Plasmodium falciparum var gene expression dynamics and its relevance in malaria disease in children from Papua New Guinea

INAUGURALDISSERTATION

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Mirjam Elisabeth Kästli
aus St. Margrethen, SG

Basel, Januar 2005

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel auf Antrag der Herren Prof. Dr. Hans-Peter Beck, Prof. Dr. Niklaus Weiß und Prof. Dr. Tom Bickle.

Basel, den 17. Januar 2005

Prof. Dr. Hans-Jakob Wirz

Dekan

To my family,
Elisabeth, Christina and Rolf
Kästli

Table of contents

Acknowledgements

First of all, I am deeply grateful to Hans-Peter Beck, my supervisor, for giving me the opportunity to carry out my Ph.D. in his group and for giving me the chance to experience malaria far from the lab in the real world - more than anything else, this broadened my mind. A big thank you for his continuous support in scientific questions but also in organisational matters, be it in visa hunting or in taking reddish CV pictures. A big thank you for his enthusiasm, especially during never ending cloning times and doubts, and above all, for his constant open door of which I often made use for numerous questions and spiriting discussions.

This work would not have been possible without the cooperation of the children and parents from Papua New Guinea who took part in our projects and who always welcomed us with warmth and interest and let us take part in their lives for this short period of time. A big thank you and I would not want to have missed this enriching experience.

I wish to thank Prof. N. Weiss and Prof. T. Bickle for being part of my thesis committee and for their support. Since I understand the inconvenience and time involved in processing a thesis and writing a report, I highly appreciate the effort.

I am deeply grateful to the whole Molecular Parasitology group who always contributed so much to a pleasant and lively atmosphere also beyond the lab - to Ingrid for her intellectual support and valuable ideas, to our *var* gene gang for exciting new ideas on how A, B and C cohabit: Matze (my consultant in scientific, sportive and delicate matters), Sebi (my worthy successor in RNase hunting), Alex, Dieter and Selina (we provided final evidence that it is possible to get lost in Olten); to Jutta, my valuable scientific information source and dear office mate with whom I shared so many cheers and wails, to Chrigu (thanks for all your scientific, magical, enjoyable and sportive input!) and to Conne (thanks for always taking care of a great "interactive" lab atmosphere - both scientifically and socially!) - to Eric (thanks so much for all your help! (and I am not a slave driver)), to Dania, to Sylvia (our good faithful soul), to Nicole and Julia - my entertaining lab companions, and to Andreas and Jean who introduced me to real-time PCR.

Many thanks also to Shinji, my prime victim for intricate immunology questions and ear matters, to Scheuri for the stoical provision of culture parasites and for the introduction to a disastrous FCB event, to Martin, Sonja (cooking artist and adventuress), Moni (definitely the most memorable parties ever), Ki and Marc (thanks for all your assistance !!!), Rolli, Michi, Nina, Sonja, Diana, Valentin, Simona and Elisabetta and many more with whom I shared delicate Mensa diners and amusing and interesting discussions - Thanks!

A special Thank you to Igor, Till and Tobi for being my mentors from the very first hour - for enduring my endless questions and oddish train of thoughts with a smile and valuable advice - and thank you for introducing me to billard albeit I have to admit, I did not really advance in that one…

I am deeply thankful to Alfred Cortés for his tremendous support and advice and spiriting discussions during my stays in PNG. Numerous thanks also to Ian Cockburn, to Judy Longo and Wilbert Jake and to the whole team of the Institute of Medical Research in Madang, especially to Kay and Manasseh Baea, Marta Mellombo, Moses Lagog, Alice Ura, Martin Imbran, Livingston Tavul, Moses Baisor, and Maggie and Andrew Raiko, who all took part in bringing these studies to a successful end. Many thanks extend to Kelley Vanbuskirk, Jennifer and Aaron Cole-Tobian, Ariadne Benet, Lucy Snow, Ged Casey and Neil Peacock for their contribution to an unforgettable time in PNG.

I wish to say thank you to Christine Walliser for readily supporting me in organizational matters and Heidi Immler and the Dicel team for rescuing me more than once from nerve-racking and mysterious computer affairs. Yvette, thanks so much for your heavenly Cafeteria and divine cakes!!!

For numerous relaxing hours while swimming, diving, indulging in exquisite desserts or strolling through Camden and much much more - for all your support and humorous and diverting hours - thank you so much and I won't name you fearing to forget someone - but you know who you are - thank you!!!

I am deeply grateful to my parents and my sister Christina for their love, support, patience and trust in me! Our Alpstein hikes, wonderful hours on the Bodensee and "Across-Switzerland-pilgrimages" let me go full speed ahead in my Ph.D. ☺ Merci vilmool ☺

Zusammenfassung

Malaria ist nach wie vor eine Krankheit von immenser gesundheitspolitischer Tragweite, vor allem in den Tropen und Sub-Tropen und verursacht jedes Jahr 300 Millionen Grippe-ähnliche Erkrankungen und tötet dabei eine Million Menschen. Die sich schnell ausbreitenden Medikamenten-Resistenzen, klimatische Veränderungen, aber auch nicht mehr intakte Gesundheitsversorgungen und bewaffnete Konflikte tragen zu einem steten Anstieg von Malaria bei. Dabei wird ein Impfstoff gegen diese Infektionskrankheit noch für längere Zeit nicht verfügbar sein.

Malaria wird von dem einzelligen Protozoen *Plasmodium* ausgelöst und durch die weibliche Anopheles Mücke übertragen. Von den 4 *Plasmodien* Arten, die Menschen infizieren können, ist *Plasmodium falciparum* bei weitem die virulenteste und auch für einen Grossteil der tödlichen Ausgänge verantwortlich. Diese ausgeprägte Virulenz von *P. falciparum* ist vor allem auch auf ein spezielles Phänomen zurückzuführen, das Zytoadhärenz heißt. Diese umfaßt die Sequestrierung und Bindung von infizierten Erythrozyten (IE) an die Endothelzellen des Mikrokapillarsystems. Zytoadhärenz ist für den Parasiten ein klarer Vorteil, da er die Eliminierung der IE in der Milz verhindert. Für den Wirt, den Menschen also, kann diese Adhäsion jedoch verheerende Konsequenzen haben, denn sie führt zu schwersten Behinderungen im Blutfluß, zu schlecht durchblutetem Gewebe und Hypoxia und trägt damit wesentlich zu schweren Krankheitsmanifestationen bei.

Nebst der Sequestrierung, beinhaltet die Zytoadhärenz auch noch die Rosetten-Bildung. Dies ist die Bindung von IE an uninfizierte Erythrozyten. Diese Klumpen-Bildung führt ebenfalls zu vermindertem Blutfluß und wurde mit schwerer Malaria assoziiert.

Hauptverantwortlich für diese Bindungsprozesse ist die auf der Oberfläche von IE exprimierte und vom Parasiten abstammende Proteinfamilie "*Plasmodium falciparum* erythrocyte surface protein 1" (PfEMP1). Durch die Exposition von PfEMP1 an der Erythrozyten Oberfläche setzt sich der Parasit aber auch dem Immunsystem aus und wird dadurch angreifbar. Der Parasit entgeht dieser Attacke durch Antigen Variation, d.h. durch die abwechselnde Exprimierung von verschiedenen Mitgliedern einer Proteinfamilie, in unserem Fall PfEMP1.

PfEMP1 wird von ungefähr 60 *var* Genen pro haploides Genom kodiert. Diese diversen und auch großen Gene werden in verschiedene adhäsive, semi-konservierte Domänen

strukturiert. Fast alle *var* Gene können zudem in verschiedene Gruppen aufgeteilt werden, d.h. in *var* Gruppe A, B oder C, gemäß ihren sehr unterschiedlichen, aber innerhalb einer Gruppe, konservierten untranslatierten Regionen.

Trotz der großen Einwirkungen von PfEMP1 auf die Malaria Pathogenese und auf das Parasiten Überleben, haben nur wenige Projekte *var* Gene und PfEMP1 *in vivo* untersucht. Das läßt sich vor allem auf die große, nicht abschätzbare Vielfalt der *var* Gene *in vivo* zurückführen, die eine Analyse äußerst erschweren.

Wir haben 2 Studien über die *var* Gen Expression in natürlich infizierten Kindern in Papua Neuguinea durchgeführt.

In einer longitudinalen Studie haben wir die Antigen Variation von *var* Genen in *P. falciparum* von älteren, semi-immunen Kindern analysiert. Mittels reverser Transkription, PCR, klonieren und sequenzieren wurde über 4 Monate in Zeitabständen von 2 Wochen die zeitliche Dynamik und Verteilung von *var* Transkripten ermittelt. Dabei haben wir ein äußerst dynamisches Bild der *var* Expression erhalten. Zum größten Teil wurden alle 2 Wochen neue *var* Transkripte ermittelt, wobei einige in einem Zeitrahmen von 10 Wochen wiederholt auftraten. Die Zahl der detektierten *var* Transkripte korrelierte mit der Anzahl von *P. falciparum* Stämmen, die ein Kind gleichzeitig infizierten. Im Durchschnitt wurden 1.7 verschiedene *var* Transkripte pro Kind und *P. falciparum* Stamm gefunden. Zudem wurde die rekombinogene Natur der *var* Genfamilie durch die Analyse von 286 verschiedenen Sequenzen von ausgesuchten *var* Gen Domänen bestätigt.

In einer zweiten Studie untersuchten wir die Frage, ob sich die strukturelle Gruppierung der *var* Gene auch in unterschiedlichen Funktionen widerspiegelt und sich in einer heterogenen Virulenz äußert. Die Exprimierung verschiedener *var* Gen Gruppen könnte demnach verschiedene pathologische Auswirkungen auf den Wirt haben. In einer Malaria Fall Kontroll Studie untersuchten wir daher die quantitative Verteilung der *var* Transkripte in den *var* Gen Gruppen A, B und C in Kindern mit schwerer Malaria, in Kindern mit milder Malaria und in asymptomatischen Kindern. Durch die Anwendung der real-time quantitativen PCR, fanden wir tatsächlich einen wesentlichen Unterschied in der Expression von *var* Genen zwischen Parasiten von kranken Kindern und Parasiten aus asymptomatischen Kindern. Dies äußerte sich in einer signifikanten Aufregulierung von Genen der *var* Gruppe B in Kindern mit klinischer Malaria verglichen zu asymptomatischen Kindern, in denen vor allem die *var* Gruppe C aktiv war. In Kindern mit klinischer Malaria fanden wir keine signifikanten Unterschiede der *var* Gen Expression

zwischen Kindern mit milder und schwerer Malaria. Nicht zuletzt stellten wir auch eine Aufregulierung der *var* Gene der Gruppe A in den Parasiten fest, die Rosetten bildeten.

Zusammenfassend, sind diese Studien über die *var* Gen Expression die ersten in ihrer Art, die in natürlich infizierten Kindern aus einem Malaria endemischen Gebiet durchgeführt wurden. Sie ermöglichen uns einen guten Einblick in die Dynamik und in die Auswirkungen der *var* Gen Expression *in vivo*. Zusammen mit früheren Studien, sind sie auch ein weiterer Beweis für den substantiellen Einfluß von PfEMP1 auf die Malaria Pathogenese.

Summary

Malaria is a tremendous global public health problem. While especially hitting the poorest countries in the world, malaria elicits each year 300 million febrile illnesses and up to 1 million deaths. Widespread drug resistances, climatic changes, but also disintegrated health services and armed conflicts have contributed to a global increase of malaria while a vaccine will not be at hand for many more years to come.

Malaria is caused by the protozoan parasite *Plasmodium* and transmitted by the female *Anopheles* mosquito. Of 4 *Plasmodium* species infecting humans, *Plasmodium falciparum* is by far the most harmful parasite responsible for nearly all mortality. The increased virulency of *P. falciparum* can be ascribed to special immune evasion strategies inherent of this species. This mainly refers to a process called cytoadherence, the sequestration and adhesion of infected erythrocytes (IE) to endothelial cells of the microcapillary system. To evade spleen dependent killing, cytoadherence is a benefit for the parasite, but detrimental to the host by leading to poorly diffused tissues and hypoxia in the upstream segments and thus, contributing substantially to severe manifestations.

Related to sequestration is a process called rosetting, the binding of IE to uninfected erythrocytes. This leads to erythrocyte clusters impeding local blood flow and accordingly, rosette formation was also associated with severe disease.

On the surface of IE, the parasite derived protein family *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) is thought to be the key mediator for sequestration and rosetting.

However, by exposing a parasite derived antigen on the surface of IE, the parasite gets vulnerable to immune attack. Therefore, *P. falciparum* evades the immune system by a process called antigenic variation, the switching of the expression between different members of PfEMP1.

PfEMP1 is encoded by approximately 60 *var* genes per haploid genome. The highly diverse and large *var* genes are structured into several adhesive, semi-conserved domains. Most *var* genes can be subgrouped into *var* group A, B and C according to their diverse, but within one group highly conserved untranslated regions.

Despite of the substantial contribution of PfEMP1 to malaria pathogenesis and parasite survival, few studies on *var* genes and PfEMP1 have been carried out *in vivo*. This is mainly due to their immense diversity interfering with most study designs.

We conducted 2 studies on *var* gene expression in naturally infected children from Papua New Guinea.

In a longitudinal study over a 4-month period in older, semi-immune children, we studied antigenic variation of *var* genes, namely the dynamics and distribution of *var* transcripts over time. Diversity and patterns of full-length *var* transcripts were evaluated by magnetic bead-anchored reverse-transcription polymerase chain reaction (RT-PCR), cloning and sequencing. We identified a highly dynamic picture of *var* expression with mostly new *var* transcripts at a 2-weeks interval but with some *var* transcripts recurring for up to 10 weeks. The number of detected *var* transcripts correlated with the number of infecting *P. falciparum* strains. On average, 1.7 different *var* transcripts were detected per child and infecting strain. The analysis of 286 different sequences of selected *var* gene domains confirmed the recombinogenic nature of *var* genes.

In a malaria case-control study on children from Papua New Guinea, we quantitatively compared the distribution of *var* transcripts among *var* groups A, B and C in children with severe malaria, with mild malaria and in asymptomatic children. The sub-division of *var* genes into these *var* groups raises questions about the biological or clinical significance of these structural differences. Upon expression, different *var* groups might have different pathological implications on the host leading to distinct virulences and different clinical outcomes. By using real-time quantitative PCR, we found a major expression difference between parasites causing clinical attack and parasites leading to asymptomatic infections. A significant up-regulation of *var* group B transcripts was evident in children with clinical malaria (mild and severe) while *var* group C genes were mainly switched on in asymptomatic children. No change in the distribution of *var* transcripts was detected between mild and severe disease. Finally, we found a significant up-regulation of *var* group A genes in parasites conferring the formation of rosettes.

Together, these studies on *var* gene expression are the first of its kind, conducted in naturally infected children in an endemic area. They are a step towards the comprehension of the dynamics and impacts of *var* gene expression *in vivo*. Together with previous studies, our data emphasize the substantial implications of PfEMP1 in malaria morbidity.

Abbreviations

ATS	acidic terminal segment
CIDR	cysteine rich interdomain region
CR1	complement receptor 1
CSA	chondroitin sulphate A
DBL	duffy binding like domain
GPI	glycosylphosphatidylinositol
HDMECs	human dermal microvascular endothelial cells
HLECs	human lung endothelial cells
ICAM-1	intercellular adhesion molecule 1
IE	infected erythrocytes
iNOS	inducible form of nitric oxide synthase
KAHRP	knob associated histidine rich protein
NTS	N-terminal segment
PAM	pregnancy associated malaria
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PNG	Papua New Guinea
RT-PCR	reverse transcription polymerase chain reaction
TM	transmembrane domain
TNF	tumour necrosis factor
TrHBMEC	transformed human bone marrow endothelial cell
VSA	variant surface antigen

1. Introduction

Malaria is found in most tropical and sub-tropical regions of the world affecting approximately 40% of the world's population and leading to annually 300 million acute illnesses and up to 1 million deaths (Roll Back Malaria; www.rbm.who.int.; 2004). 90% of malaria deaths occur in South Sahara African children. In Africa, malaria also presents a major obstacle to social and economic development, claiming annually up to US$ 12 billion economic costs. [1] Malaria is caused by the protozoan intracellular parasite *Plasmodium* and transmitted by the female mosquito of the genus *Anopheles.* Of the 4 *Plasmodium* species infecting humans (*P. falciparum, P. vivax, P. ovale, P. malariae*), *P. falciparum* is responsible for a high proportion of the morbidity and nearly all the mortality. *Plasmodium* belongs to the large phylum Apicomplexa which also includes opportunistic pathogens such as *Cryptosporidium* or *Toxoplasma* and some veterinary pathogens important for husbandry and agriculture, for instance *Babesia* and *Theileria* infecting cattle.

1.1. Malaria morbidity

Most of malaria morbidity is the result of the unrestrained asexual parasite amplification. The vast majority of malaria cases presents as non-specific acute febrile illness [2] and only about 1% proceed to severe manifestations of this disease. Traditionally, severe malaria was subdivided into cerebral malaria and severe anemia (haemoglobin < 5g/dl). [3] The latter is the consequence of haemolysis and an inappropriate bone marrow response. [3] In cerebral malaria, parasites obstruct the cerebral microcirculation contributing to increased cerebral capillary permeability and cerebral oedema (see also section 1.2.). [4] Recent studies indicated that also metabolic acidosis leading to the clinical picture of respiratory distress, is an important feature of severe malaria being also the strongest predictor of death in severe malaria. [5] Hypovolemia, hyperlactatemia and impaired renal function have been shown to mainly contribute to metabolic acidosis. [5] More and more, severe malaria is thought to be a complex multisystem disorder with many similarities to sepsis syndromes. [2] Immunopathogenic processes with excessive proinflammatory cytokine productions such as IFN-γ and tumour necrosis factor (TNF) play an important role. TNF is also known to induce the generation of the inducible form of nitric oxide synthase (iNOS). iNOS produces nitric oxide in vast amounts which was implicated in cerebral malaria by

interfering with neuro-transmission and leading to vasodilation of cerebral vessels. [6] The rupturing of erythrocytes to release new merozoites is suggested to trigger excessive proinflammatory cytokine cascades and oxygen free radicals. Furthermore, the glycosylphosphatidylinositol (GPI) of *P. falciparum* has been shown to act as malarial toxin contributing to the cytokine cascades. [7]

However, probably the most important virulence factor of *P. falciparum* has not been mentioned yet - this is a process called cytoadherence.

1.2. Cytoadherence and pathophysiological consequences

Cytoadherence is the adhesion of infected erythrocytes (IE) to other host cells during the trophozoite and schizont stage, which is the last half of the parasites' asexual blood stage replication cycle. Cytoadherence can be differentiated into sequestration - the adhesion of IE to endothelial cells in the post-capillary venules [8] - into rosetting - the adhesion of IE to other non-infected erythrocytes [9] - and into clumping - the platelet mediated binding of IE to other IE. [10] Cytoadherence is believed to confer a fitness benefit to the parasite such as by evasion of spleen dependent killing. For the host, these microvascular obstructions have clear pathogenic consequences leading to poorly perfused host tissues, anaerobic metabolism and hypoxia which further contributes to metabolic acidosis. An association has been found between binding of IE to intercellular adhesion molecule 1 (ICAM1) and cerebral malaria. [11,12] ICAM1 is prominently expressed on endothelial cells in the brain and involved in local inflammatory responses allowing the passage of leukocytes into the perivascular space. Binding of IE to ICAM1 was hypothesized to mimic the binding of leukocytes eliciting signalling cascades which result in the leakage of plasma proteins into the perivascular space contributing to cerebral oedema. [4,13] It has been shown that TNF enhances ICAM1 expression on endothelial cells and thus, high levels of TNF are thought to contribute further to parasite sequestration in the brain. It was also shown that binding of IE to human lung endothelial cells (HLECs) induces apoptosis in these cells. [14] Apoptosis of endothelial cells in the brain might lead to lesions and could further contribute to blood-brain barrier dysfunctions.

Another important receptor for IE binding is CD36, a scavenger class B receptor which is found on phagocytic cells and on endothelial cells in muscle tissues. Most isolates have been shown to bind to CD36, whereas binding of isolates to ICAM1 is only occasionally seen and is also of weaker strength. Controversial data exist on the contribution of IE binding to CD36 to malaria pathogenicity. Binding to CD36 was implicated with disease

by sequestration and by conferring platelet mediated clumping which correlates with severe disease. [10,15] Conversely, in a field study, CD36 binding isolates were found to be associated with non-severe disease, [12] it was proposed to be protective by sequestering IE in non-essential muscle tissues (reviewed by Serghides et al. [16]) and by promoting non-inflammatory phagocytosis of IE. [17,18] Also immune-modulatory functions were correlated to CD36 binding such as the down-regulation of dendritic cell activity by binding of IE to dendritic cells via CD36. [19]

A number of other receptors such as heparin, TSP, non-immune immunoglobulins, P-selectin, PECAM (CD31) or $\alpha_v\beta_3$-integrin have all been shown to bind to IE *in vitro*, but its implications *in vivo* and role in malaria pathology are not yet clear (reviewed by Sherman et al.[20]). However, it is known that the binding of IE to chondroitin sulfate A (CSA) in the placenta plays an important pathological role in pregnancy associated malaria (PAM). [21]

Finally, a process called rosetting - the binding of IE to uninfected erythrocytes - has been shown to be associated with severe malaria in Africa. [22] Rosetting has been found to be serum dependent [23] and complement receptor 1 (CR1) was shown to be the main ligand on uninfected erythrocytes. [24] Rosetting is thought to be advantageous for the parasite by providing improved invasion of merozoites into uninfected erythrocytes or by shielding IE from host phagocytic cells or antibodies. [25]

1.3. PfEMP1

The highly polymorphic *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) family contributes mainly to cytoadherence. These are large proteins (200-350 kDa) deposited on the surface of infected erythrocytes (IE) from approximately 18 hours post invasion onwards. [26] PfEMP1 is located on the infected erythrocyte surface on knob-like structures. The parasite derived proteins KAHRP (knob associated histidine rich protein) [27] and probably PfEMP3 play a role in knob formation and anchoring of PfEMP1 to the erythrocytic cytoskeleton. [28] It has been shown that large amounts of PfEMP1 molecules remain within the erythrocyte in vesicle like structures, the Maurer's clefts, suggesting possible post-translational control or slow and inefficient transport of this protein to the surface. [28,29]

1.4. *var* genes

In 3D7 *P. falciparum* culture strain, PfEMP1 is encoded by 59 highly diverse *var* genes, each of 8 to 14 kb of length. [30-33] They have a 2 exon-structure with exon 1 encoding the highly diverse extracellular part of PfEMP1 and a predicted trans-membrane domain (TM) and exon 2 encoding the conserved intracellular acidic terminal segment (ATS) anchoring the protein to the cytoskeleton. PfEMP1 molecules are structured into several semi-conserved domains namely a N-terminal segment (NTS), duffy binding like (DBL) domains, cysteine-rich interdomain regions (CIDR) and in some instances a "constant 2" (C2) region (see Figure 1)(reviewed by Smith et al.[34]). DBL domains belong to a family of receptor binding motifs, which were previously described in merozoite proteins involved in erythrocyte invasion, such as *P. vivax* duffy binding proteins or *P. falciparum* EBA-175. [35] DBL and CIDR domains are numbered in order of their location from the 5'end of PfEMP1. Furthermore, based on sequence similarities, mostly conserved cysteine residues, different subclasses of DBL domains (α to ε, x) and CIDR domains (α-γ) have been identified. The most N-terminal DBL1α and CIDR1α form the structurally conserved head structure which is found in almost all PfEMP1 molecules.

Binding to specific receptors has been associated to various domains of PfEMP1, such as DBL1α to CR1 in rosetting [36], CIDR1α to CD36 [37], DBL2β-C2 to ICAM1 [38] and DBLγ and CIDR1α to CSA [39,40].

Recently, a recombinant peptide which corresponds to the minimal CD36-binding domain of PfEMP1 was shown to bind to human dermal microvascular endothelial cells (HDMECs) which activated a signalling pathway in these cells, namely the Src-family kinases and downstream the mitogen-activated protein (MAP) kinase pathway. [41] It was suggested that this activated a dephosphorylation process of CD36 molecules on these endothelial cells which could result in an increase of IE adherence to CD36.

However, not all PfEMP1 molecules with a certain adhesive domain also bind to the corresponding receptor and it has also been shown, that only 3 amino acid changes in a CIDR1α domain greatly reduced binding to CD36. [42] This indicates that binding abilities also rely on tertiary folding structures and are not obvious from primary sequence. Additionally, binding characteristics of a single domain of PfEMP1 might not correspond to the binding abilities of the whole protein. This discrepancy is seen in pregnancy-associated malaria (PAM), where contradictory results exist on serological data, *in vitro* CSA binding studies on domains of PfEMP1 and *in vivo* binding characteristics of PAM

associated parasites. There is also evidence for non-specific or cross-reactive binding of antibodies to CSA-selected parasites. This questions previous reports of successful CSA binding inhibition assays or immunizations with CSA binding recombinant PfEMP1 domains (reviewed by Rowe et al. [43]).

1.5. *var* gene subgroups

Most *var* genes can be grouped regarding their different, but within each group, highly conserved 5' upstream sequences (upsA, upsB and upsC) [30,44] (see Figure 1 [after Smith et al. [45,46]]). In 3D7, the majority of *var* genes belong to *var* group B flanked by an upsB sequence and located subtelomerically on the chromosomes. This region is recombinogenic with various repetitive elements which was suggested to support ectopic recombination by clustering heterologous subtelomeric stretches (including *var* genes). [47] It might be speculated that *var* genes generate their immense diversity in this region by various recombination events. *var* group A consists of 10 larger *var* genes with a distinct domain structure. *var* group A genes are also located subtelomerically, however, they are transcribed towards the opposite direction as *var* group B genes, which was proposed to reduce potential DNA exchange between *var* group A and B.

