al., 1984). The protein, independently isolated by different groups, was described under various names as Pf140 (Perrin et al., 1984), p113 (Chulay et al., 1987), p126 (Deplace et al., 1987) or SERP (Knapp et at., 1989). All identified a gene with a long stretch of repeated serine residues in the N-terminal domain to which the family owes its name (Bzik et al., 1988). At the central domain, SERA possess a motif which align with two active site-determining regions of cysteine proteinases. The secreted protein was described to accumulate in the parasitophorous vacuole, and released into the culture medium at the time of schizont rupture. Notably, before the sequence of *P. falciparum* Chromosome 2 was opened, Knapp et al. (1991) discovered a SERA homolog (*serp H*) lacking the characteristic serine homopolymer, and subsequently Fox and Bzik (1994) reported SERA as one of three consecutive series of homologous genes. The complete genome sequence of *P. falciparum* revealed that SERA belongs to a multigene family (Gardner et al., 1998). The originally described SERA gene was renamed SERA5 according to

the gene arrangement order in *P. falciparum* (Aoki et al., 2002; Miller et al., 2002). It is interesting to note that SERA5 is the only member with repeated serine residues among nine gene members. The characteristic of the family is not the richness in serine residues but motifs that generate the framework of a cysteine protease. SERA homologs were identified in other *Plasmodium* species. Kiefer et al. (1996) found five SERA genes from another human parasite, *P. vivax* and Gor et al. (1998) identified three SERA genes from the rodent parasite of *P. vinckei*. Completed or ongoing genome projects of eight *Plasmodium* species: two human malaria parasites, *P. falciparum* and *P. vivax*; chimpanzee parasite *P. reichenowi*; macaque parasite *P. knowlesi*; three rodent parasites *P. berghei*, *P. yoelii* and *P. chabaudi*; and avian parasite *P. gallinaceum*, confirmed the gene organization and allowed phylogenetic analysis of the SERA gene family (Burgon et al., 2005; Arisue et al., 2007). In addition, Arisue et al. (2011) newly identified SERA genes in 11 *Plasmodium* species that further elaborate the genome organization of the gene family.

2.2 Processing of *P. falciparum* SERA5

The *in vitro* observation that *P. falciparum* SERA5 was released into the culture supernatant at the time of schizont rupture/melozoite release corresponds to its specific processing into several polypeptides. The full-length 120 kDa precursor accumulates in the parasitophorous vacuole during late trophozoite and schizont stages. As shown in Fig. 1., during the course of schizont rupture/merozoite release, SERA is proteolytically processed into a 47 kDa N-terminal (P47), a 50 kDa central (P50), an 18 kDa C-terminal (P18) and a 6 kDa domain (Delplace et al., 1987, 1988; Debrabant et al., 1992; Li et al., 2002a). The N-terminal P47 fragment is further processed into two 25 kDa fragments (P25n and P25c) in some allelic types (Li et al., 2002a). P47 is linked with the C-terminal P18 via disulfide bond that is localized at the merozoite surface that is localized at the merozoite surface (Delplace et al., 1987; Li et al., 2002a; Okitsu et al., 2007). The proteolytic processing is mediated by subtilisin-like serine protease subtilase 1 or SUB1 (Yeoh et al., 2007). Inhibition of SERA maturation blocks parasite egress from the host erythrocyte (Li et al., 2002b; Yeoh et al., 2007).

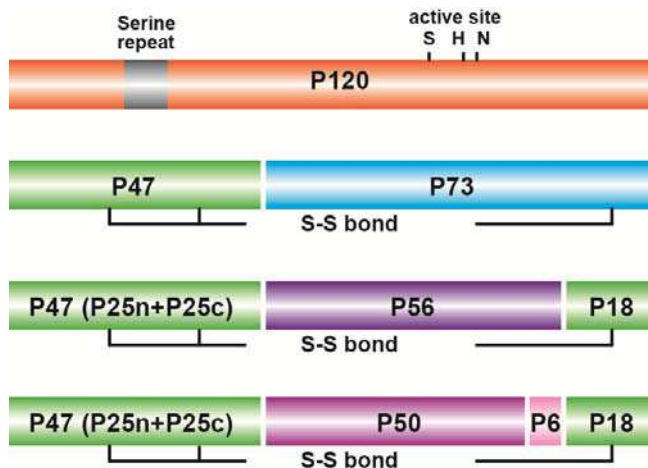


Fig. 1. Processing of *P. falciparum* SERA5 during parasite egress from host erythrocyte.

2.3 SERA genes in the database

The discovery of the SERA gene family in *P. falciparum* sparked the subsequent identification of a number of SERA genes in different *Plasmodium* species. Currently known SERA genes in the PlasmoDB database (http://plasmodb.org/plasmo/) are summarized in Table 1 and accession numbers of SERA genes found in the public database (NCBI, http://www.ncbi.nlm.nih.gov/) are summarized in Table 2.

We opted not to list SERA genes of *P. reichenowi* and *P. gallinaceum* in either Table, but for our analysis, their SERA gene sequences were assembled from various reads in the partial genome shotgun database of *Plasmodium* at The Sanger Institute. Blast programs in the following web sites were used to search SERA coding reads for: *P. reichenowi*: http://www.sanger.ac.uk/cgi-bin/blast/submitblast/p_reichenowi; and *P. gallinaceum*: http://www.sanger.ac.uk/cgi-bin/blast/submitblast/p_gallinaceum). Identified SERA gene sequences could be referred to in Arisue et al. (2007).

2.4 Characteristic features of the SERA multigene family

Almost all SERA genes were aligned in a tandem cluster between the conserved hypothetical protein gene and the iron-sulfur assembly protein gene. According to genetic background, SERA genes can be categorized into Groups I to IV (Arisue et al., 2007, 2011). Characteristic features of each Group are summarized in Fig. 2. Group I to Group III SERA genes possess the protease motif that includes an active site cysteine residue, in contrast to Group IV SERA genes where the cysteine residue is replaced by a serine (Bourgon et al., 2005; Arisue et al., 2007, 2011). The mRNA transcription and/or protein expression of Group I SERA genes were observed in the mosquito vector, while those of Group II to Group IV SERA genes were observed in the vertebrate host (Ali & Matuschewski, 2005; Arisue et al., 2007, 2011; Putrianti et al., 2010). The difference in the gene repertoire among species is due to the number of Group IV SERA genes. SERA gene repertories are summarized in Table 3.