Figure 1: *var* genes and domain structure

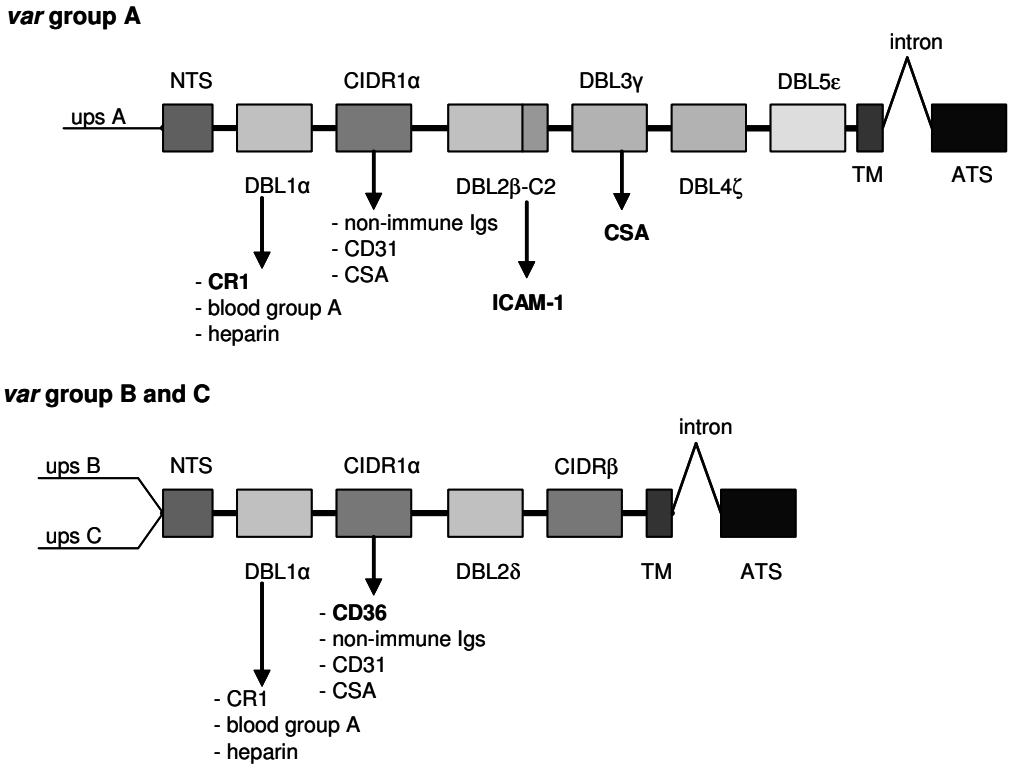


Figure 1: Characteristic domain structure of a *var* group A gene and a *var* group B or C gene. Domains which have been shown to be involved in binding are indicated. See text for explanations

Finally, 13 *var* genes called *var* group C are clustered centrally on chromosomes with a similar domain structure and length than *var* group B genes.

Not much is known about the evolution of *var* genes which are to date only found in *P. falciparum*. It can only be speculated that an ancestral *var* gene had incorporated adhesive, cysteine rich domains such as the domain DBL. This domain seems to be of older origin since it is found in different *Plasmodium* species (see above). Insertion of this domain and its further duplication together with the assembly of other adhesive domains rendered the *var* gene to a sticky molecule. It is questionable if a stable chromosomal environment could have been the basis for the relatively fast evolution of a large gene family not present in other *Plasmodium* species. Therefore, it seems more likely that the original *var* gene was located subtelomerically embedded in a highly repetitive and recombinogenic region subject to a fast evolution force. *eba-175* which is a potential source of the acquired DBL domain is also located on the subtelomers. Duplication event of this ancestral *var* gene likely generated the 3 subgroups. The highly conserved flanking sequences of *var* genes and phylogenetic analyses suggest that further duplications and recombinogenic events were mainly concentrated within *var* groups.

Apart of *var* groups A, B and C, there are 2 particular *var* genes in 3D7 - *var*1 and *var*2. They belong to *var* group A according to their chromosomal location and transcriptional direction. [45] However, both have a distinct 5' upstream region (upsD and upsE). *var*1 is similarly structured than *var* group A genes, but has no ATS domain in 3D7. *var*2 shows an unusual domain structure with a DBL1x and no CIDR. Both *var* genes have been shown to be highly conserved among field isolates and have been associated with pregnancy associated malaria (PAM) and CSA binding [48,49](reviewed by Rowe et al.[43]).

1.6. PfEMP1 and potential immune-modulatory functions

PfEMP1 has also been discussed to mediate immune-modulatory processes. Blood stage parasites are thought to be mainly controlled by innate immune responses such as by phagocytosis of IE by splenic macrophages, and by adaptive responses such as antibodies which either inhibit cytoadherence, erythrocyte invasion or mediate antibody dependent cytotoxicity and cellular inhibition (reviewed by Artavanis et al. and by Urban et al. [50,51]). To prime the adaptive immune response, antigen presenting cells such as dendritic cells play an important role in stimulating T cells. As mentioned above, there is evidence that binding of IE to dendritic cells down-regulates the activation of these cells and reduces

their ability to stimulate T cells. [19] This is thought to be mediated by binding of PfEMP1 to CD36 on dendritic cells. Phagocytosis of IE by macrophages was also shown to lead to their inactivation. Apart of the involvement of lipoperoxides generated by the haemozoin of the ingested parasite (reviewed by Urban et al. [51]), it was found that macrophages which phagocyte IE involving binding to CD36 do not elicit the usual pro-inflammatory cascade. [17,18] At last, the CIDR1 domain of PfEMP1 was recently identified to be a poly-clonal B cell activator similar to protein A of *Staphylococcus aureus* and able to divert specific antibody responses. [52]

1.7. *var* genes - expression regulation and antigenic variation

To evade the immune system, antigenic variation occurs in *P. falciparum*. This is the switching of expression between members of a surface antigen family. In the case of *P. falciparum* this is the switching of *var* genes. Antigenic variation is characteristic to parasites maintaining chronic infection and is also found in African trypanosomes, *Neisseria, Borrelia* or *Giardia lamblia* (reviewed by Kyes et al.[53]). There are still many questions regarding *var* gene regulation. It has been shown that every *var* gene represents a single transcriptional unit capable of *in situ* activation believed to involve epigenetic mechanisms. [54,55] Parasites are thought to express one PfEMP1 variant on the IE surface. By northern blots and reverse-transcription polymerase chain reaction (RT-PCR), a "leaky" transcription from the 5'end of most *var* genes in ring stage parasites was detected. [56,56,57] It was shown that the *var* gene encoding the expressed PfEMP1 is transcribed for about the first 24h of the erythrocytic cycle and might involve some form of epigenetic programming. [56,57] Voss et al. performed *var* promoter analyses by transient transfections which resulted in the identification of promoter motifs involved in *var* group specific repression and silencing (see Supplement III [58]). In the same study, a slightly shifted period of transcription was found between *var* group B and *var* group C genes raising questions on potentially different functional features of these 2 groups.

1.8. Molecular mechanisms of *var* switching

S-phase dependent chromatin assembly has been suggested to be implicated in silencing of *var* genes involving the interactions between the *var* intron and a *var* upstream region. [55] By transient transfection assays, Voss et al. [58] identified a *var* upstream motif which seems to be involved in silencing (see Supplement III). Calderwood et al. [59] showed that the *var* intron also possesses promoter activities speculating that the intron acts as an insulator

element building a boundary between transcriptionally active and silent chromatin regions. Moreover, boundary elements have been shown to be actively transcribed in *Drosophila* and in yeast. This could also explain the presence of sterile *var* transcripts consisting of the *var* intron and *var* exon 2. These were discovered together with *var* genes [31] but could not be explained to date.

1.9. Antigenic variation and cytoadherence in *Plasmodium*

Various other *Plasmodium* species are provided with multi-gene families potentially mediating antigenic variation such as the well studied SICA*var* genes with 50-100 copies in the simian malaria *P. knowlesi.* [60] In SICA*var* expression regulation, post-transcriptional gene silencing involving the 3'UTR was suggested to be involved. [60] Despite of highly conserved 3'UTR sequences within *var* groups, no such mechanisms were found for *var* genes. Another large multi-gene family of up to 1000 copies is called *vir* and was described in *P. vivax*. Homologs are found in various *Plasmodium* species such as in *P. knowlesi, P. yoelii, P. chabaudi, P. bergheii* or *P. cynomolgi* but not in *P. falciparum* (reviewed by Kyes et al. [53]). Antibody responses against these proteins were shown to be elicited *in vivo*. But despite of their location on the erythrocyte surface and potential antigenic variant character, in contrast to PfEMP1 most of them do not mediate cytoadherence. However, heavy loads of late-stage *P. knowlesi* were shown in the placenta of pregnant macaque monkeys suggesting some sort of sequestration. Furthermore, *P. knowlesi* parasites in splenectomized monkeys did not express SICA*var* anymore and were less virulent. [61] This classifies SICA*var* as a potential virulence factor of *P. knowlesi* similar to PfEMP1 in *P. falciparum*. But no homology to *var* genes could have been shown apart of the existence of cysteine rich domains. These analogous multi-gene families seem to have evolved independently but display a similar function. Also *P. chabaudi* IE in mice were shown to sequester, but mainly in the liver. 10 multi-copy families were recently identified in this species, but the family responsible for sequestration was not identified yet and no *var* gene homologs were found. [62] It is a matter of discussion why these surface antigen families evolved and whether they first evolved the ability for cytoadherence followed by antigenic variation to escape the immune system (reviewed by Kyes et al. [53]). However, since most *Plasmodium* strains apart of *P. falciparum* and few others do not sequester, it is likely that cytoadherence is a particular virulence factor characteristic of *P. falciparum* which evolved at a later time point. Also, the evolutionary driving force of cytoadherence is not yet totally clear. Evasion of spleen dependent killing is the most

accepted assumption. This is supported by reports on late stage *P. falciparum* parasites circulating in splenectomized humans. [63] In another study, *P. falciparum* parasites lost the ability to sequester in splenectomized Saimiri monkeys but displayed a different antigenic variant phenotype suggesting the involvement of another multi-gene family (reviewed by Kyes et al. [53]). Other multi-gene families have been detected in *P. falciparum* such as the immunogenic rifins encoded by about 200 *rif* genes per haploid genome. [64] It is still unclear whether they are expressed at the surface of erythrocytes. Apart of spleen dependent killing, other factors have been proposed to drive the evolution of cytoadherence such as the provision of a favourable microaerophilic venous environment for parasite maturation, specific sequestration of gametocytes ensuring long-term transmission or the shielding of IE from host phagocytic cells or antibodies by binding to uninfected erythrocytes (rosetting).

But why does the parasite send proteins onto the surface if not by rendering the surface of IE sticky? Apart of potential immune-modulatory functions, [19,52] Kyes et al. [53] also suggested a shielding function of PfEMP1 molecules covering the infected senescent erythrocyte and prevent recognition by autologous IgGs or complement. The human erythrocyte anion exchanger 1, called band 3, is the main target of anti-senescent erythrocyte antibodies in its modified form. It is believed that denatured haemoglobin forms hemichrome which cross-links to the cytoplasmic domain of band 3 eliciting aggregates of band 3. [65] The fact that these modified band 3 proteins cluster in the knobs together with PfEMP1 might not be coincidence. Finally, Saul et al. [66] even argued that parasites deliberately present parasite-derived antigens to the immune system to generate an immune response. By restricting parasite growth a premature death of host and thus aborted transmission is prevented. However, this hypothesis is inconsistent with data on *P. falciparum* and *P. chabaudi* mutants showing no detectable surface antigens while being less virulent. [53]

To sum up, antigenic variation is a common feature of *Plasmodium* species ensuring long-term survival. Cytoadherence seems to have evolved in addition in few species representing a devastating virulence factor for the host. Nevertheless, advantages for the parasite seem to outweigh a potential premature death of the host.

1.10. PfEMP1 *in vivo*

Apart of studies on binding abilities of field isolates or recombinant PfEMP1 peptides (see section 1.4.), not much is known about pathological impacts of different PfEMP1 or their role in eliciting a specific immune response. However, serological studies on PfEMP1 have shown that this family is a target of naturally occurring antibodies which also provides variant specific protection. [67,68] Important findings were obtained by Bull et al. [69] on the potentially different virulent nature of variant surface antigens (VSA), of which PfEMP1 is the best characterized. By comparing agglutination frequencies of parasites of Kenyan children, they found that children with a low antibody repertoire against VSA were infected with parasites expressing a restricted virulent subset of VSA which was commonly recognised by heterologous community plasma while parasites of hosts with a large anti-VSA antibody repertoire expressed rarer VSA which were rarely recognised by heterologous plasma. The question arises if these VSA subsets coincide with *var* groups, namely *var* group A, B or C. Recent studies indicated an involvement of *var* group A in severe malaria such as the finding of larger PfEMP1 on the surface of parasites conferring cerebral malaria [70] and the detection of a subgroup of DBL1α sequences of *var* transcripts isolated of parasites eliciting severe disease. [71] The involvement of *var* group A was further confirmed by the findings of Jensen et al. [72] who found an up-regulation of several *var* group A genes in 3D7 culture strains which were *in vitro* selected for severe malaria VSA phenotype by panning parasites on plasma pools of semi-immune children and by panning parasites on transformed human bone marrow endothelial cells (TrHBMEC) which express various receptors such as VCAM-1, P-selectin or ICAM-1, but not CD36.

Apart of the open question about the virulent nature of different PfEMP1s, there is very little known about expression and dynamics of *var* gene switching *in vivo*. *In vitro*, high switching rates of 2% per generation were measured. [73]

The only data on *var* switching *in vivo* were received by volunteers artificially infected with 3D7. [74] Surprisingly, the same *var* gene was found early in infection in different volunteers suggesting some form of imprinting. The initial switching rate was estimated to be at 16% but decreased thereafter also suggesting different switching rates for different *var* genes and host conditions.

1.11. Aim of this study

PfEMP1 encoded by *var* genes is an important virulence factor of *P. falciparum* successfully evading the immune system. However, few studies have been carried out to study the dynamics and expression of *var* genes *in vivo*. We carried out a longitudinal study during a 4-month period in malaria semi-immune children from Papua New Guinea. The longitudinal distribution and structure of *var* transcripts was analyzed by RT-PCR, cloning and sequencing. Furthermore, by amplification of the conserved 5' upstream sequences of *var* groups, we also analyzed *var* transcripts with respect to their affiliation to *var* group B or C.

Moreover, in a malaria case-control study in children from Papua New Guinea we evaluated the potential difference in the virulence of *var* group specific PfEMP1. The distribution of *var* transcripts among *var* group A, B and C was analyzed by quantitative real-time PCR and compared in children with severe malaria, mild malaria and asymptomatic malaria.

These studies are an important step towards understanding the effect and the dynamics of *var* gene expression in naturally infected individuals.

2. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea

Mirjam Kaestli [1], Alfred Cortes [2], Moses Lagog [2], Michael Ott [3], Hans-Peter Beck [1]*

Affiliations of authors:

1 Swiss Tropical Institute, Socinstr.57, Postfach, 4002 Basel, Switzerland

2 PNG Institute of Medical Research, PO Box 378, Madang, MP 511, Papua New Guinea

3 F. Hoffmann-La Roche AG, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

Reprints or correspondence:
Hans-Peter Beck, Swiss Tropical Institute, Socinstrasse 57, CH-4051 Basel, Switzerland
Tel: +41-61-284 8116
Fax: +41-61-271 8654
E-mail: hans-peter.beck@unibas.ch

Published in *The Journal of Infectious Diseases* (2004), 189:1942-51

MAJOR ARTICLE

Longitudinal Assessment of *Plasmodium falciparum* *var* Gene Transcription in Naturally Infected Asymptomatic Children in Papua New Guinea

Mirjam Kaestli,[1] Alfred Cortes,[3,a] Moses Lagog,[3] Michael Ott,[2] and Hans-Peter Beck[1]

[1]Swiss Tropical Institute and [2]F. Hoffmann-La Roche AG, Basel, Switzerland; [3]Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea

Sequestration and antigenic variation are essential for *Plasmodium falciparum* survival in vivo contributing to severe pathologic findings and, also, chronic infection. Both are conferred by *P. falciparum* erythrocyte membrane proteins encoded by ~60 *var* genes. To study the dynamics of *var* gene expression, we conducted a 4-month longitudinal study of semi-immune children from Papua New Guinea. By use of magnetic bead–anchored reverse-transcription polymerase chain reaction analysis performed over 5 *var* regions, as well as cloning and sequencing, the longitudinal distribution of full-length *var* transcripts was analyzed. We identified a dynamic picture of *var* gene expression with rapid switches but with identical *var* transcripts recurring for up to 10 weeks. The number of *var* transcripts was correlated to the number of infections, with a mean of 1.7 *var* transcripts identified per sample and infecting strain. Analysis of 158 different Duffy binding–like 1α sequences confirmed the recombinogenic nature of *var* genes. This is the first report of the dynamics of *var* gene expression in chronically infected children.

Cytoadherence and sequestration of *Plasmodium falciparum*–infected red blood cells (RBCs) are considered to be among the most important factors associated with the pathogenicity and virulence of *P. falciparum* malaria. Cytoadherence is mediated by the polymorphic *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is located on the surface of infected RBCs. PfEMP1 is encoded by 1 of ~60 *var* genes, each of which is 8–14 kb in length [1–3]. These large proteins of 200–350 kDa mediate binding to various cell surface receptors (reviewed in [4]). Cytoadherence is thought to prevent spleen-dependent killing, but it also has been shown that binding of PfEMP1 to CD36 presented on dendritic cells down-regulates dendritic cell activity, suggesting an immune modulatory role for PfEMP1 [5]. PfEMP1 has also been implicated in rosetting [6], is a target of naturally occurring immune responses, and shows antigenic variation [7, 8]. The expression of different PfEMP1 variants is accompanied by changes in the adhesive phenotype of infected RBCs [7].

PfEMP1 proteins are structured into several semiconserved domains—namely, an N-terminal segment (NTS); various Duffy binding–like (DBL) domains; a cysteine-rich interdomain region (CIDR); in some instances, a "constant 2" (C2) region; a transmembrane domain; and the conserved, C-terminal acidic terminal segment (ATS), which represents the intracellular part of PfEMP1 that anchors the protein to the cytoskeleton (reviewed in [9]). Different subclasses of DBL domains (α–ε and x) and CIDR domains (α–γ) have been identified. The most N-terminal DBL1α and CIDR1α form the conserved head structure of the protein. This head structure is found in almost all PfEMP1 molecules. Binding has been associated with various domains of PfEMP1 [6, 10–15], such as DBL1α to CR1 (in rosette formation), CIDR1α to CD36, or DBLβ-C2 to intracellular adhesion molecule 1 (ICAM1). Binding to chondroitin sulphate A in samples from placental ma-

Received 5 September 2003; accepted 26 October 2003; electronically published 29 April 2004.

Financial support: Swiss National Science Foundation (grants 031-059 064 and 031-67211.01).

Present affiliation: MRC National Institute for Medical Research, Mill Hill, London, United Kingdom.

Reprints or correspondence: Dr. Hans-Peter Beck, Swiss Tropical Institute, Socinstr. 57, CH-4051 Basel, Switzerland (hans-peter.beck@unibas.ch).

The Journal of Infectious Diseases 2004;189:1942–51

0022-1899/2004/18910-0020$15.00

laria has been shown to occur not only with DBL-γ but, also, with CIDRα [12–14].

Most *var* genes can be classified into 3 groups, on the basis of their different but, within each group, highly conserved 5′ upstream sequences [16, 17]. The majority of *var* genes are located subtelomerically and possess a upsB-type upstream region. These *var* genes are located in a region that is highly recombinogenic with various repetitive elements that support ectopic recombination by clustering of heterologous subtelomeric stretches (including *var* genes) [18]. Another small set of *var* genes is arranged in chromosome internal clusters; these genes possess upsC-type upstream regions. A third group of *var* genes consists of subtelomerically located *var* genes, which are transcribed toward the telomeres (upsA-type regions [17]). That *var* gene transcription of the upsB- and upsC-type regions is regulated differently [19] raises questions about the different functional features of these 2 *var* gene groups.

The detailed mechanism of *var* gene regulation is still unknown, but it has been shown that every *var* gene represents a single transcriptional unit that is capable of in situ activation involving epigenetic mechanisms [20, 21]. The *var* genes are transcribed for the first 24 h of the erythrocytic cycle [4]. Whereas relaxed transcription of multiple truncated *var* genes was found at the ring stage, only one full-length *var* transcript was found at the early trophozoite stage [22]. A switching rate of 2.4%/generation was calculated in vitro [23], but little is known about *var* gene expression and switching in vivo. In a longitudinal study involving nonimmune adults who were artificially infected with the 3D7 laboratory-adapted strain, Peters et al. [24] showed that the first transcribed *var* gene in the erythrocytic stage was identical in different adults. The initial switching rate was estimated to be 16%, but it decreased thereafter, suggesting different switching rates for different *var* genes and host conditions.

In the present study, we describe *var* gene expression in naturally infected semi-immune children from Papua New Guinea during a 4-month period. The longitudinal distribution and structure of the expressed *var* transcripts were analyzed by reverse-transcription polymerase chain reaction (RT-PCR), cloning, and sequencing. We also analyzed *var* transcripts, with respect to the chromosomal location using the conserved 5′ upstream regions for amplification. Using this approach, we describe the dynamic nature of *var* gene expression in several asymptomatic children. Many *var* genes were transcribed simultaneously with switches at short intervals, but some identical transcripts recurred in the same child, even after 10 weeks.

MATERIALS AND METHODS

Study area and collection of blood samples. From April to August 2001, the period of transition from the wet season to the dry season, we performed a longitudinal study at the Maiwara Primary School on the Madang North Coast in Papua New Guinea. In this region, where malaria with perennial transmission is endemic, infections with *P. falciparum* and *P. vivax* are common. *P. malariae* and *P. ovale* are also present in this area.

Written, informed consent was obtained from the children's parents or guardians. The study was approved and ethical clearance was given by the Medical Research Advisory Committee of Papua New Guinea. According to national treatment guidelines, children with parasites but without malaria symptoms were not treated.

After informed consent was obtained from parents, we obtained blood samples, by fingerprick (0.2–0.5 mL) or venopuncture (2 mL), from 11 children who had asymptomatic *P. falciparum* infections. The children were 8–10 years of age. Samples were obtained from 8 children every 2 weeks for 4 months and from 3 children every 5 days for 1 month.

Assessment of* P. falciparum *infections. Giemsa-stained blood slides were analyzed by microscopy. For blood samples that were found to be positive, by microscopy, for *P. falciparum*, the number of *P. falciparum* infections was determined by *msp2* genotyping, as described elsewhere [25]. In brief, 30 μL of full blood was spotted on filter papers (Isocode Stix; Schleicher & Schuell) and was dried for 20 min at 80°C. After washing, *msp2* PCR was performed directly on the filter papers, and restriction fragment–length polymorphism (RFLP) analysis of nested PCR products was used to record the number of infecting strains.

Isolation of full-length* var *transcripts and RT-PCR. Total RNA was extracted using TRIzol (Invitrogen), according to the manufacturer's instructions. Extraction with TRIzol was performed twice, to decrease DNA contamination. After RNA was treated with 3 U of RQ1RNase-free DNase (Promega), another extraction with TRIzol was performed. To obtain only full-length *var* transcripts, RNA was dissolved in binding buffer (0.5 mol/L LiCl, 1 mmol/L EDTA, 10 mmol/L Tris, pH 7.5), and 1 pmol of biotinylated oligonucleotide complementary to the ATS domain (Biotin-5′-GGTTC(A/T)A(A/G)TAC(C/T)ACTTC(A/T)AT(C/T)CCTGGT(A/G)CATATATATCATTAATATCCAATT-CTTCATA(C/T)TCACTTTC(T/G)GA(A/T/G)GA-3′) was added and was incubated at a temperature gradient from 65°C to 4°C over 30 min. One hundred fifty micrograms of Dynabeads M-280 streptavidin, washed according to the manufacturer's protocol and dissolved in 5.5 mol/L LiCl, was added to the RNA. After undergoing rotation for 30 min at 37°C, the beads were washed 3 times with washing buffer (10 mmol/L Tris, 1 mmol/L EDTA, 0.15 mol/L NaCl, pH 7.5) and 1 time with 10 mmol/L Tris. RT was performed on the captured hybrids, primed by 400 ng oligo$(dT)_{12–18}$, and was done by use of Sensiscript (Qiagen) reverse transcriptase, according to the manufacturer's protocol, in a final volume of 20 μL. An aliquot without reverse transcriptase was used as a negative control. After RT, cDNA was treated with RNase A, and 1 μL was used for each of the various

PCR amplifications (table 1) with Advantage cDNA polymerase (Clontech), by use of the primers shown in table 1. The PCR conditions were 35 cycles for 30 s at 95°C, for 1 min at the annealing temperature (table 1), and for 70 s at 64°C. One microliter of negative control (without reverse transcriptase) was amplified in parallel. If this negative control yielded a product, then the positive sample was discarded and was excluded from the analysis.

To exclude cross-contamination of RT-PCR products, sequence-specific primers were designed for those sequences that occurred in >1 child. Sequence-specific PCR was performed on the genomic DNA of the parasites of those children, to confirm the presence of each particular sequence.

Cloning and sequencing of PCR products. PCR products were cloned into pGEM-T vector (Promega) or pGEM-3Zf(+) vector (Promega), according to the manufacturer's instructions, and they were transfected into *Escherichia coli* SURE cells (Stratagene). An average of 20 positive clones was processed for sequencing (Montage Plasmid Miniprep$_{96}$ Kit [Millipore]; 96 capillary ABI Prism automated sequencing system [Applied Biosystems]). Multiple alignment of identical sequences derived from the same probe allowed the exclusion of PCR-derived mutations. Two sequences were considered to be identical when <3 single-nucleotide polymorphisms (SNPs) were detected.

Sequence analysis. DNA sequence analysis was performed using DNASTAR (version 4; http://www.dnastar.com/), BLAST (from the National Center for Biotechnology Information Web page [http://www.ncbi.nlm.nih.gov/BLAST/]), plasmoDB, CLUSTALW (http://searchlauncher.bcm.tmc.edu/multi-align/ or http://www.ebi.ac.uk/clustalw/), and BioEdit (version 5; http://jwbrown.mbio.ncsu.edu/BioEdit/bioedit.html). Phylogenetic analyses were performed using PHYLIP (version 3.6; http://evolution.genetics.washington.edu/phylip.html) or Molecular Evolutionary Genetics Analysis (MEGA, version 1.02; The Pennsylvania State University [http://evolgen.biol.metro-u.ac.jp/MEGA]), with neighbor-joining or maximum parsimony methods and accompanied by bootstrap analysis with 1000 replicates. Predictions of secondary structure were performed using PredictProtein (http://cubic.bioc.columbia.edu/predictprotein/). Population comparison analysis was performed using ARLEQUIN (version 2; analysis of molecular variance; Genetics and Biometry Laboratory, University of Geneva [http://anthropologie.unige.ch/arlequin/]). Nucleotide sequence data are available in GenBank (accession nos. AY462581–AY462851).

RESULTS

var *Gene transcription in individual children.* We studied *var* gene transcription longitudinally in asymptomatic children in Papua New Guinea for 4 months. RT-PCR was performed on full-length *var* transcripts over 3 adhesive domains (DBL1α, CIDR1, and DBLβ) and 2 upstream regions (subtelomeric upsB-type 5′ untranslated region [UTR] to DBL1α and central upsC-type 5′UTR to DBL1α) (figure 1*A*). RT-PCR was followed by cloning and sequencing of the PCR products. The DBL1α domain yielded the most-informative PCR product, because it contains conserved blocks that enable the design of universal primers and thus ensures the amplification of the majority of *var* transcripts with minimal bias [22]. However, we are aware that, with this approach, potential bias in the identification of sequences might occur.

We observed a highly dynamic and variable picture of *var* transcription, which, in the next section, is described in detail for 4 representative children (figures 1–4). The longitudinal dis-

Table 1. Oligonucleotide primers used for amplification of *var* gene regions.

var Gene region	Length of amplified product	T_{anneal}	Name	Primer sequence	Reference
upsB-type 5′UTR-DBL1α	1 kb	53°C	*var* 4A3-5′ [a]	5′-CTCAT(A/T)TATAATTTTACAAAATATATAAAAC-3′	[16]
			var 4A3-3′ [b]	5′-CC(A/T)AT(A/G)GC(A/G/T)GCAAAACT(G/C/T)CG(A/T)GC-3′	[16]
upsC-type 5′UTR-DBL1α	1 kb	54°C	*var* 5B1-5′	5′-CACATATA(A/G)TACGACTAAGAAACA-3′	[16]
			var 4A3-3′	5′-CC(A/T)AT(A/G)GC(A/G/T)GCAAAACT(G/C/T)CG(A/T)GC-3′	[16]
DBL1α	400 bp	52°C	DBLα-5′	5′-GCACGAAGTTTTGCAGATAT(A/T)GG-3′	
			DBLα-3′	5′-AA(A/G)TCTTC(T/G)GCCCATTCCTCGAACCA-3′	
	400 bp	51°C	αAF	5′-GCACG(A/C)AGTTTTGC-3′	[22]
			αBR	5′-GCCCATTC(G/C)TCGAACCA-3′	[22]
CIDR1α	400 bp	45°C	CIDR1-5′	5′-GGT(A/T/G)(A/C/T/G)(A/C)TGATATGTTA(A/C)A(A/C)GATTC-3′	
			CIDR1-3′a	5′-T(C/T/G)TAGTAATTTATC(A/T/C)ATTGT-3′	[14]
			CIDR1-3′b	5′-T(C/T/G)TAATAAGAATTCGATTGC-3′	
DBLβ	500 bp	50°C	DBLβ-5′	5′-CGACGT(C/G)AACA(C/T)ATGTGTACATC-3′	
			DBLβ-3′	5′-CA(C/T)TC(T/G)GCCCA(C/T)TC(A/T)GTCATCC-3′	

NOTE. CIDR, cysteine-rich interdomain region; DBL, Duffy binding–like sequence; T_{anneal}, annealing temperature; UTR, untranslated region.

[a] Forward primer.

[b] Reverse primer.

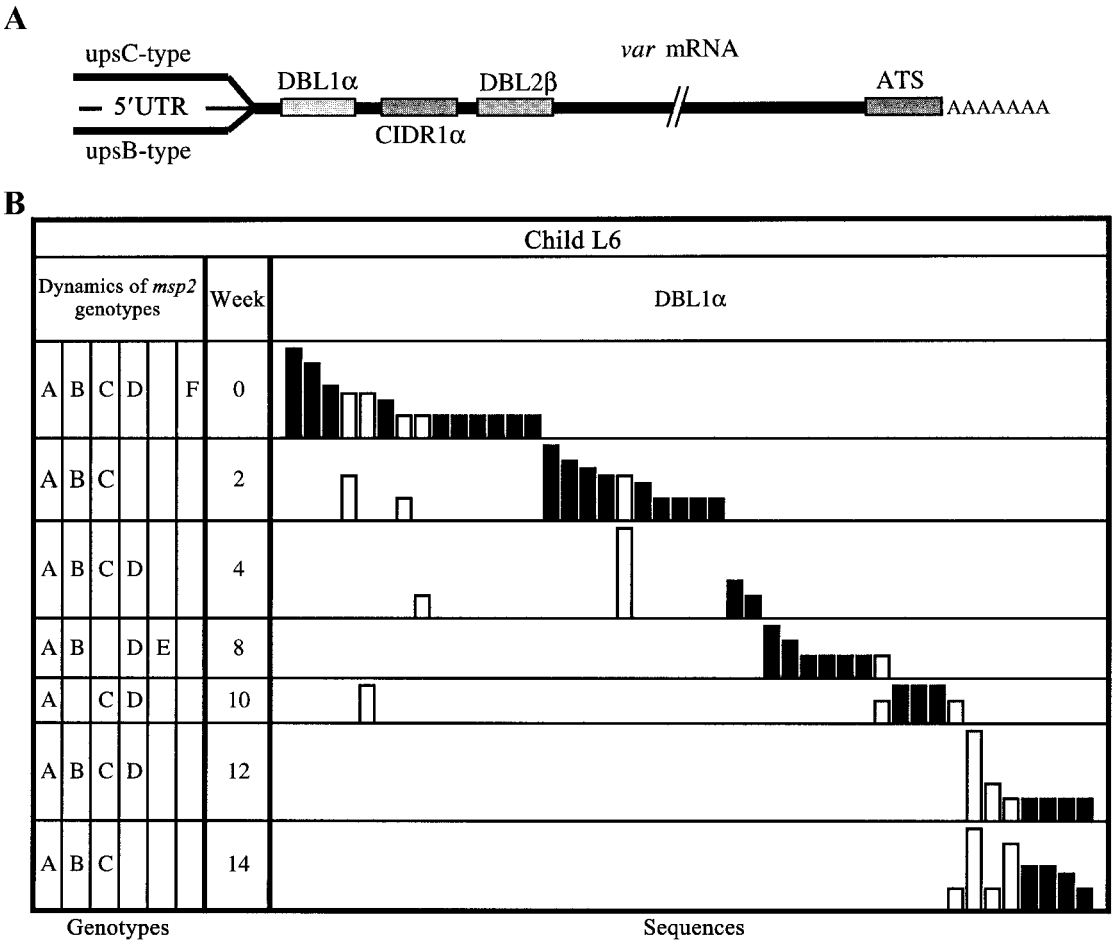


Figure 1. *A,* Schematic representation of sequenced *var* domains and upstream regions. *B,* Longitudinal distribution of Duffy binding–like 1α (DBL1α) sequences in child L6. The first major column shows the *msp2* genotyping data from the polymerase chain reaction–restriction fragment–length polymorphism analysis. The second major column indicates time points (in weeks) when samples were found to be positive for *Plasmodium falciparum* by microscopy. Different strains *(A–F)* are indicated in the minor columns. The third major column shows the longitudinal distribution of *var* gene sequences. *Bars,* individual sequences of DBL1α. Recurring sequences are indicated as white bars. The size of the bar reflects the relative frequency of a particular sequence ($\log_{10}$).

tribution of *var* sequences and *msp2* genotypes, for all time points and among all children, is shown in figure 5, which is available in the online version of this article, at the *Journal*'s Web site (http://www.journals.uchicago.edu/JID/journal/home.html).

In most children, a large number of different sequences were identified from single blood samples, and, at subsequent time points, mostly new sequences that had not been previously detected were identified. For example, for child L6, the majority of DBL1α sequences identified, which were isolated in 2-week intervals, were different at every time point (figure 1*B*). During 14 weeks, 48 different DBL1α sequences were identified. Some sequences were still present after 2 weeks or even recurred at later time points. It is important to note that child L6 was constantly infected with a large number of different *P. falciparum* strains, having 2–5 infections.

In contrast, some individuals, such as child L11, showed a less diverse *var* gene transcription, with only 1–3 different *var* transcripts identified per time point (figure 2). In child L11, all 5 domains were amplified from samples that were found to be positive for *P. falciparum*, according to microscopy, at weeks 0, 10, and 12, when the child was infected with only 1 or 2 strains. At week 10, 1 DBL1α, 1 CIDR1α, and 1 DBLβ sequence were found, suggesting that the origin of these sequences was in a single expressed *var* gene. The observed DBLβ sequence had already been observed 10 weeks earlier (figure 2). To test whether this reemerging sequence derived from 1 *var* gene or whether it was an identical domain of an otherwise different *var* gene, we amplified and sequenced genomic DNA with sequence-specific primers that targeted the domains 5′UTR-DBL1α and DBL1α-DBL2β. Identical sequences over this stretch were obtained, suggesting the presence of the identical *var* gene (data not shown).

In child L8, both upstream types (upsB-type and upsC-type

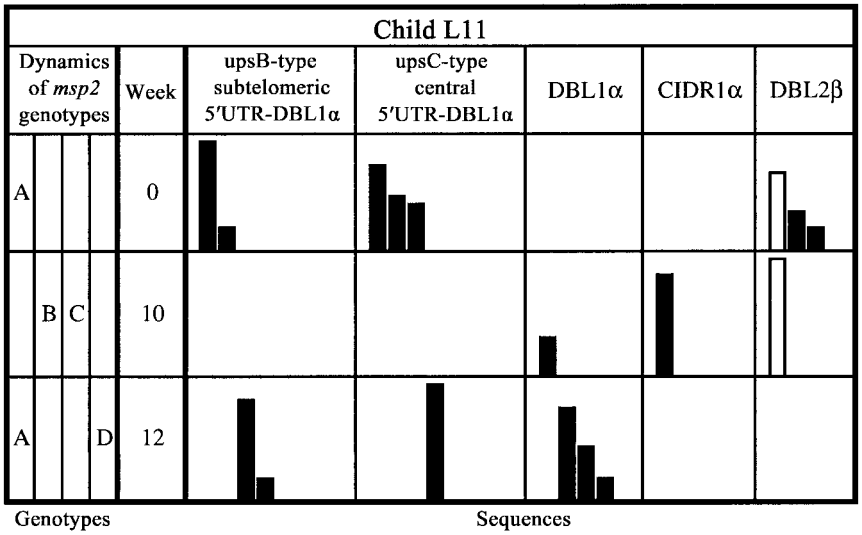


Figure 2. Longitudinal distribution of 5 different *var* gene regions in samples obtained from child L11. For details, see figure 1.

regions) were found at time points when samples were positive for *P. falciparum* by microscopy (figure 3). Between 1 and 11 different sequences of the upsB-type 5′UTR-DBL1α region were detected, whereas >1 upsC-type 5′UTR-DBL1α sequence was never found.

In child L12, identical sequences reemerged several times (figure 4). A upsC-type 5′UTR-DBL1α sequence recurred 3 times, at weeks 2, 4, and 12, and a CIDR1α sequence was found at weeks 0 and 4. Amplification of cDNA from week 2, by use of sequence-specific primers, also revealed the presence of this CIDR domain, and this indicates continuous transcription of this *var* gene for 4 weeks.

To test whether sequences from different domains of 1 sample were derived from the same *var* gene and to test the quantitative distribution of these sequences, the most abundant upsB-type 5′UTR and DBL1α sequences of child L6 (from the sample obtained at week 4) and child S12 (from the sample obtained at day 10) were linked with genomic DNA by PCR. Specific forward primers were designed in the upsB-type 5′UTR domain, and degenerated reverse primers were used for DBL1α. Subsequent sequencing showed that, in both children, the previously identified most abundant DBL1α sequence was connected with the most abundant upsB-type 5′UTR-DBL1α sequence, indicating their origin in the same *var* gene.

Distribution of* var *sequences in different children. Between 1 and 15 different sequences per domain were found in a single child. The average number of DBL1α sequences found per child was 5.2, and between 1.5 and 3.7 sequences were found for the other amplified domains (table 2).

Thirty-two identical sequences, which differed by no more than 1–3 SNPs, were found in >1 child (table 2). Twenty-three of those sequences were DBL1α domains. An additional 2 DBL1α sequences from child L3 and L8 differed by only 10 SNPs. DBLβ transcripts could be amplified from only 33% of samples in which we identified DBL1α transcripts.

Using the upstream sequence of transcribed *var* genes to determine their chromosomal location, we found that more *var* transcripts derived from subtelomerically located genes (upsB-type *var* transcript) than from centrally located genes (upsC-type *var* transcript). A mean of 1.3 upsB-type *var* gene transcripts was detected per infecting strain, compared with a mean of 0.7 upsC-type *var* gene transcripts (table 2).

Multiple infections and* var *gene transcription. Using *msp2* genotyping, we found, in 86% of all cases, multiple infections (2–5 infections), with an average of 2.8 infecting strains. When the average of 5.2 DBL1α sequences per sample was adjusted to the number of infections present, we found an average of 1.7 DBL1α sequences per infecting strain (table 2). Linear regression analysis revealed a significant correlation between the number of infecting strains and the number of DBL1α sequences identified ($P = .0002$; $r^2 = 0.31$; confidence limits, ± 0.45). In some instances, the number of *var* gene sequences observed was less than the number of infecting strains present. This finding might be ascribed either to technical limits, such as sensitivity or primer bias, which we cannot exclude while amplifying such a diverse gene family, or to the lack of the respective domain in a transcribed *var* gene (e.g., no DBLβ domain). At all time points when *var* sequences reemerged, identical strains were detected, indicating the presence of the same parasite. These identical strains were also detected in samples obtained between both time points (figures 1–4).

Analysis of* var *domain sequences. All multiple alignments

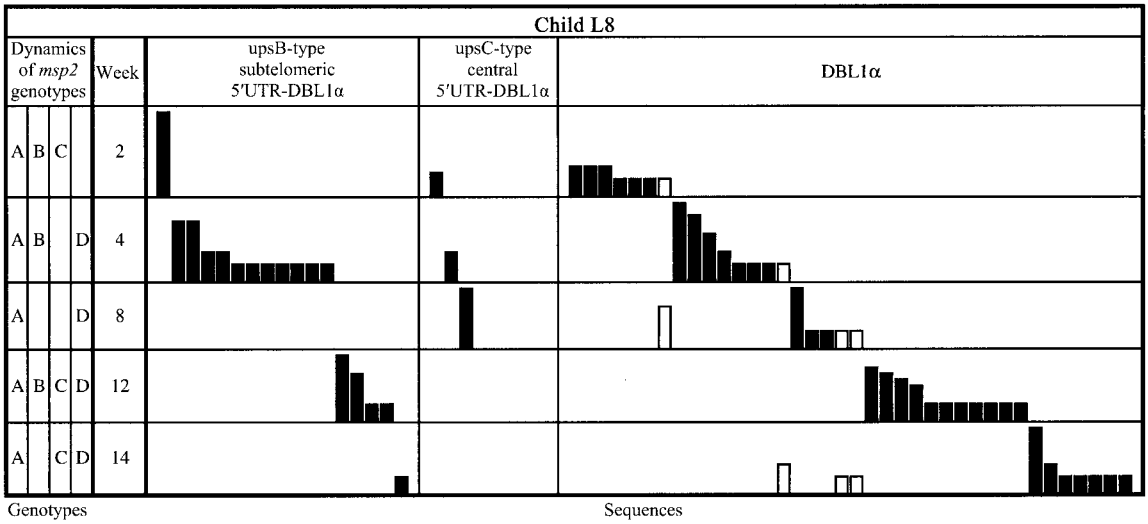


Figure 3. Longitudinal distribution of *var* gene sequences of both upstream regions and Duffy binding–like 1α (DBL1α) in samples obtained from child L8. For details, see figure 1.

of the different domains can be found in figure 6, which is available in the online version of this article, at the *Journal*'s Web site (http://www.journals.uchicago.edu/JID/journal/home.html). Multiple alignment of 150 DBL1α sequences confirmed the existence of "universally" conserved blocks (reviewed in [9]). We found clear evidence for recombination—for example, 2 sequences in samples from child S12 and child L3, which differed in the first 180 bp and which were identical after the universally well-conserved homology block F. Furthermore, for child L6, over the first 255 bp to homology block F, a DBL1α sequence from a sample obtained at week 12 was identical to another DBL1α sequence obtained at the same time point, whereas, after this 255-bp stretch, the sequence was identical to a DBL1α sequence isolated from a sample obtained from the same child 2 weeks later.

Sequences similar to the previously described DBL-α_1 subtype [26, 27] were found 40 times (25%). Also, the var_{COMMON} type [28], which has been shown to be constitutively transcribed in 60% of malaria-infected Gabonese children, was found 3 times (twice in samples from child L6 that were obtained 2 weeks apart and once in a sample obtained from child L8).

Alignment and phylogenetic analysis of the Papua New Guinea–derived DBL1α sequences with 50 previously sequenced DBL1α sequences from African *P. falciparum* strains did not show any separate clustering of PNG or African samples (data not shown). In contrast, stretches of 10–20 amino acids in the polymorphic region of DBL1α occurring in only 1 PNG sample recurred in 1 of the African samples. No geographic patterns could be detected, which is consistent with the results of other studies [29, 30].

Multiple alignments with 40 upsB-type 5′UTR-ATG sequences and 9 upsC-type 5′UTR-ATG sequences confirmed the conserved character of upsB- and upsC-type upstream sequences [16]. However, 150 bp upstream of ATG, some sequences showed deletions of up to 100 bp. One upsB-type 5′UTR-DBL1α sequence occurred in 1 child and in 3D7 parasites as *var* group B/C (PF08′0103), and it showed small differences from the characteristic upsB-type upstream sequence. It had a different length of poly(dA-dT) and homopolymeric (dA:dT) tracts and a deletion of 117 bp in an otherwise well-conserved stretch, located 250 bp upstream of the ATG.

In multiple alignments and in phylogenetic and protein-structure analysis of 38 upsB-type and 15 upsC-type *var* genes over the NTS region and the first 100 amino acids of DBL1α, no difference was observed between subtelomeric and central *var* genes. NTS and DBL1α sequences derived from either subtelomeric or central locations did not cluster separately in phylogenetic analyses or by computing population comparison tests, and the difference in this stretch of sequence between subtelomeric and central *var* genes was not significant (fixation index, 0.015; $P = .088$).

Sequences identified in field isolates and from 3D7 in the genome project. We identified 1 upsB-type 5′UTR-DBL1α sequence and 3 DBL1α sequences that were identical to *var* gene domains in 3D7 parasites. One particular 3D7 *var* DBL1α sequence was detected 3 times (once in child L13 and twice in child L6). According to PlasmoDB, this DBL1α sequence represents a *var* pseudogene (PFL1970w) with a premature stop codon 4275 bp downstream of the ATG. However, when we designed PFL1970w-specific primers to sequence over this stop codon, PCR on genomic DNA of 3D7, followed by cloning and sequencing of 10 clones, revealed an insertion relative to the 3D7

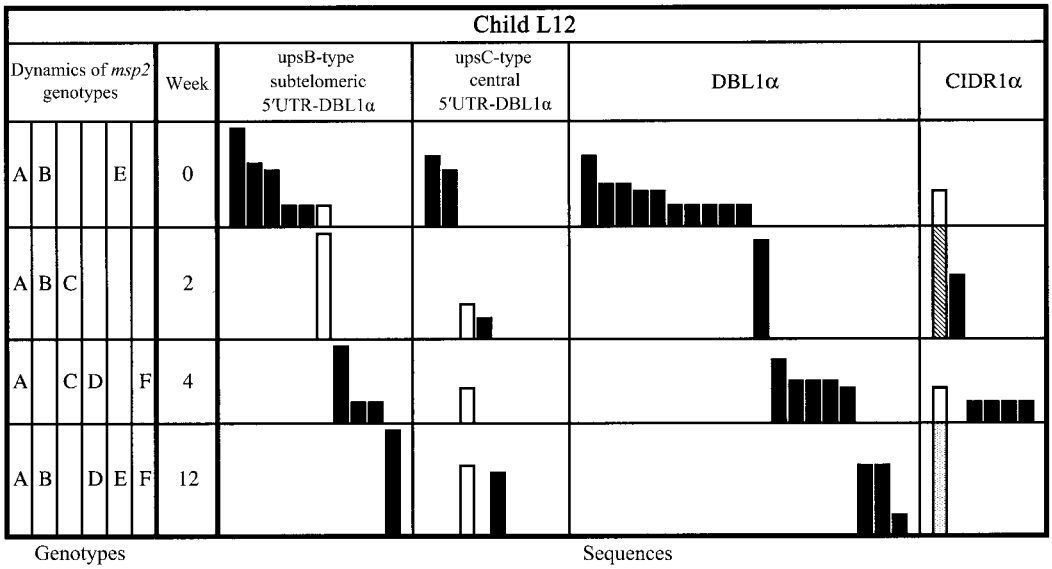


Figure 4. Longitudinal distribution of both upstream regions, Duffy binding–like sequence 1α (DBL1α), and cysteine-rich interdomain region–α (CIDRα) in samples obtained from child L12. The hatched and dotted bars in the "CIDR1α" panel denote 1 CIDR1α clone that was detected at weeks 0 and 4 *(white bars)* and that was also detected by sequence-specific primers on cDNA at week 2 *(hatched bar)* and on genomic DNA at week 12 *(dotted bar).* For details, see figure 1.

sequence producing a frameshift mutation (4233 bp downstream of ATG), resulting in a continuous open-reading frame.

DISCUSSION

During the past years, much information on *var* gene transcription has been gained; however, most of this information has been based on the findings of in vitro studies, and few studies have looked at *var* gene transcription in vivo [24, 26, 31]. To our knowledge, this is the first longitudinal study of *var* gene transcription in naturally infected children. We generated cDNA and cloned and sequenced 3 adhesive *var* gene domains (DBL1α, CIDR1, and DBLβ) and 2 *var* 5′UTR stretches (upsB-type and upsC-type regions) from blood samples obtained, over 4 months, from asymptomatic children living in an area where malaria is endemic. Using this approach, we identified a large number of different sequences, and we observed a dynamic picture of *var* gene transcription. However, despite such a dynamic transcription pattern, some sequences persisted or recurred for up to 10 weeks.

It has often been argued that RT-PCR could identify the smallest amounts of RNA, and, in the case of *var* gene transcription, it is indeed unclear whether cDNA represents only functional full-length mRNAs, because incomplete and 3′-truncated *var* transcripts have been observed [22]. By selecting *var* transcripts that contained the 3′ATS domain before RT-PCR was performed, we are confident that the number of incomplete mRNAs was reduced to insignificant levels. This was confirmed by the use of tags, other than the anti-ATS, that resulted in no product at all (data not shown). Furthermore, when Peters et al. [24] compared the number of *var* transcripts from 3D7 between the ring stage and the trophozoite stage, they observed no difference in the number and proportion of transcripts, and they even questioned the relaxed transcription in vivo. Moreover, by use of single-cell RT-PCR for trophozoite-stage cells, up to 5 different transcripts were observed in 3D7 [32], which questions the hypothesis of mutually exclusive *var* gene transcription that was previously suggested by results of cultures selected for receptor binding [20, 33].

We observed the largest diversity within the DBL1α sequences, with an average of 5.2 different sequences per blood sample. This finding is similar to previous reports of 3–15 different *var* gene transcripts in natural infections or from laboratory-adapted strains [22, 24, 31, 32, 34]. One explanation for the large number of observed *var* transcripts at 1 time point might be the presence of multiple concurrent *P. falciparum* infections. It is noteworthy that the mean number of *var* gene transcripts per sample very much resembles the mean number of concurrent infecting strains in this age group [35]. It has been speculated that multiple infections provide protection against hyperinfection by stimulating the immune system with a broad range of diverse antigens, such as PfEMP1 [35, 36]. Concomitantly expressed PfEMP1 variants could also explain the findings of Bull et al. [37], who showed that children with asymptomatic infections had a greater repertoire of variant-specific antibodies.

Most children had multiple infections, and multiple *var* transcripts were detected in these children. This clearly added to the complexity of the observed *var* gene expression dynamic.

Table 2. Overview of analyzed sequences of different transcribed *var* gene regions.

General overview of the no. of sequences	*var* Gene regions				
	Subtelomeric upsB-type 5′UTR-DBL1α	Central upsC-type 5′UTR-DBL1α	DBL1α	DBLβ	CIDR1α
No. of sequences (no. of positive blood samples)	417 (20)	133 (17)	789 (39)	97 (13)	118 (8)
No. of different sequences	66	23	158	22	17
Average no. of different sequences per child and time point	3.7	1.5	5.2	2.0	3.0
Average no. of different sequences per infecting strain, child, and time point	1.3	0.7	1.7	0.8	1.7
No. of transcribed different sequences in 3D7 culture strain	7	1	8	5	ND
No. of recurring sequences in the same child					
After 10 days	2	…	2	…	…
After 2 weeks	2	2	7	…	…
After 4 weeks	…	…	2	…	1
After 5 weeks	1	…	…	…	…
After 6 weeks	…	…	4	…	…
After 8 weeks	…	1	…	…	…
After 10 weeks	…	1	2	1	…
No. of identical sequences in different children					
In 2 children	3	…	16	4	…
In 3 children	…	1	4	…	…
In 3D7 culture strain	1	…	3[a]	…	…

NOTE. CIDR, cysteine-rich interdomain region; DBL, Duffy binding–like sequence; ND, not done; UTR, untranslated region.

[a] One of these 3 sequences was detected in 2 children.

However, in samples obtained from children who had single infections, only 1 DBL1α sequence was found to be abundant or present. There was a linear correlation between the number of detected transcripts and the number of infecting strains present. The relatively low number of *var* transcripts in children who had a single infection suggests a tight regulation that allowed the transcription of only 1 or a few *var* genes at 1 time in the clonal parasite population. However, the observation of rapidly changing *var* transcripts suggests high switching rates. Peters et al. [24] reported different switching rates between initial and subsequent switching events in laboratory-induced infections. Our data suggest different switching rates for some *var* genes that were transcribed for only a short period, whereas others persist for weeks or recur. In the present study, we cannot distinguish between different switching rates or selection forces acting on previously expressed PfEMP1 molecules. We also cannot completely rule out that recurring *var* gene sequences are derived from a closely related parasite strain that expresses a highly similar *var* gene. However, the detection of these recurring *var* sequences at intermediate time points (figure 4) and the presence of the same *P. falciparum* strain suggest that these *var* sequences are derived from a constantly expressed *var* gene.