Species (strain)	*P. vivax* (SalI)	*P. falciparum* (3D7)	*P. knowlesi* (H)	*P. berghei* (ANKA)	*P. yoelii* (17XNL)	*P. chabaudi* (AS)
SERA1	PVX_003850	PFB0360C	PKH_041200	PB000108.03.0	PY00291	PCAS_030730
SERA2	PVX_003845	PFB0355C	PKH_041210	PB107093.00.0	PY00292	PCAS_030720
SERA3	PVX_003840	PFB0350C	PKH_041230	PB000107.03.0	PY00292	PCAS_030710
SERA4	PVX_003835	PFB0345C	PKH_041250	PB000352.01.0	PY02062 PY00294	PCAS_030700
SERA5	PVX_003830	PFB0340C	PKH_041260	PB000649.01.0	PY02063	PCAS_030690
SERA6	PVX_003825	PFB0335C	PKH_041270			
SERA7	PVX_003820	PFB0330C				
SERA8	PVX_003810	PFB0325C				
SERA9	PVX_003805	PFI0135C				
SERA10	PVX_003800					
SERA11	PVX_003795					
SERA12	PVX_003790					

*The gene region of PY02062 and PY00294 was re-annotated and used as *P. yoelii* SERA4.

Table 1. GeneID of SERA genes in the PlasmoDB (http://plasmodb.org/plasmo/).

Species (strain)	Accession No.	Gene	Reference
P. vivax (SalI)	U51723	V_SERA1-V_SERA5	Kiefer et al., 1996
P. vinckei vinckei	U59860	$SERA_{vin}$-1	Gor et al., 1998
P. vinckei vinckei	U59861	$SERA_{vin}$-2	
P. vinckei vinckei	U59862	$SERA_{vin}$-3	
P. malariae (Kisii67)	AB576870	SERA1-SERA10	Arisue et al., 2011
P. ovale (Nigeria II)	AB576871	SERA1-SERA7	
P. cynomolgi (Mulligan)	AB576872	SERA1-SERA11	
P. fieldi (N-3)	AB576873	SERA1-SERA9	
P. simiovale	AB576874	SERA1-SERA9	
P. inui (Celebes)	AB576875	SERA1-SERA7	
P. hylobati (WAK)	AB576876	SERA1-SERA7	
P. coatneyi (CDC)	AB576877	SERA1-SERA7	
P. knowlesi (ATCC30158)	AB576878	SERA1-SERA6	
P. fragile (Hackeri)	AB576879	SERA1-SERA5	
P. gonderi	AB576880	SERA1-SERA4	
P. gonderi	AB576881	SERA5-SERA9	

Table 2. Accession numbers of SERA genes in the public NCBI database.

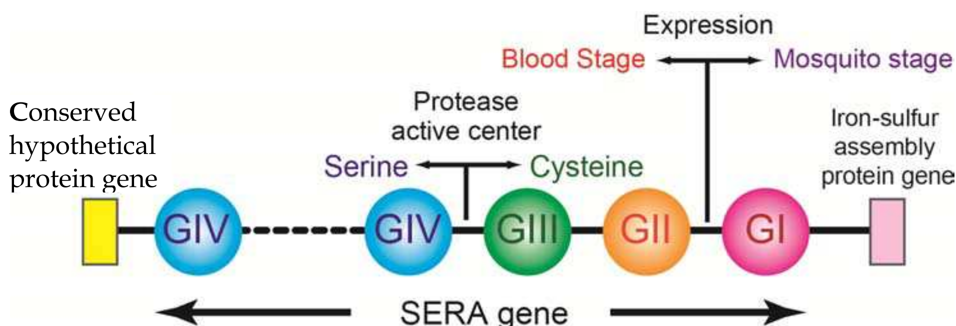


Fig. 2. Organization of *Plasmodium* SERA multigene family and the characteristic features of each SERA group.

Because of limited information for *P. reichenowi*, *P. gonderi* and *P. vinckei*, gene numbers are tentative for these species. The total number of SERA pseudogenes, truncated gene and gene fragments in Group IV SERA gene region are still undetermined in primate parasites (Arisue et al., 2011). Except for *P. gallinaceum*, all parasite species have one each of Group I to III SERA genes. The number of Group IV SERA genes remarkably increased in the primate parasite lineage. The bird parasite, *P. gallinaceum,* has the least number of SERA genes: two from Group I, and one from a branched common ancestor of Group II to IV SERA gene.

Species	Natural host	Number of SERA gene in each group				
		I	II	III	IV	Degenerate*
P. falciparum	human	1	1	1	6	0
P. vivax	human	1	1	1	9	2
P. malariae	human	1	1	1	7	3
P. ovale	human	1	1	1	4	1
P. knowlwsi	human/macaque	1	1	1	3	2
P. cynomolgi	macaque	1	1	1	8	3
P. coatneyi	macaque	1	1	1	4	3
P. fragile	macaque	1	1	1	2	3
P. simiovale	macaque	1	1	1	6	3
P. fieldi	macaque	1	1	1	6	3
P. inui	macaque	1	1	1	4	5
P. hylobati	gibbon	1	1	1	4	1
P. berghei	rat	1	1	1	2	0
P. yoelii	rat	1	1	1	2	0
P. chabaudi	rat	1	1	1	2	0
P. gallinaceum	bird	2			1	0
P. gonderi	mangabey, guenon	1	1	1	6	?
P. reichenowi	chimpanzee	1	1	1	5?	?
P. vinckei	rat	?	1	1	1?	?

Table 3. The number of SERA genes that belong to each group from several *Plasmodium* species. 'Degenerate'(*) denotes defective gene copies, *i.e.*, the total number of pseudogenes, truncated gene and gene fragments found.

2.5 Primary structure of SERA molecules and genes

Schematic representation of SERA gene structure is shown in Fig. 3A. Group I SERA genes code for around 700 amino acids. They share a similar six exon and five intron structure, except for *P. falciparum* SERA8 and *P. vivax* SERA12 that both lack one intron. Group II to Group IV SERA genes code for about 1000 amino acid residues, and similar to Group I, share a common four exon/three intron structure with few exceptions. All SERA genes have the structural context of cysteine proteinases, however, it is important to note that the canonical Cys His Asn triad of the active proteinase is not present in all. The relatively small number of amino acid residues in Group I SERA resulted to a shorter N-terminal region when compared to Group II to IV SERA. Multiple amino acid sequence alignments revealed the consensus primary structure of SERA, which consists of six putative domains shown in Fig. 3B.

Amino acid sequences of the putative pro-enzyme and enzyme domains are remarkably conserved, but extensive sequence variations are found in variable domains 1 and 2. In the C-terminal cysteine rich conserved domain, seven cysteine residues are perfectly conserved in all SERA genes.

The pro-enzyme and enzyme domains of *P. falciparum* SERA5 was identified by functional genetic and structural analyses (Hodder et al., 2003, 2009). These domains, corresponding to

P50 in Fig. 1, are flanked by the reported SUB1 cleavage sites (Yeoh et al., 2007). The consensus sequence of the cleavage site is (Val/Leu/Ile)-Xaa-(Gly/Ala)-Paa, in which Xaa is any amino acid residue and Paa is any non-polar residue except for Leu (Yeoh et al., 2007). This consensus sequence is well conserved with slight modifications in all Group II to IV *Plasmodium* SERA genes which we have analyzed. *In vitro*, *P. falciparum* SERA4 (Group IV) and SERA6 (Group II) were cleaved by recombinant PfSUB1 (Yeoh et al., 2007). Peculiarly, Group I SERA genes lack most of the N-terminal variable domain 1 and SUB1 cleavage sites.