Although we have no quantitative data for the detected *var* transcripts, we were able to semiquantify the presence of *var* transcripts, to a certain degree, and to show that the most abundant sequences of 2 *var* regions in the same probe originated in 1 *var* gene (see the "*var* Gene Transcription in Individual Children" subsection in Results). We are currently addressing this question in ongoing case-control studies that use real-time PCR to quantify *var* transcripts, with respect to the various groups of upstream regions.

The previously described conserved character of dimorphic *var* gene upstream regions [16] and the different regulation mechanisms of these *var* gene groups [19] led us to speculate that they might be functionally different. Therefore, we amplified both 5′UTR stretches and the adjacent NTS-DBL1α sequences, and we then compared expression dynamics and coding sequence. Of all analyzed 5′UTR sequences, 26% were of the upsC type, which reflects the proportion of upsC-type *var* genes in the 3D7 genome, which is 22%. We never detected expression of >3 different centrally located *var* genes in a sample. We had speculated that centrally located *var* genes would be more conserved because of a location that was less prone to recombination. This is supported by the findings of Rubio et al. [38], who showed that centrally located *var* genes were more related to each other than to subtelomerically located *var* genes. Consequently, we hypothesized that upsC-type *var* genes would be recognized faster by the immune system. However, in the present study, upsC-type sequences were found to persist for 8–10 weeks, suggesting that these *var* genes can be also

expressed over a long time. Furthermore, we were unable to show any structural or phylogenetic differences within the short sequence of the NTS-DBL1α domain of both *var* gene groups. However, this stretch of sequence is known to be conserved, represents only a limited stretch of the whole *var* gene, and does not allow for further conclusions to be made.

In summary, we have shown that chronically infected children express several *var* genes simultaneously, with a mean of 1.7 different *var* genes per infecting strain. The repertoire of *var* genes circulating in a parasite population seems to be large and recombinogenic. We have shown a rapid change in *var* gene expression, but we also detected, for the first time, recurring *var* genes after 10 weeks.

Expression of many and rapidly changing *var* genes is expected in asymptomatic children, in whom parasite survival is a balance between antigenic escape and binding capacity to avoid splenic clearance. These children must be considered to be semi-immune, and the *var* gene repertoire of the infecting parasites might already be exhausted. Therefore, the parasite might be forced to rapidly switch to new *var* variants. To unequivocally understand *var* gene expression in vivo, studies involving naive individuals must be conducted. Furthermore, comparing *var* gene expression in subjects who have severe malaria with that in subjects who have mild malaria should shed more light on the complex cascade of *var* gene expression and an immunological escape mechanism.

Acknowledgments

We thank the staff of the Papua New Guinea Institute of Medical Research, Madang, for their field assistance. We appreciate the cooperation of the parents and children of the Maiwara Primary School for participating in the study. We are very grateful to Dorothee Foernzler and Patricia Riegert for sequencing, Sophia Kossida for assistance in phylogenetic analysis, and Matthias Rottmann and Christian Flueck for critical reading of the manuscript.

References

1. Su XZ, Heatwole VM, Wertheimer SP, et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*–infected erythrocytes. Cell **1995**; 82:89–100.
2. Smith JD, Chitnis CE, Craig AG, et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell **1995**; 82:101–10.
3. Baruch DI, Pasloske BL, Singh HB, et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell **1995**; 82:77–87.
4. Kyes S, Horrocks P, Newbold C. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol **2001**; 55:673–707.
5. Urban BC, Ferguson DJ, Pain A, et al. *Plasmodium falciparum*–infected erythrocytes modulate the maturation of dendritic cells. Nature **1999**; 400:73–7.
6. Rowe JA, Moulds JM, Newbold CI, Miller LH. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Nature **1997**; 388:292–5.
7. Biggs BA, Anders RF, Dillon HE, et al. Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. J Immunol **1992**; 149:2047–54.
8. Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med **1998**; 4:358–60.
9. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. Mol Biochem Parasitol **2000**; 110:293–310.
10. Krych-Goldberg M, Moulds JM, Atkinson JP. Human complement receptor type 1 (CR1) binds to a major malarial adhesin. Trends Mol Med **2002**; 8:531–7.
11. Smith JD, Craig AG, Kriek N, et al. Identification of a *Plasmodium falciparum* intercellular adhesion molecule–1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci USA **2000**; 97:1766–71.
12. Buffet PA, Gamain B, Scheidig C, et al. *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. Proc Natl Acad Sci USA **1999**; 96:12743–8.
13. Reeder JC, Cowman AF, Davern KM, et al. The adhesion of *Plasmodium falciparum*–infected erythrocytes to chondroitin sulfate A is mediated by *P. falciparum* erythrocyte membrane protein 1. Proc Natl Acad Sci USA **1999**; 96:5198–202.
14. Degen R, Weiss N, Beck HP. *Plasmodium falciparum:* cloned and expressed CIDR domains of PfEMP1 bind to chondroitin sulfate A. Exp Parasitol **2000**; 95:113–21.
15. Baruch DI, Ma XC, Singh HB, Bi X, Pasloske BL, Howard RJ. Identification of a region of PfEMP1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36: conserved function with variant sequence. Blood **1997**; 90:3766–75.
16. Voss TS, Thompson JK, Waterkeyn J, et al. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5′ flanking sequences. Mol Biochem Parasitol **2000**; 107:103–15.
17. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature **2002**; 419:498–511.
18. Scherf A, Figueiredo LM, Freitas-Junior LH. *Plasmodium* telomeres: a pathogen's perspective. Curr Opin Microbiol **2001**; 4:409–14.
19. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP. Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters. Mol Microbiol **2003**; 48:1593–607.
20. Scherf A, Hernandez-Rivas R, Buffet P, et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J **1998**; 17:5418–26.
21. Deitsch KW, Calderwood MS, Wellems TE. Malaria: cooperative silencing elements in *var* genes. Nature **2001**; 412:875–6.
22. Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI. A study of *var* gene transcription in vitro using universal *var* gene primers. Mol Biochem Parasitol **2000**; 105:13–23.
23. Roberts DJ, Craig AG, Berendt AR, et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature **1992**; 357:689–92.
24. Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci USA **2002**; 99:10689–94.
25. Felger I, Tavul L, Beck HP. *Plasmodium falciparum:* a rapid technique for genotyping the merozoite surface protein 2. Exp Parasitol **1993**; 77:372–5.
26. Kirchgatter K, Portillo HA. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med **2002**; 8:16–23.

27. Robinson BA, Welch TL, Smith JD. Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol **2003**; 47:1265–78.
28. Winter G, Chen Q, Flick K, Kremsner P, Fernandez V, Wahlgren M. The 3D7var5.2 (*var*COMMON) type *var* gene family is commonly expressed in non-placental *Plasmodium falciparum* malaria. Mol Biochem Parasitol **2003**; 127:179–91.
29. Kyes S, Taylor H, Craig A, Marsh K, Newbold C. Genomic representation of *var* gene sequences in *Plasmodium falciparum* field isolates from different geographic regions. Mol Biochem Parasitol **1997**; 87:235–8.
30. Kirchgatter K, Mosbach R, del Portillo HA. *Plasmodium falciparum:* DBL-1 *var* sequence analysis in field isolates from central Brazil. Exp Parasitol **2000**; 95:154–7.
31. Afonso NP, Wunderlich G, Shugiro TM, et al. *Plasmodium falciparum:* analysis of transcribed *var* gene sequences in natural isolates from the Brazilian Amazon region. Exp Parasitol **2002**; 101:111–20.
32. Duffy MF, Brown GV, Basuki W, et al. Transcription of multiple *var* genes by individual, trophozoite-stage *Plasmodium falciparum* cells expressing a chondroitin sulphate A binding phenotype. Mol Microbiol **2002**; 43:1285–93.
33. Chen Q, Fernandez V, Sundstrom A, et al. Developmental selection of *var* gene expression in *Plasmodium falciparum*. Nature **1998**; 394:392–5.
34. Noviyanti R, Brown GV, Wickham ME, Duffy MF, Cowman AF, Reeder JC. Multiple *var* gene transcripts are expressed in *Plasmodium falciparum* infected erythrocytes selected for adhesion. Mol Biochem Parasitol **2001**; 114:227–37.
35. Smith T, Felger I, Tanner M, Beck HP. Premunition in *Plasmodium falciparum* infection: insights from the epidemiology of multiple infections. Trans R Soc Trop Med Hyg **1999**; 93(Suppl 1):59–64.
36. al Yaman F, Genton B, Reeder JC, Anders RF, Smith T, Alpers MP. Reduced risk of clinical malaria in children infected with multiple clones of *Plasmodium falciparum* in a highly endemic area: a prospective community study. Trans R Soc Trop Med Hyg **1997**; 91:602–5.
37. Bull PC, Lowe BS, Kaleli N, et al. *Plasmodium falciparum* infections are associated with agglutinating antibodies to parasite-infected erythrocyte surface antigens among healthy Kenyan children. J Infect Dis **2002**; 185:1688–91.
38. Rubio JP, Thompson JK, Cowman AF. The *var* genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. EMBO J **1996**; 15:4069–77.

3. Differential expression of *Plasmodium falciparum var* gene subgroups is associated with virulence in a malaria case-control study of children in Papua New Guinea.

Mirjam Kaestli [1], Ian Cockburn [2,a], Alfred Cortes [3,b], Kay Baea [3], Alex Rowe [2], Hans-Peter Beck [1]

Affiliations of authors:

[1] Swiss Tropical Institute, Socinstr.57, Postfach, 4002 Basel, Switzerland

[2] Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Rd, Edinburgh EH9 3JT, United Kingdom

[3] PNG Institute of Medical Research, PO Box 378, Madang, MP 511, Papua New Guinea

Submitted to *Blood* (December 2004)

Differential expression of *Plasmodium falciparum var* gene subgroups is associated with virulence in a malaria case-control study of children in Papua New Guinea.

Mirjam Kaestli [1], Ian Cockburn [2,a], Alfred Cortés [3,b], Kay Baea [3], Alex Rowe [2], Hans-Peter Beck [1]

Running head: *var* transcription in a malaria case-control study

This work was supported by the Swiss National Science Foundation (grant number 031-67211.01 and 031-104043/1).

Affiliations of authors:

[1] Swiss Tropical Institute, Socinstr.57, Postfach, 4002 Basel, Switzerland

[2] Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Rd, Edinburgh EH9 3JT, United Kingdom

[3] PNG Institute of Medical Research, PO Box 378, Madang, MP 511, Papua New Guinea

Present address:

[a] Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, Baltimore MD 21205-2179, USA

[b] MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom

Reprints or correspondence:

Hans-Peter Beck, Swiss Tropical Institute, Socinstrasse 57, CH-4051 Basel, Switzerland
Tel: +41-61-284 8116
Fax: +41-61-271 8654
e-mail: hans-peter.beck@unibas.ch

Total text word count: 4722

Abstract word count: 187

Scientific heading: Clinical observations, interventions, and therapeutic trials

Abstract

Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) is considered to be a major pathogenicity factor in malaria by mediating cytoadherence. PfEMP1 is encoded by approximately 60 *var* genes per haploid genome. Most *var* genes can be grouped into 3 subgroups A, B and C according to their, within groups, conserved non-coding upstream region. This raises questions about the clinical significance of these structural differences. Using real-time quantitative PCR in a case-control study conducted in Papua New Guinea, we compared the distribution of *var* gene transcripts of *var* groups A, B and C among children with severe, mild and asymptomatic malaria. We found a significant up-regulation of *var* group B transcripts in children with clinical malaria (mild and severe), whereas *var* group C genes were mainly switched on in asymptomatic children. Furthermore, rosetting parasites showed a significant up-regulation of *var* group A transcripts. Our data suggest that a major difference in *var* gene expression exists between parasites causing clinical attack and parasites leading to asymptomatic infection. These findings reflect the pathological consequences of the differential expression of *var* genes which further emphasizes its substantial implication for malaria morbidity.

Introduction

In malaria endemic areas of Papua New Guinea, *Plasmodium falciparum* accounts for 4-13% of deaths in children. [1] There are various factors contributing to severe malaria pathology including cytoadherence. [2] This is the adhesion of late stage infected erythrocytes to various receptors such as CD36 or ICAM1 on the vascular endothelium, CSA in the placenta and CR1 on red cells [3] leading to microvascular obstructions in various organs. Cytoadherence is mainly mediated by the parasite derived polymorphic *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) family. [4-6] These are large antigenically variant proteins (200-350 kDa) on the surface of late stage infected erythrocytes (IE). PfEMP1 are structured into different semi-conserved, adhesive domains namely duffy binding like (DBL) domains and cysteine rich interdomain regions (CIDR) (for a review, see Smith[7]).

The sequencing of the laboratory-adapted *P. falciparum* line 3D7 enabled scientists to look at the entire *var* gene repertoire of a single strain. In these parasites, PfEMP1 is encoded by one of 59 *var* genes, each of 8 to 14 kb of length. [8] Most *var* genes can be subgrouped into 3 types (*var* groups A, B, and C) mainly according to their conserved 5' upstream sequences. [8,9] In 3D7, the majority of *var* genes belongs to the subtelomerically located *var* group B whereas a smaller set of 13 *var* genes called *var* group C are arranged in chromosome internal clusters. 10 larger, subtelomerically located *var* genes with a distinct domain structure belong to *var* group A. [10]

Many attempts have been made to find different pathological implications in the binding of IE to distinct receptors. Various studies have shown that numerous PfEMP1[11] [12] and most field isolates [13,14] bind to CD36 which was correlated with non-severe malaria. [13] While adhesion of IE to CSA was associated with pregnancy associated malaria, [15,16] binding to ICAM1 was correlated with cerebral malaria. [17] Rosetting, the binding of IE to uninfected erythrocytes, was shown to associate with severe disease in a variety of studies in Africa. [18,19] However, less work has been done to correlate the expression of distinct PfEMP1 variants with disease severity. Bull et al.[20], found by agglutination assays that parasites of hosts with a low antibody repertoire against variant surface antigens (VSA) express a virulent subset of VSAs which is prevalent in the community whereas individuals with a larger repertoire of anti-VSA responses are infected by parasites expressing rarer VSAs. The question arises if this restricted virulent VSA subset coincides with a subgroup of *var* genes, namely *var* groups A, B or C. First indications that this could indeed be the case

were the findings of larger PfEMP1 on the surface of IE of children with cerebral malaria [21] and a subgroup of *var* DBL1α sequences isolated of parasites of children with severe malaria. [22] This indicated the involvement of *var* group A genes which was further confirmed by the findings of Jensen et al. [23] who *in vitro* selected 3D7 for severe malaria antigenic properties and found an up-regulation of several *var* group A genes.

To further study the association between the expression of *var* groups A, B and C and pathologocial outcome *in vivo*, we enrolled 65 children in a malaria case-control study in Madang, Papua New Guinea (PNG), in February to May 2003. The relative contribution of *var* gene transcripts among *var* groups A, B or C was evaluated by quantitative real-time PCR and compared between children with severe malaria, children with mild malaria, and asymptomatic children. To our knowledge, this is the first study to evaluate and compare the expression of *var* gene subgroups and clinical outcome *in vivo*.

Methods and Materials

Population and study design

From February to May 2003, during the wet season, we performed a severe malaria case-control study of children admitted to the Modilon General Hospital of Madang, Papua New Guinea (PNG). During 1994-1995, malaria accounted for 15.3% of deaths in children in this hospital with a severe malaria case fatality rate of 3.6%. [24] The Madang area is holoendemic for malaria with perennial transmission, and infections with *P. falciparum* and *P. vivax* are common. *P. malariae* and *P. ovale* are also present in this area. This study was approved and ethical clearance was given by the Medical Research Advisory Committee of Papua New Guinea. Children at an age of 0.5-6y were defined as cases if they had clinical manifestations of severe malaria according to the World Health Organization (WHO) criteria for severe and complicated malaria (2000). [2] Together with the presence of asexual *P. falciparum* parasites, these included criteria such as cerebral malaria, severe anemia (haemoglobin < 5 g/dl), prostration, metabolic acidosis, i.e. respiratory distress and hyperlactataemia (lactate > 5 mmol/L), and hypoglycaemia (glucose < 2.2 mmol/L). At admission, full clinical histories were recorded, an examination was performed and biochemical tests were done. Exclusion criteria were the confirmed diagnosis of a co-infection with any other disease, malnutrition (mid upper arm circumference (MUAC) < 12 cm) or an antimalarial treatment in the last 2 weeks. After informed consent was obtained from the parents, venous blood (1 ml) was taken from 14 children with severe malaria (age range 8-60 months, mean 35.8 months) admitted to the hospital. None of these cases were fatal. 26 children of similar age (+/- 20% of age) with mild malaria were enrolled at the hospital and at the Town clinic, Madang. Mild malaria was defined as the presence of asexual *P. falciparum* and an axillary temperature ≥ 37.5 ˚C or in the absence of the latter, symptoms or a history of headache, fever or myalgia. We also enrolled 25 asymptomatic children who were positive for *P. falciparum* by OptiMAL® test and subsequent microscopy and who age and location matched to the children with severe malaria.

Assessment of P.falciparum infections

Thick blood films were stained with Giemsa and the number of malaria parasites per 200 white blood cells was counted. Parasite densities were converted to parasites per µl by

multiplication with 40 under the assumption of 8000 white blood cells by µl. In samples positive by microscopy for *P. falciparum*, the number of *P. falciparum* infections was determined by *msp2* genotyping as described previously. [25] Briefly, 30µl of full blood was spotted on filter papers (Isocode Stix; Schleicher&Schuell) and dried for 20min at 80˚C. After washing, *msp2*-PCR was performed directly on the filter papers and nested PCR products were analysed by RFLP to record the number of infecting strains.

Parasite culture and assessment of rosetting frequency

If parasitemia was above 5000 parasites/µl, parasites of children with severe or mild malaria were cultured according to standard methods in 10% heterologous AB serum. [18] When most parasites were at the late trophozoite / early schizont stage, rosettes were counted as follows. An aliquot of culture at 2% haematocrit was stained with ethidium bromide and by using a fluorescence microscope, rosettes were counted as a proportion of 200 mature stage parasites. Rosettes were defined as the binding of an infected RBC to at least 2 uninfected RBCs. The rosette frequency is indicated as the percentage of mature-parasite-infected cells found in rosettes.

Isolation of full-length var transcripts and cDNA Synthesis

Isolation of full-length *var* mRNA and reverse transcription was done as described elsewhere. [26] Briefly, total RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's instructions. RNA was treated with 3U of RQ1 DNase (Promega). To obtain only full length *var* transcripts, RNA was incubated with 1pmol of biotinylated oligonucleotide complementary to the ATS domain. 200µg Dynal beads M-280 Streptavidin were added to the RNA. After washing, reverse transcription (RT) was performed on the captured hybrids, primed by 700ng hexamers (Invitrogen) and using Sensiscript (Qiagen) reverse transcriptase according to the manufacturer's protocol in a final volume of 20µl.

An aliquot without reverse transcriptase was used as negative control. After RT, cDNA was RNase A treated.

Isolation of genomic DNA

DNA extraction was done as described elsewhere. [27] Briefly, 30µl of full blood was spotted on filter papers (Isocode Stix; Schleicher&Schuell) and dried for 20min at 80˚C. After

washing according to the supplier's instructions, PCR was performed directly on the filter papers.

Quantitative Real-time PCR

Prior to Real-time PCR, 1µl of DNA was amplified in a primary PCR to enrich material. Over *var* 5'UTR-DBL1α, DNA was amplified in 50µl volumes with Advantage cDNA polymerase (Clontech) using 400 nM *var* group specific 5'UTR forward primers and a degenerated DBL1α reverse primer (see Table 1). PCR conditions were 94˚C for 5min and 16 cycles (for cDNA) or 14 cycles (for gDNA) of 95˚C for 30sec / 52˚C for 1min / 64˚C for 1min10sec. Real-time PCR was performed over *var* group A, B and C 5'UTR using a ABI PRISM 7000 Sequence Detection System (Applied Biosystems). Reactions were done with 5µl of primary PCR product in 25µl volumes with Advantage cDNA polymerase (Clontech) using 250 nM minor groove binder (MGB) probes labelled with FAM (Applied Biosystems) and 900 nM primers for the respective sequences (see Table 1). Real-time oligos were designed according to alignments of 5'UTR *var* gene sequences of the 3D7 genome project (www.plasmodb.org; Joe Smith, Seattle Biomedical Research Institute, oral communication, May 6, 2004) and *var* gene sequences of Papua New Guinea (AY462581-AY462851). PCR conditions were 94˚C for 5min and 40 cycles of 95˚C for 30sec / 54˚C for 1min / 65˚C for 1min10sec. Electrophoresis of real-time PCR products was performed to control for single bands and equal size. All cDNA samples were run in triplicates. If all Ct values of *var* group A, B and C were above 31, the sample was discarded. Negative cDNA controls (no reverse transcriptase) of all samples and No Template Controls (NTC) (per 96well plate) were amplified in parallel. If the NTC was positive, the plate was discarded. If the negative cDNA control was positive, the sample was discarded. As positive control and plate calibrator, 2.5ng of gDNA of *P. falciparum* 3D7 were amplified in parallel per plate and *var* group. Quantification was done using ABI Prism 7000 SDS Software, version 1.1.

Standard curve and relative quantification

Standard curves were generated using a dilution series over 6 logs of 10-14 different dilutions, each dilution in triplicates. The PCR efficiency (E) was calculated using the formula, $E = 10^{(1/-slope)} - 1$. The mean efficiencies of 3 independent standard curves with high reproducibility were 100% for *var* group A, 86% for *var* group B and 95% for *var* group C. The detection limit of the system was below 50 copies (data not shown). Relative

quantification was done according to the $\Delta\Delta C_T$ method (Application guide, Qiagen AG) with following modifications: Ct values were converted to approximate copy numbers by the formula $C/E^{\Delta Ct}$, where "C" is the number of *var* gene copies in the corresponding *var* groups A, B or C of the plate calibrator (2.5ng of 3D7 gDNA). "E" is the real-time PCR efficiency of the corresponding *var* group (*var* group A, 2; group B, 1.86 and group C, 1.95) and ΔC_T is the difference of average C_T values between the sample and the corresponding *var* group plate calibrator. The numbers of *var* copies in the plate calibrator ("C") were estimated by comparing the real-time PCR oligo sequences (see Table 1) and 5'UTR *var* gene alignments (see above). Due to the *var* specific mRNA isolation, no endogenous reference gene could be used. For the statistical analysis, we used proportions of *var* group transcripts.

Verification of real-time PCR

To test the accuracy of our approach, AT Jensen kindly provided us with RNA isolated of *P. falciparum* 3D7 parasites which were *in vitro* selected for severe malaria antigenic properties and the corresponding unselected controls. [23] Jensen et al. compared *var* gene transcription in these strains by performing *var* gene-specific quantitative real-time PCR. Our results of real-time PCR, which mainly consisted of an up-regulation of *var* group A and slight down-regulation of *var* group C genes (data not shown), agreed with the findings of Jensen et al.

Statistical Analyses

Statistical analysis was performed using Stata (Intercooled Stata, version 8.2; www.stata.com). Statistical analyses were done on transcript proportions of *var* groups, i.e. on the number of transcripts of one *var* group over the sum of transcript numbers of *var* groups A, B and C. Associations between *var* group proportions and clinical outcome were analysed using Mann-Whitney U test. Logistic regression analyses and likelihood ratio tests were performed to calculate the odds ratio (OR) for disease. The ORs were calculated and compared unadjusted or in multivariate logistic regression analyses, with adjustment for parasitemia. There was no significant difference between the age and sex distribution of cases and controls (age and clinical outcome, $P = 0.6$; sex and clinical outcome, $P = 0.72$). However, parasitemia showed a significant association with "clinical outcome" ($P < 0.001$), but showed no association with ratios of *var* groups (Spearman rank $P = 0.3$). There was no significant difference between clinical outcome and collection site ($P = 0.9$)

or ethnicity of children (P = 0.27). All tests were 2-tailed and considered significant if P values were less than 0.05.

Results

65 children were enrolled in a severe malaria case-control study in Papua New Guinea (see Table 2). By real-time PCR, the distribution of *var* gene transcripts among *var* groups A, B or C was compared in 14 children with severe malaria, 26 children with mild malaria and 25 asymptomatic children.

Analyses were done on proportions of transcripts of one *var* group over the sum of all amplified transcripts of *var* groups A, B and C.

var group B transcripts are upregulated in clinical malaria

A significant increase of the proportion of *var* group B transcripts was found in children with clinical malaria (mild and severe), compared to children with asymptomatic malaria (odds ratios in Table 3) (see Figure 1). This was confirmed by the observation that both fever and headache also correlated significantly with more *var* group B transcripts. There was a further increase of *var* group B proportions in severe malaria compared to mild malaria, however this was not significant probably due to the small number of children with severe malaria. No difference in *var* group B proportions was found between children with cerebral malaria and severe malaria cases without cerebral involvement nor between severe anemic children and not anemic children with severe malaria (data not shown).

var group C transcripts are mainly transcribed in asymptomatic malaria

var group C proportions were below 10% of all detected *var* transcripts in 33 of 40 (83%) children with clinical malaria compared to only 10 of 25 (40%) asymptomatic children. In only one child of all children with clinical malaria, more *var* group C than *var* group B transcripts were found. However, this child had a high white blood cell count and only 160 parasites/µl indicating another infection and cause for the disease. Conversely, in 11 of 25 asymptomatic children, more *var* group C than *var* group B transcripts were found. The increase of *var* group C proportions in asymptomatic children was highly significant compared to *var* group C proportions in children with mild disease while it was not significant compared to children with severe malaria despite of a low odds ratio. (see Table 3, Figure 1). There was no significant change in *var* group C specific transcript numbers between children with mild or severe malaria.

Increased var group A transcription in parasites conferring rosetting

Isolates with rosetting frequencies >10% were found in 33% (10/30) of parasites from children with mild or severe disease. The occurrence of the rosetting phenotype did not correlate with the severity of disease. However, significantly more *var* group A transcripts were detected in parasites with rosetting frequencies > 10% with a 3 fold increase of the median and 13 fold increase of the geometric mean of *var* group A proportions (Mann-Witney U test, P = 0.012).

var group A transcripts were less amplified than *var* group B or C transcripts. There was no significant difference of *var* group A proportions in children with severe malaria compared to children with mild malaria and there was only a slight decrease of *var* group A proportions in children with asymptomatic malaria compared to children with mild malaria. However, due to outliers of increased *var* group A proportions in asymptomatic controls, this decrease was only significant by performing a median test (P = 0.006, continuity corrected).

Quantitative analysis of genomic DNA of var groups A, B and C

To test for the equal presence of *var* subgroups among cases and controls and to exclude primer bias, quantitative PCR analyses of *var* groups A, B and C were also performed on genomic DNA of most samples. Numbers of amplified *var* group templates correlated well with parasite loads of the corresponding children (P = 0; Spearman's rho of *var* group A = 0.72, group B = 0.84, group C = 0.77). The distribution of the 3 *var* subgroups was very similar among different clinical outcomes. On average, the amplified *var* genes were distributed among *var* groups with 10% *var* group A, 80% *var* group B and 10% *var* group C genes. However, in 1 of 59 samples, proportions of *var* group C genes were below 5% while in 5 of 59 samples, proportions of *var* group A genes were below 5% indicating the absence of *var* genes or *var* gene stretches to which the primers were targeted in these parasites.

Multiple infections and var gene transcription

Using *msp2* genotyping, we found multiple infections (2 to 3 infections) in only 32% of all children, with an average of 1.4 infecting strains. There was no association between the number of infecting *P. falciparum* strains and clinical outcome.

Association between parasitemia and proportions of var groups A, B and C

In asymptomatic children, we found a significant positive association between parasitemia and *var* group C proportions (Spearman's rho 0.51, P = 0.013), i.e. more parasites were detected in children with increased *var* group C proportions (>0.4). In contrast, in children with clinical disease, there was a negative, but not significant association between parasitemia and *var* group C proportions. Regarding *var* group A and B proportions, these correlations were exactly reversed but not significant.

Discussion

The intriguing observation of distinct *var* gene subgroups and previous reports of disease associated *var* genes called for further studies on natural infected children. To identify potential differences in virulence associated with the expression of either *var* groups A, B or C we performed a case-control study on children with severe, mild and asymptomatic malaria. By quantitative real-time PCR, we found a significant increase of parasites transcribing *var* group B genes in clinical malaria whereas parasites of asymptomatic children transcribed more often *var* group C genes. Furthermore, we detected a significant up-regulation of *var* group A transcripts in parasites conferring rosetting.

These differences in expression are probably due to different transcriptional regulations leading to distinct switching rates. This might be supported by the observation of different, but within one group highly conserved 5' and 3' untranslated regions. [10,28] A mathematical model on *var* switching suggested that *var* genes with fast switching rates might be predominantly expressed in acute disease which is followed by parasites expressing *var* genes with slower switching rates. [29,30] This was indeed observed in volunteers which were artificially infected with 3D7. [31] If we adapt this hypothesis to our findings, increased switch on rates would be ascribed to *var* group B genes. However, to date, group specific switching rates have not been demonstrated, and one report, on the contrary, showed that different switching rates of *var* genes were independent of their 5'UTR sequence. [32] Different switching rates might not apply to all members of a group, but it was impossible in our study to analyse individual *var* genes. Therefore, the observed differences might also be due to few specific *var* genes which are switched on more often in cases or in controls. In 2 volunteers artificially infected with 3D7 Peters et al. [31] detected the same *var* gene, belonging to *var* group B, expressed at the start of infection.

The prominent presence of *var* group B transcripts in clinical malaria cannot be explained by increased switching rates alone. Successful immune evasion leading to parasite growth and pathogenicity occurs at the same time. It has been shown that clinical malaria is the consequence of infecting parasites which express variant surface antigens (VSA) that are not recognized by the existing repertoire of antibodies. [33] In 3D7, *var* group B represents the largest group with 22 different *var* genes [10] providing a large diversity. As sequestration is another important function of PfEMP1 in clinical malaria, our data would suggest that group B *var* genes code for good binding PfEMP1s, in contrast to *var* group C genes. Initially this seems unlikely, because the overall length and domain structure of *var*

group B and C is very similar in 3D7. And phylogenetic analyses on *var* coding regions shows only limited separate clustering of *var* group B and C genes [10,28](own observation). But minimal differences in the amino acid sequence could result in different binding phenotypes. A substitution of 3 amino acids in the CIDR domain of a *var* gene was shown to largely reduce binding to CD36. [12] Binding studies with 3D7 *var* genes showed that in contrast to *var* group A, most *var* group B and C encoded peptides bind to CD36. [11]

The role of CD36 binding IE in pathogenicity is controversial. Binding to CD36 was suggested to contribute to disease by sequestration and by conferring platelet mediated clumping which correlates with severe disease. [34,35] Conversely, CD36 binding isolates were found to be associated with non-severe disease, [13] and it was suggested to be protective by sequestering IE in non-essential muscle tissues (reviewed in [36]) or by promoting non-inflammatory phagocytosis of IE. [37]

We have found parasites with up-regulated expression of *var* group B genes in mild but also in severe malaria. The reason that we found no significant differences could be due to different subpopulations of *var* genes within *var* group B. It could also be due to technical limitations because *var* group B primers used for quantitative real-time PCR hybridised to upsB type 5'UTR sequences, which also included *var* genes of the intermediate *var* group "B/A" in 3D7. These are hybrid forms between *var* group A and B, and encode larger PfEMP1 similar to those encoded by *var* group A genes suggesting different binding abilities of these hybrids similar to *var* group A genes. [10] To sum up, whether PfEMP1 molecules encoded by *var* group B and C genes show different binding features relevant for the outcome of this study remains speculative and our data allow no further conclusions.

In asymptomatic children, we detected a significantly higher proportion of parasites transcribing *var* group C genes. It is a surprising finding that the comparatively small *var* group C with only 14 *var* genes in 3D7 should be able to sustain chronic infection. Asymptomatic malaria is characterized by low parasite loads reflected by a broad antibody repertoire against many VSAs of the infecting parasites. Recker et al. [38] recently suggested in a mathematical model that VSA epitopes eliciting short-lived and partially cross-reactive immune responses could sustain sequential *var* gene expression needed for chronic infection. We found previously in a longitudinal study in asymptomatic children in PNG, [26] a highly dynamic picture of *var* transcription with mostly new *var* transcripts at 2 weeks interval. Such dynamic and transient mixture of parasites in asymptomatic children might be generated by the continuous elimination of parasites presenting previously

expressed PfEMP1s. The development of an effective large antibody repertoire against PfEMP1 could be accelerated by cross-reactivity between different variants. According to our hypothesis these would comprise mainly group B *var* genes and various recombination and conversion processes in *var* group B genes might have had a homogenizing effect. This is supported by phylogenetic analyses in which several *var* group B genes cluster together (own observation). The up-regulation of *var* group C transcribing parasites in asymptomatic children could suggest an increased switch-off rate of this group, which would result in a rapid turnover with low parasite density and impeding successful antibody maturation by inducing anergy. The positive correlation of parasitemia and *var* group C transcripts in asymptomatic children would reflect the fitness of parasites transcribing *var* group C PfEMP1s in semi immune children.

We also found an increase of *var* group A transcripts in parasites conferring rosetting. This is in concordance with observations with culture strains, in which the rosetting phenotype was mostly attributed to *var* group A genes (own observation). It has been shown that rosetting is correlated with severe disease in African children [18,19] but not in PNG. [39] These controversial results could be explained by the frequent presence of red blood cell traits in PNG such as α-thalassaemia [40] or CR1 deficiency which was found in up to 79% of the population of Madang, PNG. [41] It is tempting to speculate that the lack of a disease association with rosetting in PNG is due to the large proportion of CR1 deficient individuals. [41] Therefore, it is not surprising to find no up-regulations of *var* group A genes in severe malaria cases, [23] but only an association with rosetting.

The lack of a significant difference of *var* group transcripts between mild and severe cases was somehow unexpected. Yet, malaria is a multi-system disorder (for a review, see Mackintosh [42]) and not only caused by PfEMP1, but also by immune-mediated inflammatory processes (for a review, see Artavanis-Tsakonas [43]). The disease progression can be seen as continuum and the time between onset of symptoms and final presentation at the hospital is important. Children with severe disease might merely reflect children with a longer progression of the same disease status in contrast to asymptomatic children in whom parasites currently are at a steady state with the immune system. However, the small study size has also to be taken into account and more data are needed. Because of the unknown population of *var* genes and to exclude primer bias, we also performed quantitative analysis of genomic DNA of all samples. Normalisation of *var* transcript

numbers with genomic *var* copy numbers did not change the results and we are certain that primer bias does not explain the outcome.

Finally, concern has been raised whether the same distribution of parasites is found in peripheral blood as in vital organs, where virulent parasites sequester. Different adhesive phenotypes of IE were shown in peripheral IE of pregnant women compared to IE isolated of their placenta. [44] But conversely, parasites of peripheral blood of children with severe malaria showed a different agglutination pattern than parasites of children with mild malaria, indicating a restricted parasite population also in the periphery of children with severe malaria. [45]

In conclusion, here we show for the first time in a malaria case-control study conducted in an endemic area that the different *var* subgroups influence clinical outcome. In particular, we show a significant increase of *var* group B transcripts in clinical malaria. In contrast, *var* group C transcripts were significantly more often found in asymptomatic individuals. We also show that *var* group A transcripts are associated with rosetting of parasites. These findings confirm together with previous studies the importance of PfEMP1 in disease manifestation, and justifies further studies to understand its molecular development.

Acknowledgements

We wish to express our thanks to Judy Longo and Wilbert Jake for the recruitment and assessment of cases and the staff of the Papua New Guinea Institute of Medical Research, Madang, for their field assistance. We appreciated the cooperation of the parents and children who participated in the study. We are thankful to Marco Gut (Applied Biosystems) for technical assistance in real-time PCR and Anja Jensen for providing us with the selected RNA. Finally, we are very grateful to Joe Smith for scientific assistance and Tom Smith for scrutinizing the statistical analysis.

References

1. Muller I, Bockarie M, Alpers M, Smith T. The epidemiology of malaria in Papua New Guinea. Trends Parasitol. 2003;19:253-259.

2. Severe *falciparum* malaria. World Health Organization, Communicable Diseases Cluster. Trans R Soc Trop Med Hyg. 2000;94 Suppl 1:S1-90.

3. Kyes S, Horrocks P, Newbold C. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol. 2001;55:673-707.

4. Su XZ, Heatwole VM, Wertheimer SP et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell. 1995;82:89-100.

5. Baruch DI, Pasloske BL, Singh HB et al. Cloning the P. *falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell. 1995;82:77-87.

6. Smith JD, Chitnis CE, Craig AG et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell. 1995;82:101-110.

7. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. Mol Biochem Parasitol. 2000;110:293-310.

8. Gardner MJ, Hall N, Fung E et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002;419:498-511.

9. Voss TS, Thompson JK, Waterkeyn J et al. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5' flanking sequences. Mol Biochem Parasitol. 2000;107:103-115.

10. Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. Malar J. 2003;2:27.

11. Robinson BA, Welch TL, Smith JD. Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol. 2003;47:1265-1278.

12. Gamain B, Smith JD, Miller LH, Baruch DI. Modifications in the CD36 binding domain of the *Plasmodium falciparum* variant antigen are responsible for the inability of chondroitin sulfate A adherent parasites to bind CD36. Blood. 2001;97:3268-3274.

13. Newbold C, Warn P, Black G et al. Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. Am J Trop Med Hyg. 1997;57:389-398.

14. Rogerson SJ, Tembenu R, Dobano C et al. Cytoadherence characteristics of *Plasmodium falciparum*-infected erythrocytes from Malawian children with severe and uncomplicated malaria. Am J Trop Med Hyg. 1999;61:467-472.

15. Buffet PA, Gamain B, Scheidig C et al. *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. Proc Natl Acad Sci U S A. 1999;96:12743-12748.

16. Reeder JC, Cowman AF, Davern KM et al. The adhesion of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate A is mediated by P. *falciparum* erythrocyte membrane protein 1. Proc Natl Acad Sci U S A. 1999;96:5198-5202.

17. Turner GD, Morrison H, Jones M et al. An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. Am J Pathol. 1994;145:1057-1069.

18. Rowe A, Obeiro J, Newbold CI, Marsh K. *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. Infect Immun. 1995;63:2323-2326.

19. Carlson J, Helmby H, Hill AV et al. Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. Lancet. 1990;336:1457-1460.

20. Bull PC, Lowe BS, Kortok M, Marsh K. Antibody recognition of *Plasmodium falciparum* erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. Infect Immun. 1999;67:733-739.

21. Bian Z, Wang G, Tian X, Fan J. Expression of *Plasmodium falciparum*-infected erythrocyte membrane protein from cerebral malaria patients. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 1999;17:359-362.

22. Kirchgatter K, Portillo HA. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med. 2002;8:16-23.

23. Jensen AT, Magistrado P, Sharp S et al. *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A *var* genes. J Exp Med. 2004;199:1179-1190.

24. Allen SJ, O'Donnell A, Alexander ND, Clegg JB. Severe malaria in children in Papua New Guinea. QJM. 1996;89:779-788.

25. Felger I, Tavul L, Beck HP. *Plasmodium falciparum*: a rapid technique for genotyping the merozoite surface protein. Exp Parasitol. 1993;77:372-375.

26. Kaestli M, Cortes A, Lagog M, Ott M, Beck HP. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. J Infect Dis. 2004;189:1942-1951.

27. Henning L, Felger I, Beck HP. Rapid DNA extraction for molecular epidemiological studies of malaria. Acta Trop. 1999;72:149-155.

28. Kraemer SM, Smith JD. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum var* gene family. Mol Microbiol. 2003;50:1527-1538.

29. Gatton ML, Cheng Q. Investigating antigenic variation and other parasite-host interactions in *Plasmodium falciparum* infections in naive hosts. Parasitology. 2004;128:367-376.

30. Gatton ML, Peters JM, Fowler EV, Cheng Q. Switching rates of *Plasmodium falciparum var* genes: faster than we thought? Trends Parasitol. 2003;19:202-208.

31. Peters J, Fowler E, Gatton M et al. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci U S A. 2002;99:10689-10694.

32. Horrocks P, Pinches R, Christodoulou Z, Kyes SA, Newbold CI. Variable *var* transition rates underlie antigenic variation in malaria. Proc Natl Acad Sci U S A. 2004;101:11129-11134.

33. Bull PC, Lowe BS, Kortok M et al. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med. 1998;4:358-360.

34. Pain A, Ferguson DJ, Kai O et al. Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci U S A. 2001;98:1805-1810.

35. Chotivanich K, Sritabal J, Udomsangpetch R et al. Platelet-induced autoagglutination of *Plasmodium falciparum*-infected red blood cells and disease severity in Thailand. J Infect Dis. 2004;189:1052-1055.

36. Serghides L, Smith TG, Patel SN, Kain KC. CD36 and malaria: friends or foes? Trends Parasitol. 2003;19:461-469.

37. Patel SN, Serghides L, Smith TG et al. CD36 mediates the phagocytosis of *Plasmodium falciparum*-infected erythrocytes by rodent macrophages. J Infect Dis. 2004;189:204-213.

38. Recker M, Nee S, Bull PC et al. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature. 2004;429:555-558.

39. al Yaman F, Genton B, Mokela D et al. Human cerebral malaria: lack of significant association between erythrocyte rosetting and disease severity. Trans R Soc Trop Med Hyg. 1995;89:55-58.

40. Oppenheimer SJ, Higgs DR, Weatherall DJ, Barker J, Spark RA. Alpha thalassaemia in Papua New Guinea. Lancet. 1984;1:424-426.

41. Cockburn IA, Mackinnon MJ, O'Donnell A et al. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. Proc Natl Acad Sci U S A. 2004;101:272-277.

42. Mackintosh CL, Beeson JG, Marsh K. Clinical features and pathogenesis of severe malaria. Trends Parasitol. 2004;20:597-603.

43. Artavanis-Tsakonas K, Tongren JE, Riley EM. The war between the malaria parasite and the immune system: immunity, immunoregulation and immunopathology. Clin Exp Immunol. 2003;133:145-152.

44. Fried M, Duffy PE. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. Science. 1996;272:1502-1504.

45. Bull PC, Kortok M, Kai O et al. *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J Infect Dis. 2000;182:252-259.

Table 1. Oligonucleotide primers for amplification of *var* gene regions

var gene region	length of amplified product	name	primer sequence	Ref.
DBL1αrev *		DBL_FADall_rev	5'-CC(A/T)AT(A/G)(G/T)C(A/G/T)GCAAAACT(C/G/T)C(G/T)(A/T)GC-3'	
var group A, 5'UTR	150 bp	upsA1_for **	5'-AACTTACCATAAATTATCATCAAA-3'	
		upsA3_probe	5'-6FAM-AAACCTTTGGTATAGAAAAAAATATT-MGB-3'	
		upsAj_rev	5'-TCACCTACAACAAAT(A/G)TAATAAA-3'	(Smith J, pers.comm.)
var group B, 5'UTR	370 bp	17deg_for **	5'-CTCAT(A/T)TATAATTTTA(C/G)AAAATA(A/T)A(A/T)AAAAC-3'	
		upsB1_probe	5'-6FAM-TCTAACAAAAAAAAAAAACAACAATTAC-MGB-3'	
		RT-17.2_rev	5'-TTA(A/T)GGGAGTAT(A/T)GT(A/G/T)ATATGGTAGAAT-3'	
var group C, 5'UTR	230 bp	RT-5B1.1_for **	5'-AATATTCATATTCCCACATT(A/G)TCATATAT-3'	
		upsC_probe	5'-6FAM-ACATATAATACGACTAAGAAAC-MGB-3'	
		5B1.4_rev	5'-ATTATGTGGTAATATCATGTAATGG-3'	

*reverse primer which was used in the primary PCR **forward primers were identical in primary and real-time PCR

Table 2: Clinical assessment of cases and controls (% of children)

Overview of children and clinical features	Clinical outcome		
	asymptomatic (25 children)	mild (26 children)	severe (14 children)
Sex (female/male)	9/16	10/16	7/7
Age (months) (mean [range])	41.3 [14-65]	42.2 [12-84]	35.8 [8-60]
Parasites/µl (geometric mean [range])	3'715 [40-108'000]	27'241 [160-310'200]	32'928 [240-480'000]
Temp (˚C) (geometric mean [range])	36.6 [35.4-37.5]	38.5 [35.6-40]	38.7 [37.1-39.9]
Medical history of fever in the last 5 days	5 (20%)	22 (85%)	14 (100%)
Medical history of headache in the last 5 days	0	9 (35%)	6 (43%)
Medical history of diarrhea in the last 5 days	0	3 (12%)	1 (7%)
> 1 convulsion/24h	0	0	5 (36%)
Prostration	0	0	7 (50%)
Impaired consciousness ; coma ; neurological alterations	0	0	1 (7%) ; 4 (29%) ; 4 (29%)
Anormal respiration (shallow, fast, irregular or deep)	0	3 (12%)	8 (57%)
Severe anemia (hemoglobin <5 g/dl)	0	0	3 (21%)
Hypoglycemia (glucose <2.2 mmol/l)	0	0	4 (29%)
Hyperlactatemia (lactate >5 mmol/l)	0	0	0

Table 3:

Odds ratios for disease with increased proportions of *var* group specific transcripts.

Comparison between:		Odds ratios for disease		
		var group A proportions	*var* group B proportions	*var* group C proportions
Asymptomatic vs clinical (mild and severe)	unadjusted	**0.94** [0.11-8.0] 0.96	**18.95** [3.21-111.87] <0.001	**0.01** [0.00-0.22] <0.001
	adjusted	**0.76** [0.07-8.00] 0.82	**31.63** [3.68-271.71] <0.001	**0.02** [0.00-0.24] <0.001
Asymptomatic vs mild	unadjusted	**1.56** [0.18-13.35] 0.69	**10.53** [1.74-63.64] 0.006	**0.02** [0.00-0.37] <0.001
	adjusted	**1.11** [0.11-11.75] 0.93	**21.71** [2.45-192.67] 0.002	**0.02** [0.00-0.33] <0.001
Asymptomatic vs severe	unadjusted	**0.17** [0.00-8.90] 0.33	**111.47** [2.43-5108.28] <0.001	**0.01** [0.00-1.20] 0.004
	adjusted	**0.28** [0.00-43.73] 0.59	**302.16** [1.44-63248] 0.002	**0.01** [0.00-1.13] 0.006
Mild vs severe	unadjusted	**0.04** [0.00-6.67] 0.13	**12.61** [0.26-612.36] 0.15	**0.64** [0.00-159.1] 0.87
	adjusted	**0.05** [0.00-7.18] 0.15	**9.89** [0.20-494] 0.20	**1.31** [0.01-380] 0.93

OR [95%CI] P value; Odds ratios are either unadjusted or adjusted in multivariate logistic regression for parasitemia; Significant results are shaded in grey

Figure legend

Figure 1:

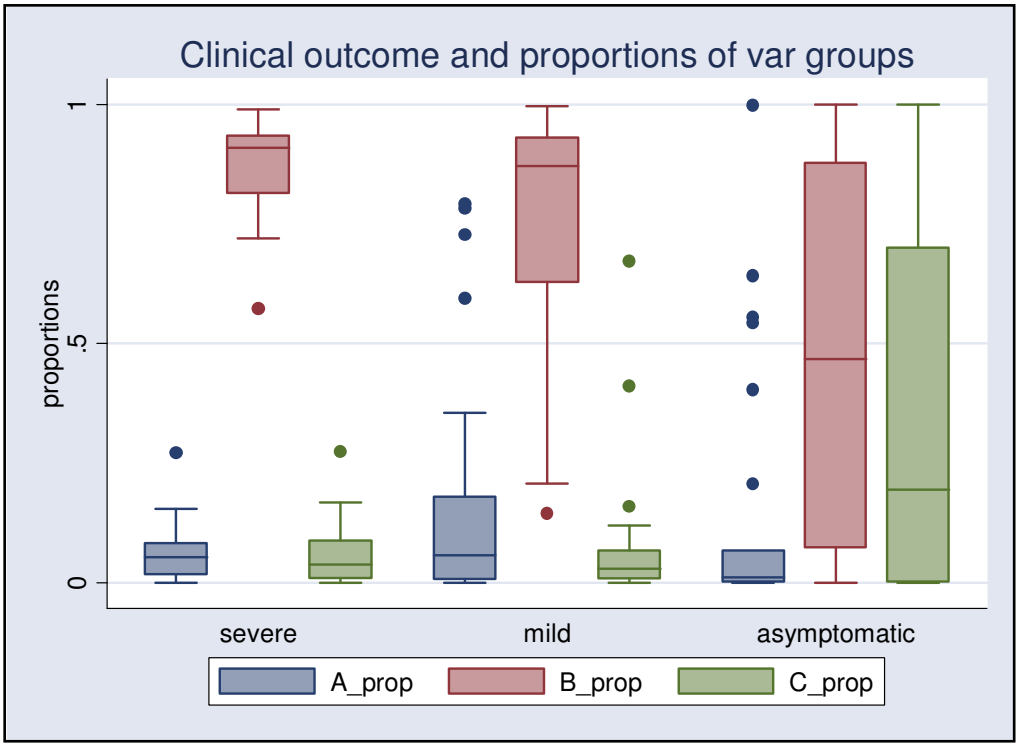
1A) Box plots of clinical outcome (severe malaria, mild malaria, asymptomatic malaria) and proportions of *var* group specific transcripts over the sum of all amplified *var* transcripts.

1B) Box plots of parasites mediating rosetting and proportions of *var* group A, B and C transcripts.

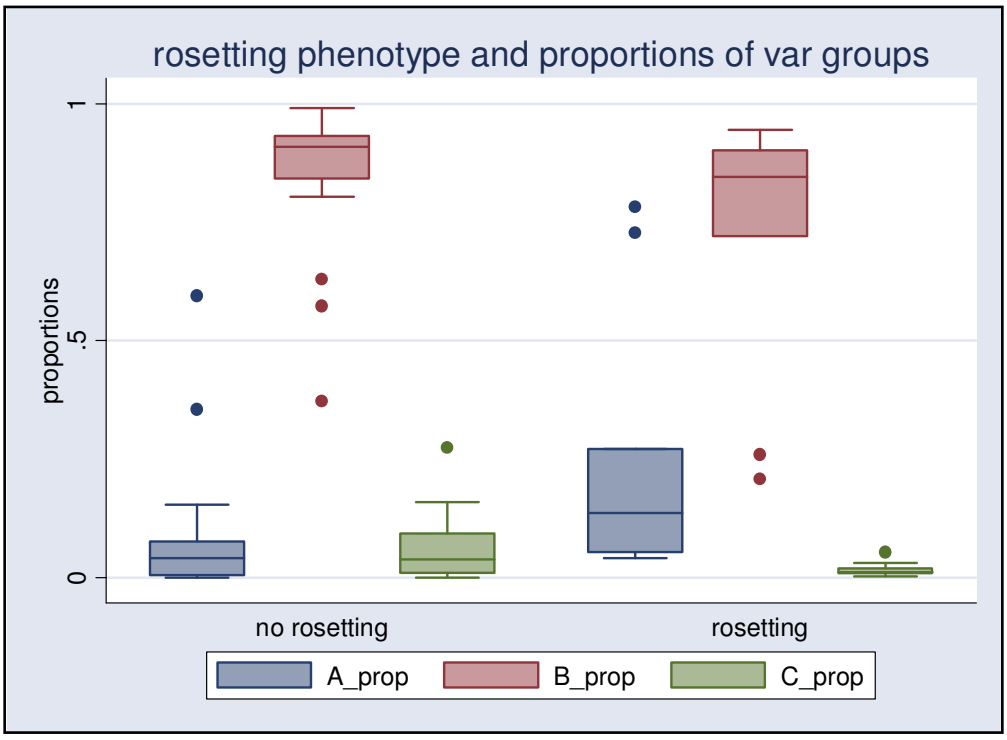
Boxes of box plots contain the middle 50% of data ranging from upper to lower quartile while the line in the box indicates the median. Vertical lines extend to a maximum of 1.5 times the inter-quartile range, points are outliers.

Figure 1

1A)



1B)



4. Discussion

PfEMP1 is an important virulence factor of *P. falciparum* not only in acute infections but also by contributing to chronic disease. Therefore, it is not surprising that this protein family has been discussed as potential vaccine target. [43,72,75] However, few studies have analyzed *var* expression dynamics *in vivo* and pathological implications of different expressed PfEMP1. [12,71,74] Here, we describe 2 studies on *var* gene expression *in vivo*. In a longitudinal study over a 4-month period, we analyzed the dynamics of *var* gene expression over time. We studied the longitudinal distribution of *var* transcripts in semi-immune children by RT-PCR, cloning and sequencing. In a case-control study, we evaluated potential differences in the virulence of PfEMP1 encoded by *var* group A, B or C. By using quantitative real-time PCR, we compared the distribution of *var* gene transcripts of *var* groups A, B and C among children with severe, mild and asymptomatic malaria.