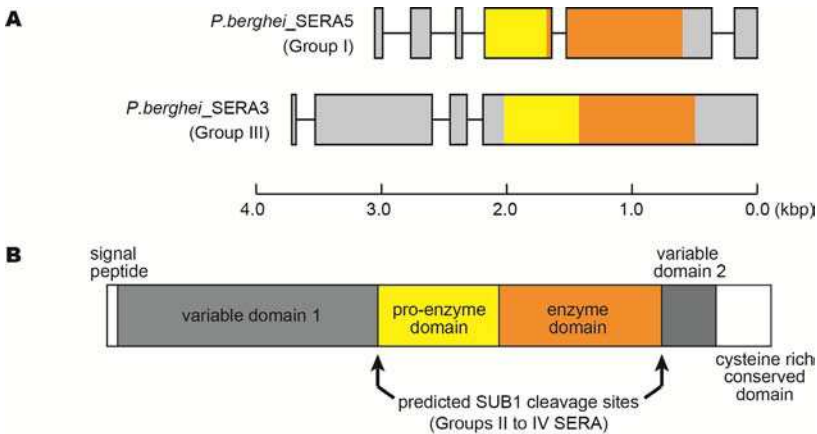


Fig. 3. Schematic representation of the SERA gene structure (A) and their putative domain organization (B).

3. Phylogenetic relationships of SERA genes

The categorization of SERA genes into four groups, Group I to IV, was based on phylogenetic analysis (Arisue et al., 2007, 2011). SERA amino acid sequences from 18 *Plasmodium* species were aligned using CLUSTAL W program (http://clustalw.ddbj.nig.ac.jp/top-j.html) under default options with manual corrections. Unambiguously aligned amino acid positions corresponding to the putative pro-enzyme domain, enzyme domain and cysteine rich conserved domain were selected and used for the phylogenetic analysis. Maximum likelihood tree was inferred using the PROML program in PHYLIP version 3.69 (Felsenstein, 1996). Except for the number of genes and number of amino acid sequences included in the analysis, the same method was used to infer the phylogenetic tree shown in Fig. 4 and 5.

A simplified maximum likelihood tree inferred from 134 SERA genes with 392 amino acid positions is shown in Fig. 4. Bootstrap proportion values were placed only on the common ancestor branch of each group. The monophyletic grouping of Group I SERA genes was supported with a bootstrap value of 100%. The long internal branch separating Group I from Groups II to IV suggests that the root of the tree is located on the branch leading to the common ancestor of Group I SERA genes. It is thus likely that Group I genes have appeared early in the evolution of the SERA gene family. *P. gallinaceum* SERA1 branches at the common ancestor of Group II to IV, suggesting that gene duplication events which produced Groups II, III and IV likely occurred after the divergence of *P. gallinaceum* from the common ancestral lineage of *Plasmodium*.

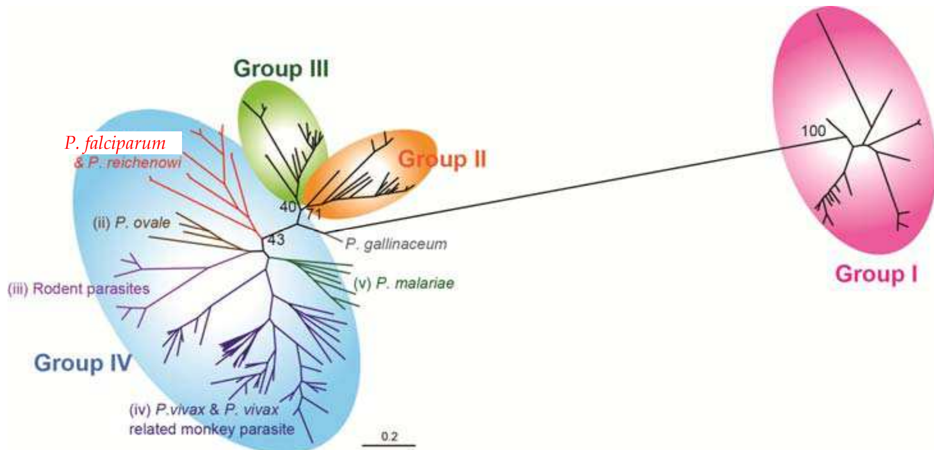


Fig. 4. Phylogenetic tree based on the alignment of 134 SERA genes from 18 *Plasmodium* species.

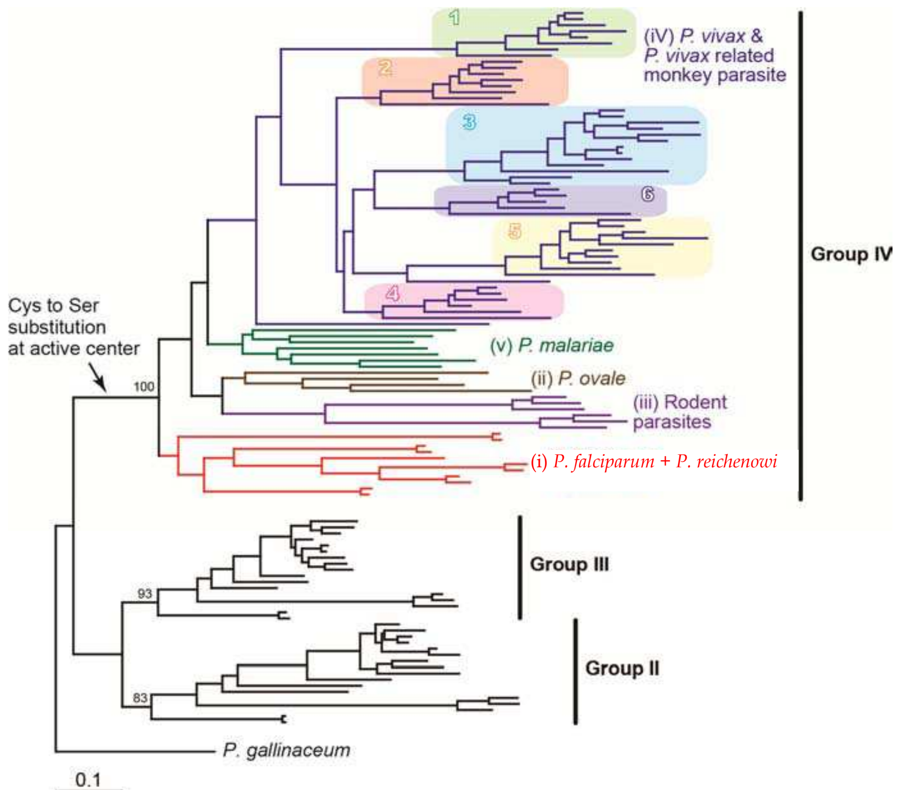


Fig. 5. Phylogenetic tree based on the alignment of 115 SERA genes from 18 *Plasmodium* species.