Both studies were carried out in Madang Province, a highly endemic area at the north coast of Papua New Guinea (PNG). Analysis of Madang hospital paediatric records of 1994-1995 showed that 19.2% of all admissions and 15.3% of all deaths in children were assigned to malaria. [76] Furthermore, the most common severe manifestations of malaria in children were shown to be severe anemia (22%) and coma (16%) and hyperlactatemia was the major predictor of death. [76]

4.1. Expression of *var* group A, B and C genes in clinical malaria

We found a significant up-regulation of *var* group B genes in children with clinical malaria (severe and mild) while *var* group C transcripts were hardly detectable in these children. No clear trend was evident between mild and severe malaria. Compared to children with mild malaria, only a further increase of *var* group B proportions was observed in children with severe malaria. In contrast to previous indications (see Introduction section 1.9.), we found no up-regulation of *var* group A in children with severe malaria but only a slight increase in children with mild malaria when compared to children with asymptomatic malaria.

The lack of different proportions of *var* group specific transcripts in mild and in severe malaria could be attributed to the limited dataset of only 14 children with severe malaria.

However, the fact that we only collected 14 severe malaria cases in 3 months also reflects the smaller contribution of severe malaria to morbidity in Madang. Severe malaria

morbidity and mortality is less frequent in PNG than in many endemic settings of Africa. [77] One reason for this reduced morbidity with *P. falciparum* in PNG might be the omnipresence of the less virulent *P. vivax* which is suggested to induce some degree of cross-protection. [78]

Apart of a fairly good access to antimalarial drugs in Madang, these reduced severe manifestations could also be attributed to the genetic background of Papua New Guineans. Various red blood cell traits such as Southeast Asian Ovalocytosis (SAO) [79], α-thalassemia [80], CR1 deficiency [81] and Gerbich-negative phenotype [82] are prevalent and have been shown to a various degree to be protective against severe malaria. A high prevalence of 80% of CR1 deficient individuals in Madang [81] could have accounted for the observed lack of up-regulation of *var* group A genes in severe disease. CR1 is known to be the main ligand on erythrocytes for rosetting which was shown to be associated with severe disease in Africa but not in PNG. [22,83] We found a significant association between parasites conferring rosetting and an up-regulation of *var* group A genes in these parasites. This is in concordance with *in vitro* observations (personal communication Alex Rowe). It is tempting to speculate that rosetting is not associated with severe disease in PNG because the main ligand CR1 is missing. [81] Not fully developed rosettes which do not withstand physiological shear stress *in vivo* could explain the non-existent correlation between up-regulation of *var* group A genes and severe disease in PNG. However, this implies that the rosetting phenotype of PfEMP1 molecules encoded by *var* group A genes contributes significantly to severe disease. But other severe manifestations including binding of IE to ICAM1 have been implicated with large *var* genes and *var* genes provided with a DBL2β-C2 domain in cerebral malaria. [38,70] In 3D7, DBL2β-C2 containing *var* genes belong to *var* group A but also in some cases to the intermediate groups B/A and B/C. [45] Since we included *var* group B/A in group B, we cannot exclude that the observed *var* group B up-regulation in severe malaria is due to *var* B/A gene up-regulation. Further sequence analysis will shed light onto this question.

It has always been discussed whether parasites circulating in the periphery represent the disease associated phenotypes. There is controversial data on phenotypes of circulating parasites and adhering parasites from human placental studies. Two studies reported identical binding phenotypes in the placenta and in the periphery, [84,85] whereas other studies showed antigenically distinct parasites. [21,86] By performing agglutination assays, Bull et al. [69] found a distinct parasite subset in peripheral blood of children with severe malaria compared to children with mild malaria. This indicates that virulent parasites also

circulate in the periphery. Since various host tissues express different receptors, the presence of different parasite subsets within one host can be expected.

But one can also speculate that severe disease is maintained by a homogenous parasite population expressing multi-adhesive variant surface antigens (VSA) or a parasite population expressing a VSA subset which mediates cytoadherence independent of the endothelium such as rosetting or platelet-mediated clumping. Although not in PNG, rosetting and clumping have both been shown to be correlated with severe disease [10,23] while multi-adhesive IE have been shown to occur *in vitro.* [87]

The observation of similar *var* transcript proportions in mild and in severe malaria raises a fundamental question about the definition of severe and mild malaria. Is there a qualitative difference between mild and severe disease? The progression of disease might be merely a factor of time and a child arriving at the hospital with mild malaria might be severely sick a few hours later without treatment. According to our data, there is no significant change in *var* gene transcript composition. However, we observed a large variation of *var* group specific transcript proportions between individuals of clinical cases and individuals with asymptomatic infections. We propose that a delay in the clinical picture in relation to the transcribed *var* subset contributed to this large deviance.

4.2. Expression of *var* group A, B and C genes in asymptomatic children

In our case-control study, *var* group C genes were found to be highly up-regulated in asymptomatic children. Conversely, on average, only half of *var* group B transcripts were detected in children with asymptomatic malaria compared to children with clinical malaria. Also *var* group A proportions were reduced overall in asymptomatic children, but 6 of these 25 children had high *var* group A proportions.

In the longitudinal study on older, semi-immune children, we did not perform quantitative analyses, but sequence analysis revealed a larger diversity of *var* group B transcripts (5'UTR-DBL1α sequences) than of *var* group C transcripts. Of the latter, never more than 3 different transcripts were found per child compared to 11 different *var* group B transcripts. On average, 3.7 different *var* group B transcripts were detected per child compared to 1.5 different *var* group C sequences. This will be discussed and incorporated into the hypotheses below. In this longitudinal study, DBL1α subtype sequences characteristic of *var* group A genes were also isolated, indicating the presence of *var* group A transcripts in semi-immune children.

We found a highly dynamic picture of *var* expression in the longitudinal study with mostly new *var* transcripts at each 2 week interval but some *var* transcripts recurred for up to 10 weeks. RT-PCR, cloning and sequencing revealed up to 15, but on average 5 different DBL1α sequences per child. When multiple infections were taken into account we calculated a mean of 1.7 DBL1α sequences per infection. This might be a good estimation of the actual number and reflects the tight control of *var* gene transcription within a single parasite population. It is not known how parasites are able to synchronize their *var* gene expression avoiding the simultaneous expression of many *var* transcripts despite the reported high switching rate. It seems unlikely that elimination of concurrent variants by immune responses could account for the observed low diversity because this would require rapid effector mechanisms and would lead to an early exhaustion of the *var* repertoire. Other explanations such as an inter-parasite communication by chemotactic signalling have not been shown yet and no signalling pathways from the erythrocyte membrane to the parasite have been found except of indications of a cAMP dependent signalling pathway being involved in initiation to gametocytogenesis. [88] However, low diversity could be obtained by an ordered switching mechanism suggesting some sort of epigenetic programming to prevent the random switching to different *var* genes in different parasites. However, this is speculative and not much is known about *var* switching control.

Although we observed only a small number of expressed *var* variants per infecting strain, we detected on average 5 different *var* transcripts in semi-immune children due to multiple *P. falciparum* infections. Different VSA subtypes presented at the same time to the immune system might be advantageous for the parasite because it might lead to immune distraction and the distribution of different VSA subsets among different *P. falciparum* strains reduces the risk of a fast exhaustion of the *var* gene repertoire. In endemic areas, larger anti-VSA antibody repertoires have been shown to occur in asymptomatic children compared to matched children with no detectable *P. falciparum* infections [89] indicating an on-going build-up of anti-VSA antibodies of which some might only be short lasting.

In a mathematical model, Recker et al. [90] showed that the combination of 2 types of immune responses against PfEMP1 variants can maintain chronic infection. In this model, short-lived immune responses (for instance due to linear, low complexity epitopes) are built up against minor, partially cross-reactive epitopes of PfEMP1 variants. But every PfEMP1 variant is provided with a major unique epitope which elicits long-lived responses. This combination results in a cascade of sequentially dominant PfEMP1 variants of which the active variant is always the most immunologically distinct variant

from the previous. Parasites expressing similar minor epitopes are suppressed by short-lived antibody responses while the development of long-lived, variant specific immune responses limits an aggressive amplification of parasites expressing immunologically distinct forms. [91]

This model is independent of different *var* switching rates but can explain the low number of simultaneously expressed *var* variants in an infecting strain. In this case, immune selection would be solely responsible for the observed low diversity.

4.3. Two hypotheses explaining our results

I will discuss our findings in the context of two hypotheses.

4.3.1. Hypothesis I: Different switching rates of *var* groups

Initially, it is difficult to imagine that the observed shifts of *var* group specific transcripts in clinical and asymptomatic malaria can only be attributed to distinct virulences of the PfEMP1 molecules in different clinical presentations. Phylogenetic analyses and basic binding studies did not show any major differences in these two *var* subfamilies [45,92](and own observation). We propose that a different transcriptional regulation of *var* groups such as different switching rates accounts for our observations. The different, but within one group highly conserved untranslated regions of *var* genes might support this hypothesis. And mathematical models indicated that a combination of innate, clonal and anti-PfEMP1 immunity could maintain chronic infection when using different *var* switching rates influencing anti-PfEMP1 immunity. [93,94] Moreover, Horrocks et al. [95] showed in isogenic parasite clones *var* genes with identical inherent "on" and "off" rates. However, in their experiments these different "on" and "off" rates were not associated with specific *var* groups. Yet, this study collected data with sampling only every 10 generations over 40 generations.

We base our hypothesis on following assumptions in relation to distinct *var* group specific switching rates:

- ***var* group C has high switch off rates**
- *var* group A and B have low switch off rates. Accordingly, *in vivo* parasites expressing these genes are mainly eliminated by the immune system

- *var* group A and B also have increased switch on rates compared to *var* group C, therefore, *var* group A and B genes have a higher likelihood to be switched on first in an infection.

The following explanations are summarized in Figure 2. Row (1) shows the typical course of parasitemia during infection and row (2) our observed proportions of *var* group specific transcripts. Rows (3) and (4) show possible scenarios which could explain our findings. The left part depicts an infection in a malaria naïve host while the right part refers to a semi-immune host with a modest anti-VSA antibody repertoir.

According to our assumptions, *var* group A and B encoded PfEMP1s are prominently expressed early in infection (row (3)) while *var* group C variants do not play a major role in acute infections as they tend to be fast switched off again (row (3): small arcs). Therefore, hardly any antibodies are mounted against these group C PfEMP1s (row (3) as depicted in the small antibody drawing to the right.

Larger PfEMP1 molecules of *var* group A and B/A are provided with more adhesive domains and were shown to have distinct binding characteristics mediating rosetting or binding to ICAM1 in the brain (see above). Therefore, these variants might bind to receptors in vital organs or might confer rosetting contributing to the progression to more severe manifestations of disease. Due to the conserved structure of *var* group A variants, the small size of this group, and their extended presentation, I propose that these variants elicit an effective variant-specific, long-lasting immune response (as depicted in the right panel of row (3) and in row (4) with large arcs). This could contribute to the protection from severe malaria already observed after only 1 or 2 clinical episodes. [96] According to the model of Recker et al. [90] (see above), I also assume that all PfEMP1 variants possess minor epitopes eliciting cross-reactive, short-lasting responses shared between different PfEMP1 variants (row (4): small arcs in all 3 *var* groups) and variant-specific epitopes mounting long-lasting responses. The existence of cross-reactive antibodies against PfEMP1 was shown after immunizing mice with 3 CIDR1 domains. [97] Since *var* group B genes have been shown to undergo numerous small recombination and conversion events [47,98] resulting in a potential homogenizing effect, *var* group B variants might share a considerable amount of minor epitopes which could increase the production of primarily short-lasting antibodies against these PfEMP1 (row (4): small arcs). The development of long- and short-lasting immune responses against *var* group B variants leads to a continuous elimination and turn over of these parasites.

Because of increased switch on but low switch off rates of *var* group A and B genes and distinct structural features of a subset of these genes, group A and B encoded PfEMP1 molecules are predominantly found in parasites causing acute infections. Since in malaria naïve hosts, an effective antibody response against PfEMP1 variants does not yet exist and parasites do sequester, unrestrained parasite amplification occurs (row (1)). Eventually, initial acute infections are either cleared by drug treatment, successful anti-parasitic immune response, or by death.

In semi-immune hosts, a modest anti-VSA antibody repertoire has been built up. This antibody repertoire might be primarily targeted against *var* group A and B variants because of their preferred exposure in previous infections (row (4), second column). This would result in the constant elimination of parasites expressing these variants, and leads to a fast and dynamic turn over of parasites expressing *var* group A and B variants (row (3), right panel). Such dynamic turn over was also evident in the longitudinal study. And due to the constant elimination of these variants, the presence of *var* group C encoding parasites increases.

I assume that the combination of previously exposed and cross-reactive *var* group B variants as well as the short presentation of *var* group C variants results in a strongly reduced variant specific immune response against PfEMP1 of *var* group C. There are also indications that *var* group C genes are more diverse than *var* group B genes, and higher fixation indexes up to 0.53 (P=0.003) have been found between *var* group C genes of different culture strains (Dieter Suter, personal communication). This could further support a sequential dominance of *var* group C derived variants, each variant being the most immunologically distinct of the previous expressed one.

We did observe asymptomatic children with high parasite loads above 50'000 parasites/µl and with high proportions of *var* group C transcripts above 90%. Since these children seem not to mount a pro-inflammatory, fever inducing immune response despite of this high parasite load, we suggest that some sort of immune unresponsiveness or anergy exists against parasites expressing these variants. Distinct binding characteristics of these PfEMP1 variants could explain why there are no obstructions in local blood flow in vital organs and accordingly, no apparent clinical signs in these children.

Figure 2

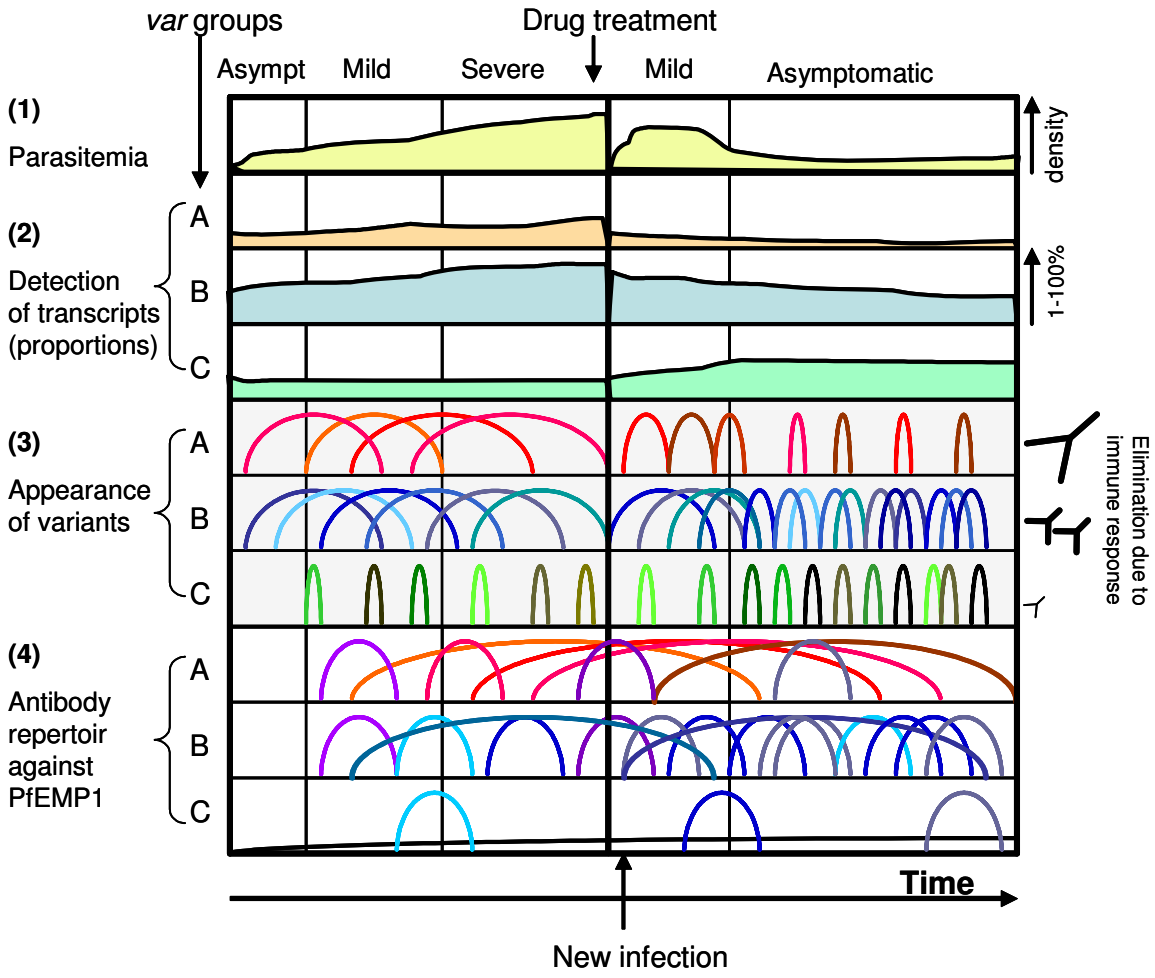


Figure 2: Model of *var* group specific transcription of a *P. falciparum* strain in a malaria naive host and in subsequent infections of the same host after build-up of a modest anti-VSA antibody response. The primary severe infection is cleared by drug treatment or by a successful anti-parasitic immune response.
Small arcs in the "Antibody repertoire" indicate short-lived antibodies against cross-reactive epitopes, while large arcs indicate long-lived variant specific immune responses. For explanations see text.

Distinct binding characteristics might lead to an accumulation of parasites in non-essential muscle tissue due to binding to CD36, constant removal by non-inflammatory phagocytosis activated by CD36 binding to macrophages (see below, 4.5.), or low binding abilities leading to a constant removal by the spleen.

This model could explain our findings of increased proportions of *var* group A and B transcripts in clinical malaria while *var* group C derived variants are most prominent in asymptomatic malaria. It also explains the large variation of our data, particularly of proportions of *var* group B transcripts in children with asymptomatic malaria. These children might well be at the beginning of an infection (left part of left panel). The 6 observed asymptomatic children with large *var* group A proportions might be explained by this delay of disease (see above).

4.3.2. Hypothesis II: Ordered switching pattern

Horrocks et al. [95] found some evidence for ordered switching events suggesting that the ability to switch to a certain *var* gene depends on the switching history of this parasite. In this second hypothesis, I propose a model which is to a certain degree a programmed ordered process and assumes that switching primarily occurs within one *var* group while switching from *var* group B to *var* group C is a rare event, in particular at the beginning of an infection. On a molecular level, an ordered switching mechanism could be explained by epigenetic mechanisms. For instance, *var* group C genes might be packed in silent chromatin after passage through mosquito and liver stages. Accordingly, the machinery for *var* gene transcription is concentrated at the chromosome subtelomers transcribing *var* group B or A genes. It might also be possible that the transcription machinery specific for *var* group C expression is not yet available at the beginning of the asexual erythrocyte stage. There is some evidence provided by Peters et al. [74] for an imprinting mechanism of *var* transcription at the beginning of a new infection, who found the same *var* group B gene switched on in different volunteers artificially infected with 3D7. Subsequently expressed *var* genes were also mainly of *var* group B type.

Hypothesis II is illustrated below in Figure 3. Letters A, B and C indicate parasites expressing variants of *var* groups A, B or C while numbers after the letter indicate different variants within one *var* group. I assume that at the beginning of an infection, a high switching rate is advantageous to guarantee the establishment of a new infection in a host, above all in high transmission areas.

Figure 3

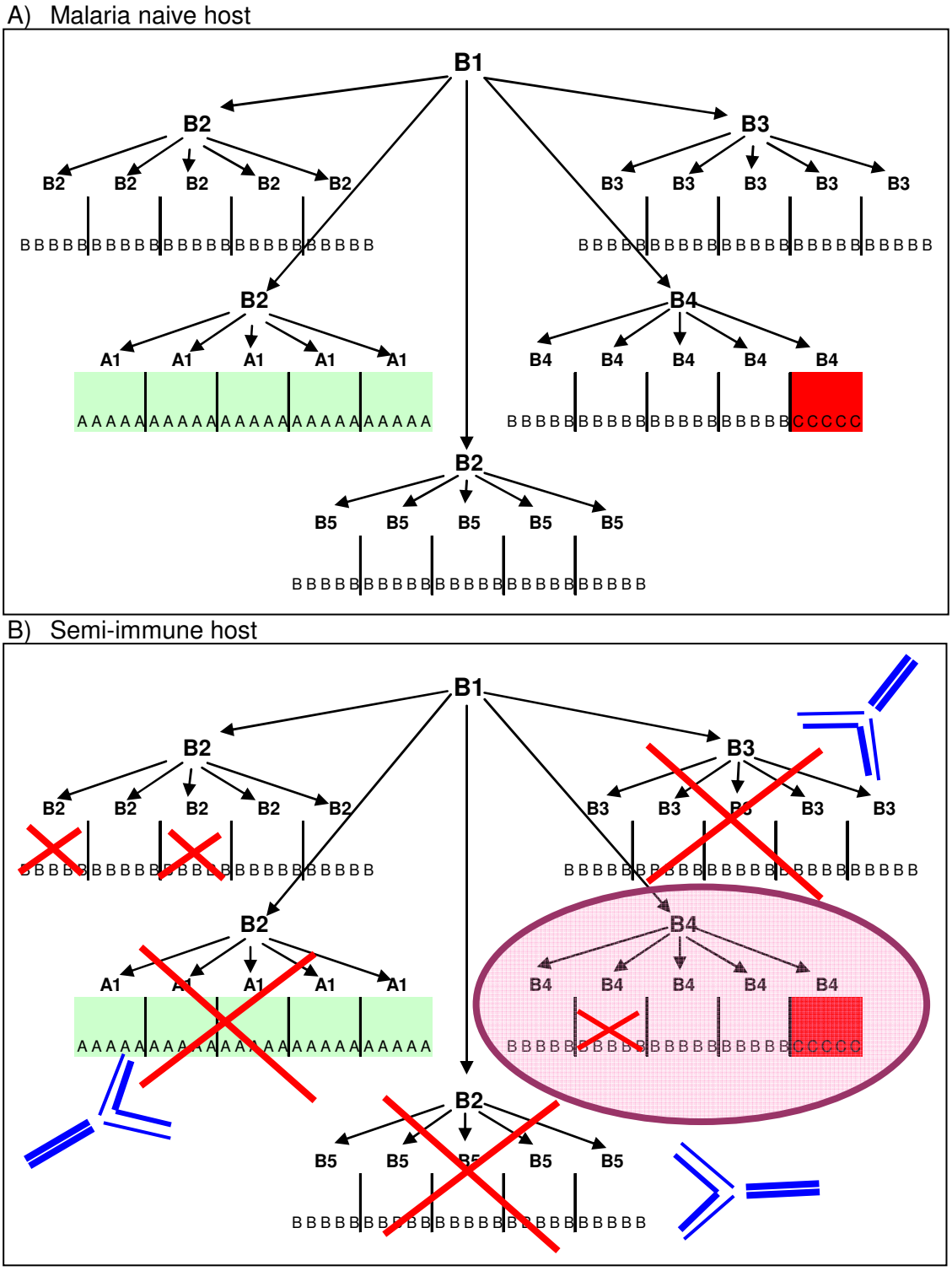


Figure 3. Switching of *var* group A, B or C variants during parasite amplification. At the beginning of an infection, a *var* group B gene is active, from which other *var* genes are switched on at high rates. Different *var* genes within one group are indicated with numbers. For simplicity, only 5 merozoites derived from one parasite are indicated, however, up to 24 merozoites are normally found.
A) amplification of parasites expressing *var* group A, B or C derived PfEMP1 in a malaria-naïve host
B) reduced amplification of parasites in a semi-immune host. A considerable part of variants is eliminated by existing immune response

A high switching rate leads to the expression of several variants (in this hypothesis primarily of *var* group B type) increasing the chance of at least one not being recognized by the host. This was indeed seen in the artificially infected volunteers in whom fast switching was observed after the expression of the initial *var* gene.
In hosts with a modest anti-PfEMP1 antibody repertoire, a considerable part of parasites expressing previously exposed *var* group B variants is immediately eliminated by immune responses. This would increase the proportion of *var* group C expressing parasites (see Figure 3, B)).

Quantitative real-time PCR analysis in 3D7 and FCR3S1.2 culture parasites, showed an up-regulation of *var* group C genes similar to the situation in asymptomatic hosts (own observation). This was not expected, as in culture no immune selection is present. Relating to hypothesis II, I assume that the initial imprinting of *var* group B expression after liver stage passage is lost in cultured parasites.

As mentioned above, sequence analysis of the longitudinal study revealed a smaller number of *var* group C variants in semi-immune children and in culture parasites than *var* group B variants. Although *var* group C genes seem to be up-regulated, only very few variants of *var* group C genes are transcribed simultaneously. High switch off rates for *var* group C genes or an ordered switching mechanism could account for that. But, this could also simply be ascribed to the lower number of *var* group C genes per haploid genome compared to *var* group B genes. In 3D7, our *var* group specific primers targeted 33 *var* group B genes but only 9 *var* group C genes.

4.4. Structural differences in PfEMP1 of *var* group B or C

Despite of very similar DNA sequences of *var* group A and B, protein structure differences could also contribute to our findings. *var* group B and C genes most likely have diverged early in *P. falciparum* history, and phylogenetic analyses have shown that DNA exchange is reduced between these groups. [45,99] This is in contrast to the ectopic recombination currently observed within groups. Small *var* group specific differences in amino acid sequence might have evolved which are not apparent by broad phylogenetic analyses. But even minor changes in amino acid sequence have been shown to be able to lead to major changes in binding properties or immunogenicity of proteins. [42,100] Therefore, it could be possible that both groups show differences in these features. For instance, *var* group B

derived PfEMP1 molecules might elicit more efficient immune responses compared to *var* group C derived PfEMP1. On the other hand, *var* group C derived PfEMP1 might show weaker binding abilities leading to a continuous but partial removal of these parasites by the spleen.

var group C genes are located centrally on chromosomes in more stable regions. One could speculate that these *var* variants are of older origin and thus are adapted to a host-parasite co-existence guaranteeing long term transmission of the parasite. This could explain the apparently reduced virulence of *var* group C variants. Presumably, the combination of *var* variants contributing to high parasite loads but also to long term survival might provide the highest benefit to the parasite.

.

4.5. Outlook

Despite of the importance of *var* genes as morbidity and mortality driving factor in malaria disease, few studies were performed *in vivo*. Our studies on *var* gene expression in naturally infected children provide a good insight into the complex and dynamic expression of *var* genes *in vivo*. We found that *var* subgroups do not merely reflect different recombination assemblies but also lead to differences in disease upon expression, and we showed the highly dynamic nature of *var* transcription in semi-immune children. However, the observed transcriptional shifts of *var* groups contain a set of *var* genes which are unknown apart of their affiliation to the corresponding *var* group. Subsequent sequence analyses of *var* transcripts will concentrate on these individual genes. Phylogenetic analyses and protein structure predictions will be performed to find potential sequence conservations of *var* transcripts of parasites conferring severe disease.

Based on these initial results, we have to confirm our findings in a larger case-control study, enrolling a larger number of children, but also compare *var* gene expression in different geographical locations. Currently, a larger case-control study is carried out in Tanzania addressing similar questions on *var* group specific expression in children with severe and mild malaria. To test our hypotheses, *var* gene expression could be compared in young children in longitudinal studies in subsequent acute infections in combination with serological studies, analyzing the nature of antibody response in acute and reconvalescence phase, the degree of cross-reactivity and duration of antibody response. Since in young children, the number of previous clinical episodes and build-up of anti-VSA repertoire cannot be easily evaluated, longitudinal studies in infants would be highly interesting following *var* gene expression in malaria naïve hosts. Alternatively and far easier, but not longitudinally, *var* gene expression could also be evaluated in malaria naïve hosts such as in tourists returning from an endemic area who contracted malaria or in volunteers of malaria vaccine studies.

When relating to *in vitro* analysis, more information on regulations of *var* gene expression and *var* switching is desperately needed. Sophisticated techniques such as elaborate transfection protocols, FISH analysis, chromatin immuno-precipitation or yeast-one hybrid systems might dissect the regulatory mechanisms and nuclear proteins involved in *var* gene regulations.

Another central open question addresses the molecular background of our findings. Do *var* group specific protein structures or distinct expression regulations account for the observed *var* group specific expression pattern in different clinical stages? Potential differences in *var* switching rates will be studied in our laboratory in a longitudinal, chip based approach analysing the switching pattern of single *var* genes.

Because of technical limits, most studies on *var* gene expression are based on RNA level to date. However, potential post-translational regulations and transport limitations should be taken into account. Preliminary MS/MS analyses on PfEMP1[101] raises hopes that also advanced proteomics technologies can be applied. Though, the low abundance of PfEMP1 remains a challenge.

Most studies on *var* gene structuring and regulation are based on *in vitro* analysis of culture strains, above all of 3D7 parasites. It is highly tempting to accept these findings as a general rule valid for all *var* genes while disregarding the fact that this strain only reflects a small subset of the enormous *var* diversity. In addition, 3D7 was in culture for a long time and might not reflect parasites *in vivo* anymore under constant survival pressure. Therefore, findings of initial *in vitro* studies should also be challenged *in vivo.*

Moreover, most original findings on binding abilities of PfEMP1 variants were done by static binding assays on individual recombinantly produced domains. [36,38,40,42,87,102] However, static binding conditions do not reflect physiological shear pressure in blood vessels *in vivo* while the analysis of single *var* domains disregards potential tertiary protein structures which are also influenced by numerous conserved cysteines in the adhesive domains yielding disulfide bonds among potentially different parts of the protein.

Therefore, some technical approaches might have to be reassessed and some dogmas should be re-opened for discussion.

To overcome the enormous diversity of *var* genes, large scale screening approaches on binding or serology characteristics of culture and field derived *var* genes in combination with *in vivo* studies on PfEMP1 pathology might reveal distinct *var* genes relevant for disease or conserved immunogenic *var* epitopes as a potential target for an anti-morbidity malaria vaccine. Our studies confirmed the important role of this protein family in virulence and survival of *P. falciparum* and narrowed the more virulent subset of *var* genes down raising hopes that PfEMP1 and thus malaria morbidity can indeed be successfully combated.

Supplement I: Analysed sequences of Longitudinal Study

Dynamics of *msp2* genotypes and recurring *var* transcripts

child	week	dynamics of *msp2* genotypes	number of different *var* sequences per infecting strain					recurring *var* sequences				
			upsB-5'UTR	upsC-5'UTR	DBL1α	DBL2β	CIDR1α	upsB-5'UTR	upsC-5'UTR	DBL1α	DBL2β	CIDR1α
S07	0		1.3		2.3	*0.7						
	1.5				1.0	*0.5						
	2					*0.5						
S08	0		3.5		1.0							
	4				*2.0							
	5		0.5		1.0							
S12	0		*1	*0.2	0.8							
	1.5		1.3	0.3	3.8							
	3		*0.5		1.5							
L02	0			*1.5	*1.0	*1.0						
	4				1.0							
	10		*1.5	*0.5	*1.5	*0.5						
	12			*0.5								
L03	6			*0.3			0.6					
	10				*1.3							
L06	0				2.8							
	2		*2.7		4.0	*0.7						
	4		0.3		1.0							
	8				*1.8	*0.3						
	10				*2.0							
	12				1.8							
	14				2.7							
L08	0				1.3	*0.3						
	2		0.3	*0.3	*2.3							
	4		3.7	*0.3	2.7		1.5					
	8			*0.5	*3.0							
	12		1.0		2.8		6.0					
	14		*0.3		3.3							
L11	0		2.0	3.0		*3.0						
	10				*0.5	0.5	0.5					
	12		1.0	0.5	1.5							
L12	0		1.7	1.0	3.3		0.5					
	2		0.3	*0.7	0.3		*0.3					
	4		*0.8	*0.3	1.3		*0.8					
	8				*0.7							
	12		0.2	*0.4	0.6	*0.8						
L13	2				*2.0							
	4				*1.0							
	6				*1.5							
	12				*2.0							
	14				*0.5	*0.5						
L19	4				1.0		*2.0					
	14			*1.0								

Legend to figure "dynamics of *msp2* genotypes and recurring *var* transcripts"

Summary of distribution of *var* sequences per child, time point and domain. The number of different analyzed *var* sequences is divided by the number of present *P. falciparum* infections. Time points and regions with less than 20 sequences analyzed are marked with an asterisk. Recurring *msp2* genotypes within one child are indicated as filled boxes in the same column. Recurring *var* sequences within one region are indicated as filled boxes in the same column.