Group IV SERA genes which diverged after the cysteine to serine substitution in the catalytic site were further divided to five monophyletic sub-Groups: (i) *P. falciparum* and *P. reichenowi*, (ii) *P. ovale*, (iii) rodent *Plasmodium* species, (iv) *P. vivax* and *P. vivax*-related monkey parasite species, and (v) *P. malariae*. These indicate that genes have duplicated independently in each of the sub-group lineages. To increase the resolution of the tree, the long branched Group I genes were excluded from the dataset and the maximum likelihood tree was re-constructed from 115 SERA genes categorized into Group II to Group IV. The resultant tree is shown in Fig. 5.

Interestingly, Group IV SERA genes of *P. vivax* and *P. vivax*-related monkey parasites (10 species) were further categorized into six orthologous gene groups, namely, Clade 1 to Clade 6; and each clade has 5 (Clade 6) to 10 (Clade 5) parasite species. The number of SERA genes analyzed varied from 5 (*P. fragile*) to 12 (*P. vivax*). This does suggest that a common ancestor of *P. vivax* and related monkey malaria parasites had at least 6 SERA genes of Group IV; and that gene duplications and gene deletions occurred in each lineage. Orthologous relationships between SERA gene members and their relative arrangement are shown in Fig. 6.

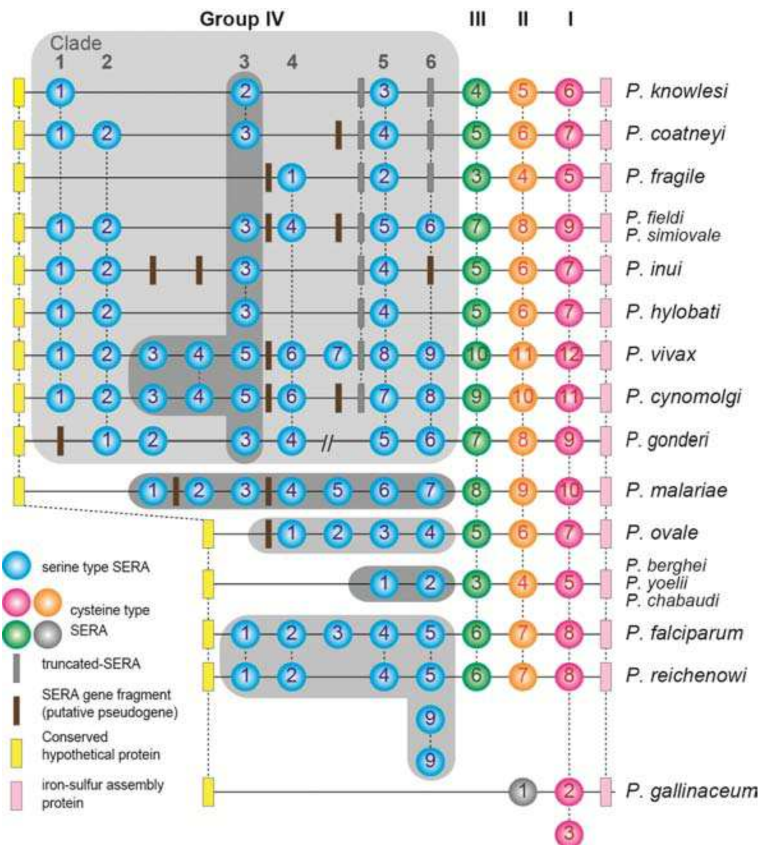


Fig. 6. Organization and phylogenetic relationship of SERA gene in 18 *Plasmodium* species.

In addition to several gene members characterizing Group IV, there are multiple SERA gene fragments and pseudogenes containing multiple stop codons. Taken together, these extensive gene duplications, gene deletions as well as pseudogenization/truncation are evident only in the serine type SERA gene (Group IV) of *P. vivax* and related monkey malaria parasites.

4. Transcription analyses of SERA genes

Transcription analyses revealed, likewise, some discordance among *Plasmodium* species. Transcription profile of the SERA gene family was analyzed first by Aoki et al. (2002) in *P. falciparum*. Genes were most actively transcribed at the late trophozoite to schizont stages of the parasite with SERA5 predominantly transcribed among the family (Fig. 7.).

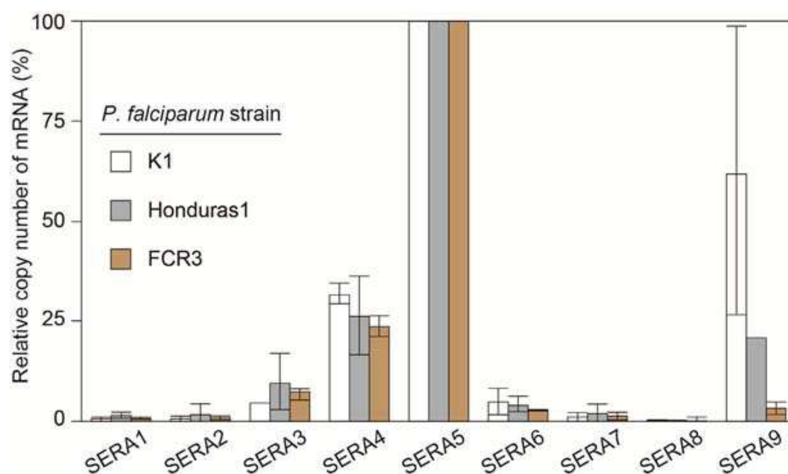


Fig. 7. The relative abundance of mRNA for each *P. falciparum* SERA gene during late trophozoite to schizont stages of the parasite.

Similar to *P. falciparum*, multiple number of SERA genes were transcribed at late trophozoite to schizont stages when transcription analysis was done for the SERA gene family in other *Plasmodium* parasites: human parasite *P. vivax* (Palacpac et al., 2006); rodent parasite *P. berghei* (Arisue et al., 2011); and three monkey parasites *P. knowlesi*, *P. cynomolgi* and *P. coatneyi* (Arisue et a.l, 2011). A representative summary of SERA gene transcription analysis is shown in Fig. 8. In malaria parasite species infecting humans, one of Group IV SERAs of *P. falciparum* (SERA5) and *P. vivax* (SERA4) showed the highest transcription level among other gene members. In the rodent malaria parasite *P. berghei*, Group III SERA gene (SERA3) was predominantly expressed. In three monkey malaria parasites, the abundantly expressed genes are members of Group IV Clade 3: SERA3 and SERA5, in *P. cynomolgi*; SERA3 in *P. coatneyi*; and SERA2 in *P. knowlesi*. These results show that SERA genes were differently expressed between rodent and primate parasites. Based on the malaria parasite mitochondrial genome, *P. falciparum* belongs to the primate parasite group 1 lineage whereas *P. vivax* and the three macaque parasites belong to primate parasite lineage 2. Phylogenetic analysis showed that these two lineages are not closely related; and the rodent parasite lineage is positioned between them (Hayakawa et al., 2008). Note, however, that

both primate lineages showed similar transcription pattern of SERA gene which might suggest a possible relationship between SERA function and host specificity.

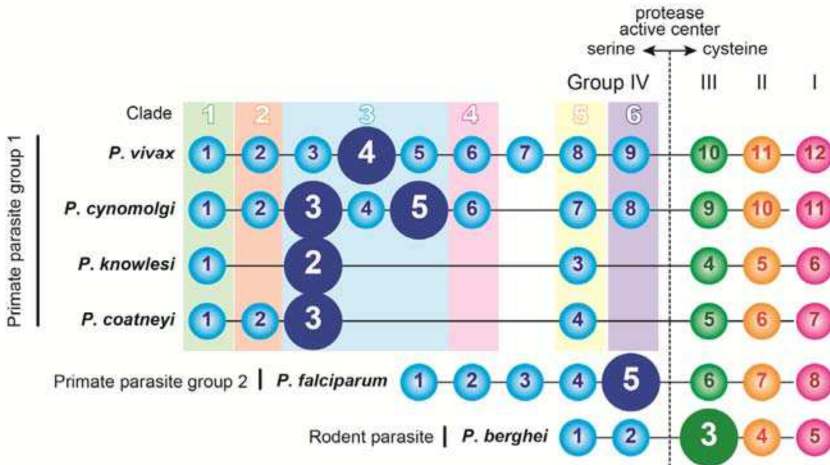


Fig. 8. Representation summary of SERA gene transcription analyses in primate and rodent *Plasmodium* spp. Abundantly transcribed SERA genes are differentiated using enlarged solid circles.

5. Duplication in the multigene family

In general, duplicated genes undergo either (i) concerted evolution or (ii) birth-and-death evolution (Nei & Rooney, 2005). In the concerted evolution model, all members of a gene family evolve as a unit rather than independently. When a mutation occurs in a gene, mutation spreads to all other gene members by unequal crossover or gene conversion. As a result, all members of the gene family show identical sequence to each other. The evolution of rRNA multigene families in vertebrates is a classic example of concerted evolution. Analysis of MHC genes in mammals (Hughes & Nei, 1989; Nei et al., 1997; Nei & Hughes, 1992), other immune system related genes (Hughes & Nei, 1990; Ota & Nei, 1994) and disease-related genes (Zhang et al., 2000) show a quite different evolutionary pattern. The birth-and-death evolution model was proposed to explain differential duplication/independent diversification processes that result to subsequent loss or maintenance of genes in a multigene family. Thus, some duplicated genes are maintained in the genome for a long time while others are deleted or became pseudogenes through deleterious mutations. This model applies to rRNAs of *Plasmodium* species in marked contrast to the concerted evolution of rRNAs in most organisms. The model aptly explains the observation that rRNA genes in *Plasmodium* were structurally and functionally distinct (Rooney, 2004; Nishimoto et al., 2008).

The observed gene duplication and gene deletion found in the *Plasmodium* SERA genes are clearly in concordance with the birth-and-death model, although traits of gene conversion are detected in a few of Group IV SERA genes. The birth-and-death model has, likewise, been recently proposed for gene duplication/gene deletion of merozoite surface protein 7, an immune target parasite surface antigen gene (Garzón-Ospina et al., 2010). It, thus, seem

most probable that diversification of *Plasmodium* SERA multigene family was also driven by the birth-and-death evolution. Inferred gene duplication events in the evolution of the *Plasmodium* SERA gene family are shown in Fig. 9.

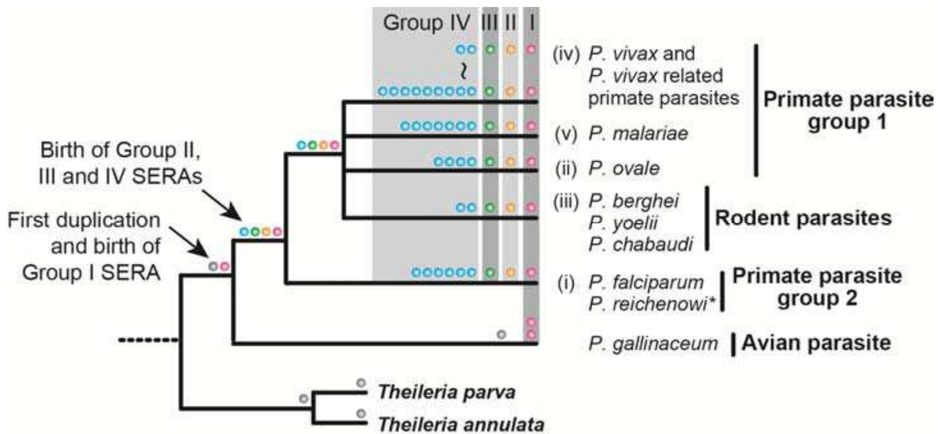


Fig. 9. Inferred gene duplication events in the evolution of the *Plasmodium* SERA gene family. Asterisk in *P. reichenowi* denotes that the SERA gene number in this species is tentative.

Theileria is the only other genus to possess a SERA homolog gene. The apicomplexan parasite is closely related to *Plasmodium*. Sequence similarity search against *Theileria* genome database at The Sanger Institute identified a single gene from both *T. parva* and *T. annulata* which has similarity with cysteine-type SERA (Arisue et al., 2007; McCoubrie et al., 2007). Based on Fig. 9, as all *Plasmodium* species have multiple SERA genes, we likely infer that the first duplication occurred at the common ancestor lineage of *Plasmodium*. Because every *Plasmodium* species has Group I SERA gene, the first duplication event is from Group I SERA gene. The duplication events which gave rise to Group II and IV SERA genes occurred after the divergence of *P. gallinaceum* from the branch leading to a common ancestral species of other *Plasmodium* species since *P. gallinaceum* has no Group II to IV SERA gene. The rest of the 17 *Plasmodium* species might have diverged into five lineages of (i) *P. falciparum* and *P. reichenowi*, (ii) *P. ovale*, (iii) rodent *Plasmodium* species, (iv) *P. vivax* and *P. vivax*-related monkey parasite species, and (v) *P. malariae,* and duplications of Group IV SERA genes occurred independently on each lineage. In addition, gene deletions as well as pseudogenization/truncation occurred frequently in *P. vivax* and *P. vivax*-related primate parasite lineage.

6. Conclusion and open issues in SERA study

Multigene families are believed to provide an organism with a set of related genes that allow fine tuning of its biological function with possibly different temporal or topologic expression patterns. SERA gene duplications during *Plasmodium* evolution generated four types of SERA genes: Group I to Group IV. The speculated function of SERA during the parasite life cycle is summarized in Fig. 10.

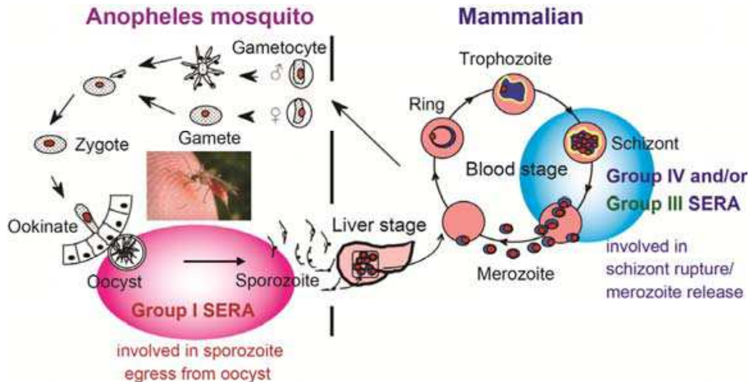


Fig. 10. The life cycle of the malaria parasite and inferred role(s) of SERA.

Group I SERA gene, bearing the canonical cysteine, was shown to be transcribed in the oocyst stage and its gene product was a protease required for sporozoite egress from the oocyst (Aly & Matuschewski, 2005). Group III SERA gene was suggested to play an essential role during schizont rupture/merozoite release in the mammalian host (Yeoh et al., 2007; Arastu-Kapur et al., 2008; Putrianti et al., 2010). Group IV SERA genes bear the characteristic replacement of active-site cysteine to a serine residue. Perhaps importantly, only mammalian parasites have Group IV SERA gene; and Group IV SERA gene of primate parasite was suggested to play an essential role in schizont rupture/merozoite release together with Group III SERA gene (Yeoh et al., 2007; Arisue et al., 2011). The duplication of Group IV SERA gene occurred particularly frequent in two evolutionarily distinct primate lineages and it is intriguing to assume that duplications of SERA genes were associated with host range expansion.

The study of the SERA gene family points to its unique features reinforcing the importance of investigating other uncharacterized gene families of *Plasmodium* to further understand the evolutionary history and biology of this harmful parasite. Many questions still remain in the analysis of SERA. SERA genes are thought to be subject to birth-and-death evolution, and thus, a pattern of interspecific gene clustering is expected to characterize the SERA family whereby functional genes are maintained in the genome for a long time and others are deleted or become non-functional. Group I and Group III SERA genes are highly conserved in *Plasmodium* species. For Group II SERA genes, although maintained among *Plasmodium* species with significant sequence similarity, no function has yet been predicted. Gene disruption studies with Group II SERA gene of *P. berghei* showed no apparent phenotypic change (Arisue et al., unpublished data). Group II is similar to Group I and Group III in being a cysteine-type SERA gene which has been suggested to have proteolytic activity to cleave host membrane structure (Aly & Matuschewski, 2005; Yoeh et al., 2007). The papain-like cysteine protease motif in its amino acid sequence suggests the possibility that Group II SERA act as a protease sometime in the parasite life cycle. Parasite egress from the host cell is an important process that remains poorly understood.

P. falciparum SERA5 is a vaccine candidate molecule now on clinical trial in Uganda (Horii et al., 2010). Serum antibodies against the N-terminal domain of *P. falciparum* SERA5 in individuals living in malaria endemic area protect infants from clinical malaria and inhibit *in vitro* parasite growth (Okech et al., 2001, 2006; Aoki et al., 2002; Horii et al., 2010). During

blood stage growth, all SERA gene family member of *P. falciparum* are transcribed most actively at trophozoite and schizont stages. SERA5 is the most abundantly expressed gene family member, with expression levels estimated to be approximately 0.5-1.5% of the whole mRNA at schizont stage (Aoki et al, 2002). However, sero-positivity rate against the N-terminal domain of *P. falciparum* SERA5 was observed to be relatively low (Fig. 11.; Aoki et al., 2002; Horii et al., 2010).

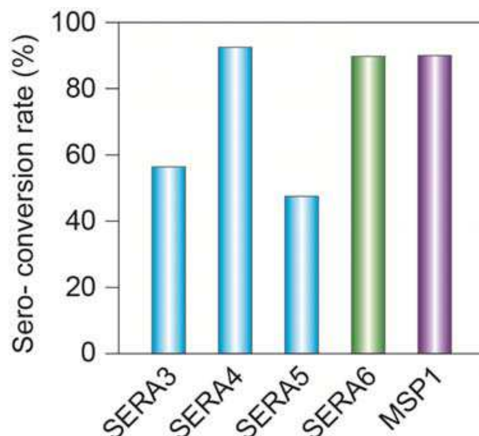


Fig. 11. Relative sero-conversion rates of *P. falciparum* SERA3 to SERA6 and merozoite surface protein 1 (MSP1) in a malaria endemic area of Uganda.

Since the SERA gene family does not show antigenic variation to evade host immune response (Fig. 7.), there may possibly be another mechanism of host parasite evasion/molecular mimicry/interference or competition.

The host range or host specificity of *Plasmodium* is believed to be restricted, although, primate malaria parasites generally infect multiple hosts. For example, it has been reported that, *P. knowlesi* and *P. cynomolgi* have the ability to infect a wide variety of macaques and human (Coatney et al., 1971); additionally, two human parasites *P. malariae* and *P. ovale* have been detected in chimpanzees (Hayakawa et al., 2009; Duval et al., 2009). It may be probable that duplications of Group IV SERA genes that occurred frequently in both primate parasite lineages may be associated with host range expansions. To date, no experimental support lends credence to this speculation.

As described above, the molecular function of SERA genes in each group, the relationship of immune evasion mechanism and the SERA gene family, and the association of host range with Group IV SERA genes remain important issues that needs to be addressed. The importance of SERA genes in parasite egress and their role in host-parasite interactions serve to propel further studies in understanding this multigene family.

7. Acknowledgment

This work was supported by **KAKENHI (18073013 and 20390120)** from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

8. References

Aly, ASI. & Matuschewski K. (2005) A malarial cysteine protease is necessary for *Plasmodium* sporozoite egress from oocysts, *The Journal of Experimental Medicine*, Vol. 202, No. 2, pp. 225-230.

Aoki, S.; Li J, Itagaki S et al. (2002) Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of *Plasmodium falciparum*, and the acquired antibody titers correlate with serum inhibition of the parasite growth, *The Journal of Biological Chemistry*, Vol. 277, No 49, pp. 47533-47540.

Arastu-Kapur, S.; Ponder EL, Fonovic UP et al. (2008) Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*, *Nature Chemical Biology*, Vol. 4, No. 3, pp. 203-210.

Arisue, N.; Hirai M, Arai M et al. (2007) Phylogeny and evolution of the SERA multigene family in the genus *Plasmodium*, *Journal of Molecular Evolution*, Vol. 65, No. 1, pp. 82-91.

Arisue, N.; Kawai S, Hirai M et al. (2011) Clues to evolution of the SERA multigene family in 18 *Plasmodium* species, *PLoS One*, Vol. 6, No. 3, e17775.

Bourgon, R.; Delorenzi M, Sargeant T et al. (2004) The serine repeat antigen (SERA) gene family phylogeny in *Plasmodium*: the impact of GC content and reconciliation of gene and species trees, *Molecular Biology and Evolution*, Vol. 21, No. 11, pp. 2161-2171.

Bzik, DJ.; Li WB, Horii T et al. (1988) Amino acid sequence of the serine-repeat antigen (SERA) of *Plasmodium falciparum* determined from cloned cDNA, *Molecular and Biochemical Parasitology*, Vol. 30, No. 3, pp. 279-288.

Carlton, JM.; Angiuoli SV, Suh BB et al. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*, *Nature*, Vol. 419, No. 6906, pp. 512-519.

Carlton, JM.; Adams JH, Silva JC et al. (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*, *Nature*, Vol. 455, No. 7214, pp. 757-763.

Chulay, JD.; Lyon JA, Haynes JD et al. (1987) Monoclonal antibody characterization of *Plasmodium falciparum* antigens in immune complexes formed when schizonts rupture in the presence of immune serum. *The Journal of Immunology*, Vol. 139, No. 8, pp. 2768–2774.

Cortés, A.; Mellombo M, Mueller I et al. (2003) Geographical structure of diversity and differences between symptomatic and asymptomatic infections for *Plasmodium falciparum* vaccine candidate AMA1. *Infection and Immunity*, Vol. 71, No. 3, pp. 1416-1426.

Coatney, GR.; Collins WE, Warren M et al. (1971) *The Primate Malarias*, US Government Printing Office, Washington DC, USA.

Debrabant, A.; Maes P, Delplace P et al. (1992) Intramolecular mapping of *Plasmodium falciparum* P126 proteolytic fragments by N-terminal amino acid sequencing, *Molecular and Biochemical Parasitology*, Vol. 53, No. 1-2, pp. 89-95.

Delplace, P.; Bhatia A, Cagnard M et al. (1988) Protein p126: a parasitophorous vacuole antigen associated with the release of *Plasmodium falciparum* merozoites, *Biology of the Cell*, Vol. 64, No. 2, pp. 215-221.

Delplace, P.; Fortier B, Tronchin G et al. (1987) Localization, biosynthesis, processing and isolation of a major 126 kDa antigen of the parasitophorous vacuole of *Plasmodium falciparum*, *Molecular and Biochemical Parasitology*, Vol. 23, No. 3, pp. 193-201.

Duval, L.; Nerrienet E, Rousset D et al. (2009) Chimpanzee malaria parasites related to *Plasmodium ovale* in Africa, *PLoS One*, Vol. 4, No. 5, e5520.

Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods, *Methods in Enzymology*, Vol. 266, pp. 418-427.

Fox, BA. & Bzik DJ. (1994) Analysis of stage-specific transcripts of the *Plasmodium falciparum* serine repeat antigen (SERA) gene and transcription from the SERA locus, *Molecular and Biochemical Parasitology*, Vol. 68, No. 1, pp. 133-144.

Fox, BA.; Xing-Li P, Suzue K et al. (1997) *Plasmodium falciparum*: an epitope within a highly conserved region of the 47-kDa amino-terminal domain of the serine repeat antigen is a target of parasite-inhibitory antibodies, *Experimental Parasitology*, Vol. 85, No. 2, pp. 121-134.

Gardner, MJ.; Tettelin H, Carucci DJ et al. (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*, *Science*, Vol. 282, No. 5391, pp. 1126-1132.

Gardner, MJ.; Hall N, Fung E et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature*, Vol. 419, No. 6906, pp. 498-511.

Garzón-Ospina, D.; Cadavid LF & Patarroyo MA. (2010) Differential expansion of the merozoite surface protein (msp)-7 gene family in *Plasmodium* species under a birth-and-death model of evolution, *Molecular Phylogenetics and Evolution*, Vol. 55, No. 2, pp. 399-408.

Gor, DO.; Li AC & Rosenthal PJ. (1998) Protective immune responses against protease-like antigens of the murine malaria parasite *Plasmodium vinckei*, Vaccine, Vol 16, No. 11-12, pp. 1193-1202.

Hayakawa, T.; Culleton R, Otani H et al. (2008) Big bang in the evolution of extant malaria parasites, *Molecular Biology and Evolution*, Vol. 25, No. 10, pp. 2233-2239.

Hayakawa, T.; Arisue N, Udono T et al. (2009) Identification of *Plasmodium malariae*, a human malaria parasite, in imported chimpanzees, *PLoS One*, Vol. 4, No. 10, e7412.

Hodder, AN.; Drew DR, Epa VC et al. (2003) Enzymic, phylogenetic, and structural characterization of the unusual papain-like protease domain of *Plasmodium falciparum* SERA5, *The Journal of Biological Chemistry*, Vol. 278, No. 48, pp. 48169-48177.

Hodder, AN.; Malby RL, Clarke OB et al. (2009) Structural insights into the protease-like antigen *Plasmodium falciparum* SERA5 and its noncanonical active-site serine, *Journal of Molecular Biology*, Vol. 392, No. 1, pp. 154-165.

Horii, T.; Shirai H, Li J et al. (2010) Evidences of protection against blood-stage infection of *Plasmodium falciparum* by the novel protein vaccine SE36, *Parasitology International*, Vol. 59, No. 3, pp. 380-386.

Hughes, AL. & Nei M. (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals, *Molecular Biology and Evolution*, Vol. 6, No. 6, pp. 559-579.

Hughes, AL. & Nei M. (1990) Evolutionary relationships of class II major-histocompatibility-complex genes in mammals, *Molecular Biology and Evolution*, Vol. 7, No. 6, pp. 491-514.

Janssen, CS.; Phillips RS, Turner CM et al. (2004) *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites, *Nucleic Acids Research*, Vol. 32, No. 19, pp. 5712-5720.

Kiefer, MC.; Crawford KA, Boley LJ et al. (1996) Identification and cloning of a locus of serine repeat antigen (sera)-related genes from *Plasmodium vivax*, *Molecular and Biochemical Parasitology*, Vol. 78, No. 1-2, pp. 55-65.

Knapp, B.; Hundt E, Nau U et al. (1989) Molecular cloning, genomic structure and localization of a blood stage antigen of *Plasmodium falciparum* characterized by a serine stretch. *Molecular and Biochemical Parasitology*, Vol. 32, No. 1, pp. 73-83.

Knapp, B.; Nau U, Hundt E et al. (1991) A new blood stage antigen of *Plasmodium falciparum* highly homologous to the serine-stretch protein SERP, *Molecular and Biochemical Parasitology*, Vol. 44, No. 1, pp. 1-13.

Levin, ND. (1988) *The protozoan phylum Apicomplexa-volume II*, CRC Press, ISBN 0-8493-4654-1, Florida, USA.

Li, J.; Mitamura T, Fox BA et al. (2002a) Differential localization of processed fragments of *Plasmodium falciparum* serine repeat antigen and further processing of its N-terminal 47 kDa fragment, *Parasitology International*, Vol. 51, No. 4, pp. 343-352.

Li J.; Matsuoka H, Mitamura T et al. (2002b) Characterization of proteases involved in the processing of *Plasmodium falciparum* serine repeat antigen (SERA), *Molecular and Biochemical Parasitology*, Vol. 120, No. 2, pp. 177-186.

Liu, Q.; Ferreira MU, Ndawi BT et al. (2000) Sequence diversity of serine repeat antigen gene exon II of *Plasmodium falciparum* in worldwide collected wild isolates. *Southeast Asian Journal of Tropical Medicine and Public Health,* Vol. 31, No. 4, pp. 808-817.

McBride, JS.; Newbold CI & Anand, R. (1985) Polymorphism of a high molecular weight schizont antigen of the human malaria parasite *Plasmodium falciparum*, *The Journal of Experimental Medicine,* Vol. 161, No. 1, pp. 160-180.

McCoubrie, JE.; Miller SK, Sargeant T et al. (2007) Evidence for a common role for the serine-type *Plasmodium falciparum* serine repeat antigen proteases: Implications for vaccine and drug design, *Infection and Immunity*, Vol. 75, No. 12, pp. 5565-5574.

Miller, SK.; Good RT, Drew DR et al. (2002) A subset of *Plasmodium falciparum* SERA genes are expressed and appear to play an important role in the erythrocytic cycle, *The Journal of Biological Chemistry*, Vol. 277, No. 49, pp. 47524-47532.

Morimatsu, K.; Morikawa T, Tanabe K et al. (1997) Sequence diversity in the amino-terminal 47 kDa fragment of the *Plasmodium falciparum* serine repeat antigen. *Molecular and Biochemical Parasitology*, Vol. 86, No. 2, pp. 249-254.

Nei, M. & Hughes AL. (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: *11th Histocompatibility Workshop and Conference*, Tsuji, K., Aizawa, M., Sasazuki, T., editors. Vol. 2, pp. 27-38, Oxford University Press, ISBN 0-19-262217-X, Oxford, UK.

Nei, M.; Gu, X. & Sitnikova T. (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 94, No. 15, pp. 7799-7806.

Nei, M. & Rooney AP. (2005) Concerted and birth-and death evolution of multigene families, *Annual Review of Genetics*, Vol. 39, pp. 121-152.

Nishimoto, Y.; Arisue N, Kawai S et al. (2008) Evolution and phylogeny of the heterogeneous cytosolic SSU rRNA genes in the genus *Plasmodium*, *Molecular Phylogenetics and Evolution*, Vol. 47, No. 1, pp. 45-53.

Okech, BA.; Nalunkuma A, Okello D et al. (2001) Natural human immunoglobulin G subclass responses to *Plasmodium falciparum* serine repeat antigen in Uganda, *The American Journal of Tropical Medicine and Hygiene*, Vol. 65, No. 6, pp. 912-917.

Okech, B.; Mujuzi G, Ogwal A et al. (2006) High titers of IgG antibodies against *Plasmodium falciparum* serine repeat antigen 5 (SERA5) are associated with protection against severe malaria in Ugandan children, *The American Journal of Tropical Medicine and Hygiene*, Vol. 74, No. 2, pp. 191-197.

Okitsu, SL.; Boato F, Mueller MS et al. (2007) Antibodies elicited by a virosomally formulated *Plasmodium falciparum* serine repeat antigen-5 derived peptide detect the processed 47 kDa fragment both in sporozoites and merozoites, *Peptides*, Vol. 28, No. 10, pp. 2051-2060.

Ota, T. & Nei M. (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular Biology and Evolution*, Vol. 11, No. 3, pp. 469-482.

Pain, A.; Bohme U, Berry AE et al. (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*, *Nature*, Vol. 455, No. 7214, pp. 799-803.

Palacpac, NM.; Leung BW, Arisue N et al. (2006) *Plasmodium vivax* serine repeat antigen (SERA) multigene family exhibits similar expression patterns in independent infections, *Molecular and Biochemical Parasitology*, Vol. 150, No. 2, pp. 353-358.

Perrin, LH.; Merkli B, Loche M et al., (1984) Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages, *The Journal of Experimental Medicine*, Vol. 160, No. 2, pp. 441-451.

Polley, SD.; Chokejindachai W & Conway DJ. (2003) Allele frequency-based analyses robustly map sequence sites under balancing selection in a malaria vaccine candidate antigen, *Genetics*, Vol. 165, No. 2, pp. 555-561.

Putrianti, ED.; Schmidt-Christensen A, Arnold I et al. (2010) The *Plasmodium* serine-type SERA proteases display distinct expression patterns and non-essential in vivo roles during life cycle progression of the malaria parasite, *Cellular Microbiology*, Vol. 12, No. 6, pp. 725-739.

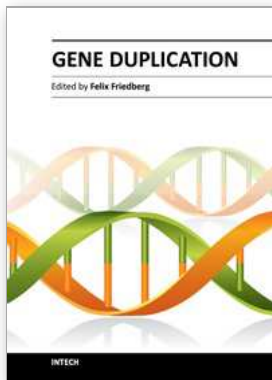
Rooney, AP. (2004) Mechanism underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in apicomplexans, *Molecular Biology and Evolution*, Vol. 21, No. 9, pp. 1704-1711.

Schmidt-Christensen, A.; Sturm A, Horstmann S et al. (2008) Expression and processing of *Plasmodium berghei* SERA3 during liver stages, *Cellular Microbiology*, Vol. 10, No. 8, pp. 1723-1734.

WHO (2010) *World Malaria Report 2010*, WHO Press, ISBN 978 92 4 156410 6, Geneva, Switzerland.

Yeoh, S.; O'Donnell RA, Koussis K et al. (2007) Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes, *Cell*, Vol. 131, No. 6, pp. 1072-1083.

Zhang J.; Dyer KD. & Rosenberg HF. (2000) Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 9, pp. 4701-4706.



Gene Duplication

Edited by Prof. Felix Friedberg

ISBN 978-953-307-387-3
Hard cover, 400 pages
Publisher InTech
Published online 17, October, 2011
Published in print edition October, 2011

The book Gene Duplication consists of 21 chapters divided in 3 parts: General Aspects, A Look at Some Gene Families and Examining Bundles of Genes. The importance of the study of Gene Duplication stems from the realization that the dynamic process of duplication is the "sine qua non" underlying the evolution of all living matter. Genes may be altered before or after the duplication process thereby undergoing neofunctionalization, thus creating in time new organisms which populate the Earth.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Nobuko Arisue, Nirianne M. Q. Palacpac, Kazuyuki Tanabe and Toshihiro Horii (2011). Clues to Evolution of the SERA Multigene Family in the Genus Plasmodium, Gene Duplication, Prof. Felix Friedberg (Ed.), ISBN: 978-953-307-387-3, InTech, Available from: http://www.intechopen.com/books/gene-duplication/clues-to-evolution-of-the-sera-multigene-family-in-the-genus-plasmodium

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.