DBL1α

```
L5.06b    (AY462708)    ARSFADIGDIVRGKDLYLGYDDKEKKR----REKLENNLKDIFAKIHEEV----TNGKKG 52
S5.12f    (AY462715)    ARSFADIGDIIRGRDLYRGNRKKNQKQT--ERDQLESKLKEVFGNIYKELKNGKTNGKKG 58
L2.06k    (AY462713)    ARSFADIGDIVRGKDLYLGNTHESTQR-----IILENNLKTIFGHIYEEL---IKNGSNG 52
L3.08a    (AY462772)    ARSFADIGDIIRGKDLYLGDKGN---------DKLEKNLKTIFGDIYKELSRTSTSGRNG 51
L2.06c    (AY462781)    ARSFADIGDIVRGKDLYIGNKKER--------ARLEENLKKIFKEIHGGL-------KDG 45
L1.02b    (AY462731)    ARSFADIGDIVRGRDMFRSNEN----------VEKGLRVVFKNIHENLRGPA------ 42
L7.13a    (AY462745)    ARSFADIGDIVRGRDMFKSNED----------VEKGLKVVFKKIQRKLNGAA------ 42
L3.02a    (AY462737)    ARSFADIGDIVRGRDMFKRNDK----------DAVQKGLRAVFKKINDNLNEKK------ 44
L4.13a    (AY462738)    ARSFADIGDIVRGRDMFKSNEK----------VEYGLRKLFKKIHENLRGPA------ 42
L8.06c    (AY462743)    ARSFADIGDIVRGIDIFKPNDK----------DEVQNGLKKVFKNIHDSLSSDV------ 44
L1.06L    (AY462746)    ARSFADIGDIVRGRDMFK--SN----------EKVEIGLKKVFDKIYNKLRPEG------ 42
L6.03c    (AY462732)    ARSFADIGDIVRGKDMFKRTDN----------DEVWKGLRAVFGKIYNGLTPQA------ 44
L7.08h    (AY462733)    ARSFADIGDIIRGKDMFKRTDK----------DYVENGLKAVFKKIYEGLKNNG------ 44
S5.12d    (AY462734)    ARSFADIGDIVRGRDMFK--SN----------DNIEKGLRVVFRKIYDGLKKS------- 41
L3.06d    (AY462735)    ARSFADIGDIVRGRDMFKSNED----------VEKGLQVVFKKIYEGLKGEG------ 42
S3.07b    (AY462736)    ARSFADIGDIVRGTDMFKPNKD----------DKVEKGLRAIFKNIYEGLKGEG------ 44
L7.08i    (AY462747)    ARSFADIGDIVRGRDMFKRTDN----------VENGLKKVFDKIYDKLGTQK------ 42
L3.08c    (AY462729)    ARSFADIGDIVRGRDMFLPNKD----------DKVQKGLQVVFEKINNGLKKIG------ 44
L1.12i    (AY462730)    ARSFADIGDIVRGRDMFKSN----------DNIEKGLRVVFRKIYDGLKKSG------ 42
L7.06b    (AY462748)    ARSFADIGDIVRGRDMFKSNNV----------DAVQEGLKVVFKKIYNRSTHHA------ 44
L6.03d    (AY462739)    ARSFADIGDIVRGIDMFKPNVH----------DKVEKGLREVFKKIHDEMEGEV------ 44
L2.08g    (AY462740)    --SFADIGDIVRGIDMFKPNVH----------DKVEKGLREVFKKIHDEMEGEV------ 43
S1.07g    (AY462741)    ARSFADIGDIVRGIDMFKPNVH----------DKVETGFREVFKKIHDGMEDEV------ 44
L1.12e    (AY462742)    ARSFADIGDIVRGRDMFKSN----------DNVENGLKAVFKKINNGLNDKG------ 42
L3.12c    (AY462744)    ARSFADIGDIVRGKDMFKPNTV----------DKVHEGLKVVFQKIHDKLKQPE------ 44
L3.08b    (AY462846)    ARSFADIGDIVRGRDLFYGNPQESAR-----RDKLEENLKRIFGKIYEELTKDKK----- 50
L6.06e    (AY462830)    ARSFADIGDIIRGKDLYLGD-KQEKD-------NLQNNLKRIFKNIYDNLK--------- 43
S3.12o    (AY462829)    ARSFADIGDIVRGKDLFLGGPSQEKI-------KLEENLKNIFENIKK---------- 41
L1.06m    (AY462828)    ARSFADIGDIVRGKDLFLGAPNKEKI-------KLEENLKKIFDNIKN---------- 41
L1.06c    (AY462834)    ARSFADIGDIIRGKDLYLGKNESEKG----TKEKLRGNLEKIFDKIRN---------- 44
L8.08g    (AY462851)    ARSFADIGDIVRGKDLFLGKNESEIK----RKEKLQRNLEKIFAQIKN---------- 44
L6.03b    (AY462831)    ARSFADIGDVIRGRDMFLGNNEKDMT----EKQKLQSNLKKIFGNFME---------- 44
L5.06e    (AY462850)    ARSFADIGDIIRGRDMFLGNNEKDMT----EKQKLQSNLKKIFEKIRN---------- 44
S3.12a    (AY462833)    ARSFADIGDIIRGKDLFLGHNHK--------KKPLLDNLEKIFNRFQK---------- 40
S3.12b    (AY462832)    ARSFADIGDIIRGKDLFLGNNKE--------NEKLQENLKRIFKNIYANLK--------- 43
PFD1235w  (AY462842)    ARSFADIGDIIRGKDLYRRDSRT---------DKLEENLRKIFANIYKELKNGK------ 45
PF11_0521 (AY462836)    ARSFADIGDIIRGKDLYLGNGDY--------KEKVSNNLRAIFNKIYENLN--------- 43
L5.06d    (AY462839)    ARSFADLGDIIRGKDLYLGHKLG--------NNKLEARLQTMFQNIKE---------- 40
L1.12b    (AY462840)    ARSFADIGDIIRGKDLYLGHEQG--------NNKLEARLEAMFDNIKK---------- 40
L2.06f    (AY462844)    ARSFADIGDIIRGKDLFLGHKQR--------KRKLEENLRNIFKNIHDHLT--------- 43
S3.12c    (AY462843)    ARSFADIGDIIRGKDLYRGGGRG--------RKQLEENLQKIFGNIYNELTRTATSGN-- 50
L3.08g    (AY462821)    ARSFADIGDIIRGKDLYLGDQQE--------KAKLENNLKNIFAKIHENLNDI------- 45
PF08_0140 (AY462835)    ARSFADIGDIVRGKDLYLGDQQE--------KLYLENNLKKIFVKIHENLND-------- 44
L1.12c    (AY462837)    ARSFADIGDIIRGKDLFLGTYQE--------KKSLEENLKNIFRKLYKELTKYKE----- 47
S1.07e    (AY462824)    ARSFADIGDIVRGKDLFLGHSKE--------KEKLEQNLKNLFKQLYEELKKNNK----- 47
L6.06b    (AY462728)    ----ADIGDIIRGKDLFIGYDKKDRN----EKKQLQQNLKEIFGKIYDNLMKDVRKDK-- 50
L1.08b    (AY462818)    ARSFADIGDIIRGKDLFLGYDKKDRD----QKKKNQDNLKNIFAKIYENLMEDLTKDN-- 54
L8.08b    (AY462826)    ARSFADIGDIIRGKDLFIGYNQKDRN----EKKQLQENLKKIFRKIHEGLTTEKR----- 51
L1.08d    (AY462845)    ARSFADIGDIIRGKDLYLGYDQKEKT----RRENLENKLKEIFGNIYEGLTKEKR----- 51
L3.12d    (AY462827)    ARSFADIGDIVRGKDLYLGTNEEKKP---------LEENLKEIFKKLYEKLNDPQK----- 47
S3.12i    (AY462823)    ARSFADIGDIIRGKDLYIGNRKEKE-------KVKLQQRLKEIFAKIHNGLPQGK------ 48
L2.06e    (AY462773)    ARSFADIGDIVRGRDLYGGNK---------KKDKLEKNLKTIFKEIYDKLDG-------- 43
L7.06a    (AY462775)    ARSFADIGDIVRGKDLYIRNK---------KKDKLQENLKKIFKEIYDKLDG-------- 43
S3.12k    (AY462774)    ARSFADIGDIVRGRDLYRLDKG--------KRDKLEENLKTTFKKIYDNLFK-------- 44
L3.12b    (AY462776)    ARSFADIGDIVRGKDLFRGNTK--------EKDRLQSTLKRIFQKIYNNLMEDLKKDP-T 51
S8.08b    (AY462777)    ----ADIGDIVRGKDLYIRNKR--------EKDRLQSTLKRIFQKIYKNLMEDLRKSSRT 48
S3.12J    (AY462778)    ARSFADIGDIVRGRDLYGGNSK--------EKVKLEKKLKKIFQKIYNNLMQDLENDKIK 52
L5.12a    (AY462770)    ARSFADIGDIIRGKDLYLGDN---------GKDKLEENLKTIFGKIYDKLDGK------K 45
L3.19a    (AY462771)    ARSFADIGDIIRGKDLYRGNN---------GKDKLENNLKEIFKKIYGKLT--------- 42
L2.06g    (AY462785)    ARSFADIGDIIRGKDLYRGNN---------RKDKLENNLKTIFGIIYEGLT--------- 42
L6.06a    (AY462780)    ARSFADIGDIVRGKDLFRGNN---------GKDKLQDNLKKIFEKIYNNLN--------- 42
S1.07f    (AY462787)    ARSFADIGDIIRGKDLFIGNN---------KRDKLEKQLQKYFKKMYEKLTDP------- 44
L1.12J    (AY462788)    ARSFADIGDIVRGKDLYIGNKG--------EKKKLEEKLKQYFKKIYNDLTKDN------ 46
L5.08e    (AY462752)    ARSFADIGDIVRGKDLYRRDNKKD---------KLEKNLKEIFQKIQG---R-Y------ 41
S5.12c    (AY462755)    ARSFADIGDIVRGKDLYIRNKKKD---------KLEDNLQKIFKMIQG---KNP------ 42
S1.12c    (AY462754)    ARSFADIGDIIRGKDLYRRDNKKD---------KLENNLKKIFKKIQN---K-N------ 41
L8.08i    (AY462750)    ARSFADIGDIVRGRDLYIRNNKKD---------KLEENLKTIFEKIKG---NNN------ 42
S3.12n    (AY462779)    ARSFADIGDIVRGRDLYRGDNKKDK----AQKLKLQDNLNKIFNDIK----NTN------ 46
L7.11a    (AY462751)    ARSFADIGDIVRGKDLYLGGNTKEK----KKREDLEKNLKNIFENIT----KGN------ 46
S3.12f    (AY462847)    ARSFADIGDIVRGRDLYRGNN--------KRRQQLEDNLKKIFENIK----NEN------ 42
L5.08b    (AY462753)    ARSFADIGDIVRGKDLFLGHNQ--------RKRKLEENLKQMFENIQS---TI------- 42
S5.12a    (AY462825)    ARSFADIGDIVRGKDLFLGHSK--------EKEKLEQNLQNIFEKLKS---SN------- 42
L3.08d    (AY462820)    ARSFADIGDIIRGKDLFLGYNKSDQ----KEKEQLQDNLEKIFRNIQD---KNR------ 47
```

```
L7.06e    (AY462841)    ARSFADIGDIVRGKDLYLGNKKKNQDI--REKEKLEDNLKTIFANIYKELSRTK------ 52
L7.06f    (AY462807)    ARSFADIGDIVRGKDLYLGNKKKNQDI--REKEKLEDNLKTIFANIYKELSRTK------ 52
L8.06g    (AY462808)    ARSFADIGDIVRGKDLYRGNKKEN-----KQRKQLDDSLKEIFENIK----ESD------ 45
L3.13a    (AY462809)    ARSFADIGDIVRGRDLYGGGGRGK------GKEKLEQKLKDIFGDIYKELSTKN------ 48
S8.08a    (AY462705)    ARSFADIGDIIRGKDLYRRDKGEE--------TKLEKNLKKIFAKIHDKLDDT------- 45
S3.12h    (AY462724)    ARSFADIGDIIRGKDLYRRDKGEE--------TKLEKNLRKIFAKIHGGLTG-------- 44
L6.11a    (AY462795)    ARSFADIGDIIRGKDLYRGNNREN--------DKLEKKLKEYFKKIYEGLKG-------- 44
L8.06f    (AY462812)    ARSFADIGDIVRGIDLYGGN---NKR-----RKQLDDKLKEIFKKIHENLG--------- 43
L7.06c    (AY462720)    ARSFADIGDIVRGKDIFIGNKKENKQ-----RQQLDDKLKDIFGKIHEGLS--------- 46
L1.12h    (AY462704)    -RSFADIGDIVRGKDLYRGNSKEKNR-----RQQLENNLKEIFAKIHEKLTT-A------ 47
S7.08c    (AY462811)    ARSFADIGDIVRGRDLYRGNTQEKKQ-----REKLDNNLKDIFAKIHGNLR--------- 46
L7.12b    (AY462719)    ARSFADIGDIVRGRDLFYGNTQEKTK-----REQLDDKLKEAFGKIHSGLT--------- 46
S3.07d    (AY462810)    ARSFADIGDIVRGRDLYRGNKQEKEQ-----REKLDEKLKEIFKKIHNGLD--------- 46
MAL6P1.4  (AY462849)    -RSFADIGDIVRGKDLFLGTNEEK--------KPLEENLKEIFKKIYENLG--------- 42
L2.06b    (AY462797)    ARSFADIGDIVRGRDLYRRDK---GKG-----DHLEKNLKKYFQQIHNDVTSTSG----- 47
L8.06e    (AY462803)    ARSFADIGDIVRGRDLYRGG----GRG----RKKLDDKLKEIFQQIHSEVTS-SG----- 46
L5.06f    (AY462805)    ARSFADIGDIIRGRDLYRGGGR--GKG----KEKLEGKLKDIFAKIYNDVTS--G----- 47
L1.08e    (AY462813)    ARSFADIGDIIRGRDLYSG----NKVK----KKKLDDSLKTIFGKIYEGLTG-G------ 45
PF08_0103 (AY462819)    ARSFADIGDIIRGKDLFIGYDKKDRVQ----KKKLQDSLKNIFGNIYNELTT-SG----- 50
L7.12c    (AY462814)    ARSFADIGDIVSGRDLYRGN-KKETNQ----REKLEENLRKIFENIYEGLS--------- 46
L7.08e    (AY462815)    ARSFADIGDIVRGRDLYRGN-KKENKQ----RDKLDENLKKIFGKIHEGLS--------- 46
L3.08f    (AY462848)    ARSFADIGDIIRGRDLYRGN-SKE-------KDYLQDKLKKYFQKIYDNLN--------- 43
S1.12a    (AY462798)    ARSFADIGDIVRGKDLYGRN-NRK-------RKQLDENLQKYFENIYYGLSN--E----- 45
L2.08e    (AY462718)    ARSFADIADIVRGRDLYSGN-SKEKKQ----RKQLDDKLKEVFTNIYNEL---RS----- 47
L7.08k    (AY462804)    ARSFADIGDIVRGRDLFYGN-PQEKDQ----RKKLQQNLKTIFKKIHGGS---TK----- 47
L1.12d    (AY462717)    ARSFADIGDIVRGRDLYLGN-PEEIKQ----SHQLDKKLKDIFAKIYEGLSRTRT----- 50
L5.06a    (AY462703)    ARSFADIGDIVRGRDLFRGYNEKDRKE----REQLQDSLKNIFGNIYEKLLEENQ----- 51
L1.12f    (AY462789)    ARSFADIGDIIRGKDLYRRDNKKDK---------LEKNLKKIFEKIYDKLLEENQ----- 46
S3.12e    (AY462714)    ARSFADIGDIIRGRDLFRG-NTYESAQ----RIILENNLKKIFRNIYDKLD---G----- 47
L3.12a    (AY462759)    ARSFADIGDIVRGRDLYGGS-KKEIKQ----RQLLDENLKTIFGKIYEELTKNEK----- 50
L3.12e    (AY462727)    ARSFADIGDIVRGRDLYRGN-KKENKQ----RKQLDDSLKKIFGKIHSGLSKG------- 48
L1.06a    (AY462757)    ARSFADIGDIVRGRDLFRGN-DEEIKQ----RQQLDKKLKEIFKKIYEGLKNG------- 48
L1.06h    (AY462801)    ARSFADIGDIVRGRDLYGGG-NN--KR----RQQLDENLKTIFKNIYDKLLEENQ----- 48
S1.07a                  ARSFADIGDIIRGKDLYLGDNGKDR---------LEENLRKIFKKIYDKLDG-K------ 44
S1.07b    (AY462767)    ARSFADIGDIVRGRDLYRGVNGNDK---------LEKNLKEIFAKIHKEVTS-T------ 44
L3.08h    (AY462769)    ARSFADIGDIVRGRDLFLGYNEKDRKE----KEQLAKNLKDIFAKIHDDVT--------- 47
L1.06J    (AY462706)    ARSFADIGDIVRGRDLYLGDNRKDREQ----REQLEKNLKDIFEKIHKEVT--------- 47
L2.08a    (AY462768)    ARSFADIGDIIRGKDLYRGNNGKDK---------LEENLKRIFQQIHKEVTS-T------ 44
L2.12a    (AY462725)    ARSFADIGDIIRGKDLFIGYNQKDREE----KNQLQENLKKIFAKIHSEVTN-G------ 49
L1.12a    (AY462822)    ARSFADIGDIIRGKDLFRGYNEKDQQE----KAKLQQSLKDIFAKIHSEVTK-GG----- 50
L2.08f    (AY462726)    --SFADIGDIVRGKDLFRGYNEKDQQG----KAKLQQNLKDIFAKIHSDVMK-T------ 47
S1.12b    (AY462709)    ARSFADIGDIIRGKDLFRGYNEKDRNE----KKQLQQNLKTIFGKIHDDVTN-G------ 49
L2.06h    (AY462710)    ARSFADIGDIIRGKDLYLGGRGRD---------QLESNLRKIFAKIYKDVTN-G------ 44
L8.06h    (AY462707)    ARSFADIGDIIRGKDLYRG-NDKEKDQ----RKKLDKNLKTIFGKIYKDVTT-N------ 48
L5.06c    (AY462783)    ARSFADIGDIVRGKDLYRG----VNG-----NDKLENKLKEIFKKIHEGLT--ST----- 44
S3.07a    (AY462765)    ARSFADIGDIVRGKDLYRG----NNRE----NDKLESKLKVIFGKIYKDLT--TT----- 45
L7.08b    (AY462784)    ARSFADIGDIVRGKDLYRGYDQKEKEQ----RDELENNLKTIFGDIYKDVTNGRN----- 51
S7.08a                  ARSFADIGDIVRGKDLFLG-NTHESAR----REKLENKLKDIFGDIYKELR--KN----- 48
L1.12g    (AY462711)    ARSFADIGDIVRGRDLYLG-NPQESTQ----RDKLENNLKTIFANIYKDVT---S----- 47
L1.02a    (AY462762)    ARSFADIGDIVRGKDPFYG-NPQESAQ----RKQLEKNLKDIFKNIYNDVT--RG----- 48
L1.06g    (AY462806)    ARSFADIGDIVRGEDLFLG-TNEE-------KKPLEENLKEIFKNIYNEVM--RG----- 45
L2.08c    (AY462790)    ----ADIGDIIRGKDLYLG-DKGEKK-------KLEENLKTIFTQIYNDVT--KG----- 41
L5.12b    (AY462760)    ARSFADIGDIVRGKDLYLG-NKQEKKQ----RKKLEDKLKDIFKNIYKDVT--RG----- 48
L5.08c    (AY462749)    ARSFADIGDIIRGRDLYRRDNKKDK------TDKLQEQLKKYFQKIYEELKN-------- 46
L8.08d    (AY462799)    ARSFADIGDIIRGKDLYLGGNKKEKNR----RDDLEDKLKNIFAKIYDNLVE-------- 48
L7.08f    (AY462817)    -------------KDLYLGYDDKEKTR----RKKLEEKLQKIFGNIYNELTR-------- 35
S3.12m    (AY462782)    ARSFADIGDIVRGKDLYLRYNRKDKT------DKLQEQLKKYFQKIHDGLTN-------- 46
L6.02a    (AY462716)    ARSFADIGDIVRGKDLYLGYDDEEKN----RRKKLQQNLKNIFGNIYKELTS--T----- 49
S1.08b    (AY462793)    ARSFADIGDIVRGKDLYLGDKEKKQNGKKTEREKLEQKLKDIFEKIHEELK--------- 51
L6.02b    (AY462756)    ARSFADIGDIVRGRDIFRGNDEEKKQ-----RDKLEENLKNIFGKIHDNLT--------- 46
L8.08c    (AY462758)    ARSFADIGDIVRGKDLFYG-NPQEKD----QREKLDENLKTIFKNIK------------- 42
S1.08a    (AY462786)    ARSFADIGDIVRGKDLYLGYDDKEKD----RRRKLEENLKTIFTQIYNDVTS--G----- 49
S5.12e    (AY462794)    ARSFADIGDIVRGKDLFLG-NDEEKK----RREKLEKNLKEIFKQIHEKLTD--Q----- 48
L5.08a    (AY462791)    ARSFADIGDIIRGKDLFLGNTYESAQ-----REKLDDKLKKIFGDIYEGLKK--K----- 48
S1.07d    (AY462764)    ARSFADIGDIVRGKDPFYGNTYESAQ-----REKLESKLKEIFAKIYEGLTT--T----- 48
S1.12d    (AY462792)    ARSFADIGDIVRGKDLFLGNTYESAQ-----RDKLEKNLKEIFGKIHSEVMK-------- 47
S3.12g    (AY462712)    ARSFADIGDIIRGKDLFLGNTYESEQ-----RIILENNLKTIFAKIHSEVTN--G----- 48
L1.06b    (AY462761)    ARSFADIGDIVRGKDPFYGNTHESAQ-----RKVLDKKLKEVFGKIHDDV---------- 45
L2.06L    (AY462763)    ARSFADIGDIVRGRDLFYGNTHESAQ-----RDQLDKKLKEVFGKIHGGL---------- 45
L5.06g    (AY462721)    ARSFADIGDIVRGRDLYRGNKKENKQ-----REKLEAKIKEYFQKIYDDLFQ-------- 47
L2.13a    (AY462722)    ARSFADIGDIVRGRDLFYGNPQESTQ-----RKQLDKKLKVIFGKIHEGL---------- 45
S1.07c    (AY462766)    ARSFADIGDIIRGKDLFYGNTQESAR-----RDELEKNLKTIFEKIHDKL---------- 45
L6.03a    (AY462723)    ARSFADIGDIVRGRDLYRRDKGE--------ETKLEKNLKTIFAKIHSEVTKGGN----- 47
L8.08e    (AY462796)    ARSFADIGDIIRGKDLYRGNNRE--------NDKLEKKLKEYFKKIYEELVK--K----- 45
L1.06k    (AY462838)    ARSFADIGDIIRGRDMWDKDGG---------STEMENNFKKIFNTIRQKLPE-------- 43
L5.08f    (AY462800)    ARSFADIGDIIRGRDLYRRDKGE--------ETKLEKNLKEIFKNIYNELTT-------- 44
L7.08g    (AY462802)    ARSFADIGDIIRGRDLYRRDKGE--------ETKLEKNLKEIFKNIYNELTT-------- 44
```

```
L5.06b    (AY462708)    ----------EIEARYK----DTTNFFQLREDWWTANRQEIWKAITCDAG-----GSQYF 93
S5.12f    (AY462715)    ----------EIETRYKGD--EANNFFKLREDWWTANRATIWEAMTCSEHL---KNSAYF 103
L2.06k    (AY462713)    ----------QIEARYKKDD-EDGNFFQLREDWWYANRATIWEALTCDAR----HDAQYF 97
L3.08a    (AY462772)    K-------KEEIERRYG----SDENFFQLREDWWDANRATVWYAITCGAG----ESDKYF 96
L2.06c    (AY462781)    -----------AKKHYQD---GSGDFFQLREDWWYANRIMVWNAITCGAG-----GSQYF 86
L1.02b    (AY462731)    ------------KNYYADDDG-SGNYSKLREAWWTTNRDQVWKAITCSAPR----DADYF 85
L7.13a    (AY462745)    ------------KSYYNADE--KGNYYKLREDWWMANRDQVWKAITCKAPT----GADYF 84
L3.02a    (AY462737)    ------------ITHYNDG---SGNYVKLREDWWTVNRDQVWKALTCSAPY----NAHYF 85
L4.13a    (AY462738)    ------------KSHYEDGDK-SGNYYKLREHWWIVNRKQLWEAITCIAPR----DAHYF 85
L8.06c    (AY462743)    ------------RKNYPDDG--FGNYYKLREAWWNANRDKVWNALTCSIPY----YANYV 86
L1.06L    (AY462746)    ------------K-NYPD-DG-SGNYYKLREAWWNVNRDQVWRAITCEAPQ----KVDYF 83
L6.03c    (AY462732)    ------------KTHYADEDG-SGNYYKLREAWWKANRDQVWRAITCSAPG----DVNYF 87
L7.08h    (AY462733)    ------------ANVHYKDDG-SGNYYKLREAWWTVNRDQVWKALTCTAPD----NVNYF 87
S5.12d    (AY462734)    ------------KITDYDND---PNYYKLRNDWWTVNRDQVWKALTCNAPD----NVNYF 82
L3.06d    (AY462735)    ------------ANVHYKEDK-DENYYKLRNDWWTANRDQVWKAITCKAPK----GANYF 85
S3.07b    (AY462736)    ------------VKVHYKENK-DGNYYKLREAWWTANRDQVWRALTCNAPY----DANYV 87
L7.08i    (AY462747)    ------------KNYYNNTGN-NVDYVKLREDWWTANRDQVWKAITCKAPH----DANYF 85
L3.08c    (AY462729)    ------------INAYNDGS---GNYSKLREVWWNVNRDQVWRAITCSAPG----DVNYF 85
L1.12i    (AY462730)    ------------INYYNDEN---GNYYKLREAWWNVNRDKVWEAITCEASK----NANFF 83
L7.06b    (AY462748)    ------------KNDYTGDH---PNYYKLREDWWNVNREQVWRAITCKASK----NANYF 85
L6.03d    (AY462739)    ------------KNYYNPDG--SGNYYKLREAWWDVNRNKVWESITCGALP----KSAYF 86
L2.08g    (AY462740)    ------------KNYYNPDG--SGNYYKLREAWWDVNRNKVWEAITCGALP----KSAYF 85
S1.07g    (AY462741)    ------------KNDYNPDG--SGNYYKLREAWWNVNRNKVWEAITCDASY----KSGYF 86
L1.12e    (AY462742)    ------------IRYYDDDG--SGNYYKLREAWWTANRDQVWKAITCGALP----KSAYF 84
L3.12c    (AY462744)    ------------KDYYNADE--KGNFYKLREDWWTANRDQVWKALTCYAPD----KANYF 86
L3.08b    (AY462846)    -------KMVEAKKKYKDEN--GGNFFKLREDWRTANRETVWKAITCAAKV----GDTYF 97
L6.06e    (AY462830)    -------DNEAVKTRYDNDEK-DGNFYQLREDWWALNRKEVWKALTCSVPY----EAYYF 91
S3.12o    (AY462829)    -------GNTELSTLP---------IEKVREYWWALNRKEVWKALTCSVPY----EAYYF 81
L1.06m    (AY462828)    -------ENAELSKLS---------LEKVREYWWAIHRKELWEALTCNAPK----GANYF 81
L1.06c    (AY462834)    -------NDKTLNNLS---------IGQVREDWWALNRKDVWKALTCSAPE----DAKYV 84
L8.08g    (AY462851)    -------NDRTLNNLS---------IGQVREAWWTANRDQVWKAITCNAPY----KAWYF 84
L6.03b    (AY462831)    -------SNANLKKHTLE---------RVREYWWALNRNDVWKALTCSAPY----DANYV 84
L5.06e    (AY462850)    -------DGRIPKNVPIG---------QVREYWWALNREDVWKALTCSADG----SEEYF 84
S3.12a    (AY462833)    -------IYEDINNLPID---------DIREYWWALNRNDVWEALTCSAPY----YADYF 80
S3.12b    (AY462832)    -------DHQALKHYKDDTKN----YYQLREAWWALNRNDVWKALTCSAPY----DANYV 88
PFD1235w  (AY462842)    -------KWAEAKEYYQDDGT-G-NYYKLREAWWALNRKDVWKALTCSAPR----DAQYF 92
PF11_0521 (AY462836)    -------DPKLKKHYQKDAPN----YYKLRDDWWNANREDVWKALTCNAPY----EAQYF 88
L5.06d    (AY462839)    ------------NNNTELGE---LTTAQVREYWWALNRVQVWKAITCKAKE----GDIYS 81
L1.12b    (AY462840)    ------------NNKKQLGE---LSTAQVREYWWALNREDVWKALTCSADG----SEEYF 81
L2.06f    (AY462844)    ----------EEAKKKCQDGD---GNYFKLREDWWYANRRQVWNAITCAAKE----EDTYS 87
S3.12c    (AY462843)    -------KGKTLQKHYKDND---KNFFKLREDWWTANRDQVWKALTCFADG----SEDYF 96
L3.08g    (AY462821)    ----------EAKTYYNSDT---PDFYKLREDWWALNRRDVWKAITCGAG---------- 82
PF08_0140 (AY462835)    ----------KIKSNYNDSE---GNYFKLREDWWTAIRDQVWKAITCNAPK----DVNYL 87
L1.12c    (AY462837)    -------NEAVIKSRYSKDD---PYFYQLREDWWNANRDQVWKAITCGADD----NDKYF 93
S1.07e    (AY462824)    -------ND-AIKTHYEKDG---QNYYKLRNDWWELNRKQVWNAITCGAPK----DAKYL 92
L6.06b    (AY462728)    ------RRQMEAQERYKKDEE-DGNYYQLREDWWDANRGTVWKAITCKAEQ----NNKYF 99
L1.08b    (AY462818)    ------IKRQEAEKRYKKDEE-DGNYYQLREDWWDANRGTVWKAITCKAEQ----NNKYF 103
L8.08b    (AY462826)    --------------SHYNGDGD---NYYKLREDWWYANRKEVWKAITCDASE----DAKYF 91
L1.08d    (AY462845)    --------------SRYKKDGD-TDNYYKLREYWWALNRDQVWKAITCSAGE----DDIYS 93
L3.12d    (AY462827)    --------------SSYNDDG--TGNYYKLREYWWNANRNDVWKAITCEAKS----DDKYK 88
S3.12i    (AY462823)    ------------NSAYLNDG---PNYSKLREDWWNANRYDVWKAITCGAPK----ESKYF 89
L2.06e    (AY462773)    KKGK------EAKDRYKD--DEANNFFKLREDWWDANRHQVWRAITCSAG-----GGRYF 90
L7.06a    (AY462775)    KKKE------EAKDYYKD--TDKN-YYKLREDWWNANRIMVWNAITCGAG-----GGKYF 89
S3.12k    (AY462774)    KKGK------EAEERYNG--DKDPNYYQLREDWWDANREKVWKAIKCNAG-----GGKYF 91
L3.12b    (AY462776)    KELK-------GKERYKD----TENYFQLREDWWNNNRKMVWYAITCGAA-----GGKYF 95
S8.08b    (AY462777)    KVDK-------AKERYEG----DKNYYQLREDWWDANRAKVWYAMTCDAA-----GGTYF 92
S3.12J    (AY462778)    VKNKGKVTKEEAKRRYKDINDIDNNFYKLREDWWDANRAKVWEAITCGHP-----GGTYF 107
L5.12a    (AY462770)    GHKS-------AKEHYKDESR---NYYQLREDWWNANRKMVWYAITCGAGQ----IDKYF 91
L3.19a    (AY462771)    -HRR-------AKEHYKDEDDKDPNYYQLREDWWNNNRIMVWRAITCGAGQ----IDKYF 90
L2.06g    (AY462785)    -DSG-------AKNHYEDATG---NYYQLREDWWNINRKKVWDAITCHVVS----GNNYF 87
L6.06a    (AY462780)    -NRK-------AKTHYEDKDP-EKNFFKLREDWWNANRQQVWYAITCGAG-----GSQYF 88
S1.07f    (AY462787)    ----------RAKKHYED--GAPE-FYKLREDWWNANRQEIWKALTCDAPNG---DVHYF 88
L1.12J    (AY462788)    IKRQ------EAQKRYQD--DKYENYYQLREDWWNANRQEIWKALTRDAPNG---DVHYF 95
L5.08e    (AY462752)    ----------STLTRLPLDK--------VREYWWALNRQEIWKAITCDAGK----NDKYF 79
S5.12c    (AY462755)    ----------SKLGALSLDQ--------VREYWWEENREKIWKAITCNAG-----GYSYF 79
S1.12c    (AY462754)    ----------EDLKSLSLDE--------LREYWWELNRETVWKAITCNAG-----GGKYF 78
L8.08i    (AY462750)    ----------STLKDLPLDA--------VREYWWDANRATVWKALTCDAHE----GDTYF 80
S3.12n    (AY462779)    ----------RRLNDLTDDQ--------IREYWWDANREKVWYAITCDAGG----G-TYF 83
L7.11a    (AY462751)    ----------TELSTLPIEK--------VREYWWELNRKEVWKAITCEAN------GTYF 82
S3.12f    (AY462847)    ----------TDLNKLTTEK--------VREYWWALNRQTVWKAITCEAY------GTYF 78
L5.08b    (AY462753)    ----------DQLNGLSLDE--------VREYWWEANRREVWKAITCHAPN----DAQYF 80
S5.12a    (AY462825)    ----------PQLTNVSLEK--------IREYWWALNRQEIWKAITCDAPD----YAKYF 80
L3.08d    (AY462820)    ----------DKLSDLSIDK--------VREYWWNANRQEIWKAITCDAPE----DAKYF 85
L7.06e    (AY462841)    ----------GAQTYYQDDNG--GNYYLLREDWWALNREKVWKAITCEAEV----EDKYS 96
L7.06f    (AY462807)    ----------GAQTYYQDDNG--GNYYLLREDWWALNREKVWKAITCDAPH----GAQYF 96
L8.06g    (AY462808)    ----------AKLTKLNDDQ--------IREYWWALNRNDVWKAITCDAPH----GAQYF 83
L3.13a    (AY462809)    ----------GAQKHYGDDE----NYYKLREDWWTANRQEIWKALTCDAPD----YAQYF 90
```

```
S8.08a    (AY462705)    ----------VVQKHYSDDDKGTKNYYKLREDWWTANRHTVWKALTCDHRL---AGAHYF 92
S3.12h    (AY462724)    ----------GAKTHYEGDD---PNYYKLRNAWWEANRQEIWKAITCGH-----PGGTYF 86
L6.11a    (AY462795)    ----------DAQNRYNGDG---DNYYKLREDWWDANRHTVWKAITCGAD----AGNKYF 87
L8.06f    (AY462812)    ---------THEKKHY-ENDT--ANYYKLREAWWALNRQDVWKALTCD------AHGTYF 85
L7.06c    (AY462720)    ---------ESAKKHYTDNKD--PNYYQLREDWWTANRATVWKAITCDEDKKL-AGASYF 94
L1.12h    (AY462704)    ---------NGVKSRY-NDAT--GNYYKLREDWWTANRETVWKAITCDD--RL-AGAHYF 92
S7.08c    (AY462811)    ---------NGVKDHY-QDAT--GNFFKLREDWWEENRETVWKALTCDEDKKL-SNASYF 93
L7.12b    (AY462719)    ---------GSAKDHYKDENG--GNYYKLREDWWNANRETIWKALTCEAP----EHASYF 91
S3.07d    (AY462810)    ---------GKAQARYNGDT---DNFYRLREDWWNANRQEIWKAITCDEENKL-ASASYF 93
MAL6P1.4  (AY462849)    ---------IQEKNHYNDTP----DYYKLREDWWIANRDQVWKAITCNAG-----GYSYF 84
L2.06b    (AY462797)    ----SNGE--PLKTRYGSDKD--NDFFKLREDWWTANRETVWKAITCN---AQGN--TYF 94
L8.06e    (AY462803)    ----SNWK--ALKTRYQNDPK--GDFFQLREDWWTANRSTVWKAITCS---EKLANYKYF 95
L5.06f    (AY462805)    ----KNLQ--ALRDRY-NDGS--GNFYKLREDWWTANRHTVWEAITCDDD-NKLGGYSYF 97
L1.08e    (AY462813)    ----------VKERY-TNDG--GDFFQLREDWWTANRETVWKAMTCS---EHLKNSAYF 88
PF08_0103 (AY462819)    ----KNVD--KAKARY-NDPK--GDFFQLREDWWALNREKVWSAITCN---AQ--GNKYF 96
L7.12c    (AY462814)    ------NN--GVKARY-EGDK--ENFYQLREDWWALNRDQVWKALTCS---DDLKDASYF 92
L7.08e    (AY462815)    ------KN--GAQKHY-KGDP--E-FFKLREDWWTANRETVWKAITCS---DELKDASYF 91
L3.08f    (AY462848)    ------GA--QTYYND-NDKD--KNYYQLREDWWIANRDQVWKAMTCSDDDKKLASAHYF 92
S1.12a    (AY462798)    ----TNGQ--ALQARY-NNDT--PYYYKLREDWWNANRETVWKAMTCS---KELDNSSYF 93
L2.08e    (AY462718)    ----TNG----VKDHY-KGDT--PDFLKLREDWWTANRATIWEALTCEAP----EHASYF 92
L7.08k    (AY462804)    ----NNGA---LKPRY-KDDA--PYYYKLREDWWTANRATIWEALTCEA------YGTYF 91
L1.12d    (AY462717)    ----KNGQ--TLQERY-KGD---PEFFQLREDWWTENRETVWKALTCEVG-----GGTYF 95
L5.06a    (AY462703)    ----KNGKNVALQARY-KDTT---NFYELREDWWTANRATVWKALTCD-EENKLRDASYF 102
L1.12f    (AY462789)    ----TNVK--TLQARY-KDTT---NFYELREDWWTVNRDQVWKALICG-APD---DAQYF 92
S3.12e    (AY462714)    ----KNGK--TLQARY-NDPK--GDFFQLREDWWTANRETVWKAITCD-AHEG---DTYF 94
L3.12a    (AY462759)    ----TNKE--AAQKRYK-DND--PNFYKLREDWWYANRKKVWDAITCKAN----DGDKYF 97
L3.12e    (AY462727)    ----------AQTYYN-DDK--ENYYQLREDWWDVNRKKVWDAITCGAP----DEGEYF 90
L1.06a    (AY462757)    ----SNGK--EAQKHYE-DDT--KNYYQLREDWWDANRETVWKAITCKAD----ENDRYS 95
L1.06h    (AY462801)    ----TNGK--TLQERYKKDED--KNFFKLREDWWTANRHTVWEAITCSAD----KGNAYF 96
S1.07a                  ----KGKKQDAKERYK-DDNG--GNYYQLREDWWNNNRKMVWYAITCGAA-----GGEYF 92
S1.07b    (AY462767)    ----RGKNMQALQTRY-GQDG--PDYYQLREDWWNNNRKMVWRAITCGAA-----GGEYF 92
L3.08h    (AY462769)    ----RGTNAEELKTRYKGDDN--NNYSKLREDWWNNNRKMVWRAVTCGVI-----GSHYF 96
L1.06J    (AY462706)    ----KGSN-GKLKARY-GSDK--ENYYQLREDWWEENRETVWKAITCNAG-----GGKYF 94
L2.08a    (AY462768)    ----SGKNKDALKTRY-QNDT--DNYYKLREDWWNNNRKMVWYAITCGAK-----GSQYF 92
L2.12a    (AY462725)    ----KNVNKQSLQARYN-DTD---NYFQLREDWWDANRKKVWDAITCKAND----GDKYF 97
L1.12a    (AY462822)    ----KNVNKQSLQERYK-DTT---NYYLLREDWWALNRNDVWKAITCGAEH----KDKYS 98
L2.08f    (AY462726)    ----SGSNGQALQTRYNGDKD--PNFFQLREDWWDANRAKVWYAMTCGAPD----NAEYF 97
S1.12b    (AY462709)    ------RNGEAAKERYK-DD---PNFFQLREDWWDANRATVWKAITCNARD----NAKYF 95
L2.06h    (AY462710)    ------GKNGELQERYK-DDT--ANYYQLREDWWDANRATVWKAITCEA------NGTYF 89
L8.06h    (AY462707)    -----GKNAEELKARYEGDKE--NYFFQLREDWWTANRETVWKALTCDAD------GSYF 95
L5.06c    (AY462783)    ----NGRNGEAAKKYYF-DPK--GNFYQLREDWWYANRETVWKAIRCSAP----TDANYF 93
S3.07a    (AY462765)    ----RGKTN-GAQTYYN-DNG--GNYYKLREDWWTANRET-------------------- 77
L7.08b    (AY462784)    ----GKTNGEAAKARYKDDTD--GNYYKLREDWWTANRHTVWKALTCKAP----QDADYF 101
S7.08a                  ----GKTNGEAAIKRYGGDGD---NYYKLREDWWDANRHTVWEAITC------------ 88
L1.12g    (AY462711)    ----SGRNGQTLQKRYE-DNG--GNYYQLREDWWTENRETVWKAITCNDDKKL-ANAQYF 99
L1.02a    (AY462762)    ----NKGKTNGAEARYGNDPD----FLKLREDWWTANRETVWKALTCEAPN----NAQYF 96
L1.06g    (AY462806)    ----RNGKTNGAEARYKEDPD--KNYSKLREDWWTANRHTVWEAITCDDDNKL-SNASYF 98
L2.08c    (AY462790)    ----G--KNVALQERYQDDTD--KNYYKLREDWWALNRDQVWKAITCGAP----DDAQYF 89
L5.12b    (AY462760)    -------KNVALQKRYNGDEA--NNFFKLREDWWTVNRDQVWKAITCSEEL---RGDAYF 96
L5.08c    (AY462749)    ----GKTNG-EVKRRYE-DNG--GNYSKLREDWWTVNRATVWKAITCGTHDG----DTYF 94
L8.08d    (AY462799)    ----KKK---EAKDHYE-DKD--GNYFQLREDWWTANRATVWKAITCKADTG----NAYF 94
L7.08f    (AY462817)    ----GKQNGQTLKTRYE-NDT--ENYFQLREDWWTANRETVWKALTC-DAPGG---ASYF 84
S3.12m    (AY462782)    ----G------VKERYK-KDG--DNFYQLREDWWEANRETVWKAITCSDRLGG---NNYF 90
L6.02a    (AY462716)    ----NG-KKSALQERYGND----PNFFKLREDWWDANRETVWKAITCGHP-----GGTYF 95
S1.08b    (AY462793)    ----NG-KKSALQERYTNDG---PEFFQLREDWWTANRHTVWKAITCDVV-----GFDYF 98
L6.02b    (AY462756)    ---------EEAKKSYSDT----TNFFQLREYWWELNRETVWEAITCGVH-----GSDYF 88
L8.08c    (AY462758)    ----RN-N-----EDLKNIP-----LPKVREYWWALNRNDVWKAITCDAK-----AFYYF 82
S1.08a    (AY462786)    ----RT-NGEAAKKHYEDATG---NFFKLREDWWYANRRQVWKAITCNAK-----GFNYF 96
S5.12e    (AY462794)    ----RA-K-----ERYKDENG--GNYFQLREDWWTANRHTVWEAITCEVKS----GSQYF 92
L5.08a    (AY462791)    ----NGKNVDTLQARYKDGKD--PDFFKLREDWWTANRHTVWEAITCNAK-----GYQYF 97
S1.07d    (AY462764)    ----NG-----VKDRYEGDAK--KNFYKLREDWWTANRHTVWEAITCNAK-----GYQYF 92
S1.12d    (AY462792)    -------NNGALQARYKGDK---ENFFQLREDWWDANRETVWKAITCEVK----SGSQYF 93
S3.12g    (AY462712)    ----KNAKKQALQKRYKKDG---PEFFKLREDWWYANRETVWKALTCDDGL---SNSKYF 98
L1.06b    (AY462761)    --------TSGAQTRYKKDGG---NYFQLREDWWTASRETVWYANTCGAG----TSDKYF 90
L2.06L    (AY462763)    --------SEEAKGHYGGDE----NYYKLREDWWTANRHTVWKALTCDAR----DNAEYF 89
L5.06g    (AY462721)    --------NKQDAKERYGSDAP---YYFQLREDWWTANRETVWKALICHAP----NDAEYF 93
L2.13a    (AY462722)    --------KDPEKTKYNDPKG---NYYLLREDWWTANRATIWEALTCHAG----QNDKYF 90
S1.07c    (AY462766)    --------DGKIKSNYNNDTK---NYYQLREDWWTANRETVWKAITCKAN----DGDKYF 90
L6.03a    (AY462723)    -------NAEELKARYKDESG---DYYQLREDWWEANRETIWRALTCRAPH----SAHYT 93
L8.08e    (AY462796)    -------YREEAKDYYKD-TE---NYFQLREDWWALNRDQVWKAITCDA-H----DSRYR 89
L1.06k    (AY462838)    ----------EIQKRYTNREN---KHLELRADWWEANRAKVWEAMKCHI----------- 79
L5.08f    (AY462800)    --------KNGTKERYND-TD---NYFQLREDWWTANRETVWKALTCNADA----SSAYF 88
L7.08g    (AY462802)    --------KNGTKERYND-TD---NYFQLREDWWTANRATIWEAITCGAPN----NAQYF 88
```

```
L5.06b    (AY462708)    RATCSDS-GDNEKHS--TLASNKCRCP-KGD-------------QVPTYFDYVP 130
S5.12f    (AY462715)    RATC----GDEENKS--TRAKDKCRCK-KKDSK-------SETDQVPTYFDYVP 143
L2.06k    (AY462713)    RPTC------NGGKS--S-TPNKCRCKDKKTGK-------DDTDQVPTYFDYVP 135
L3.08a    (AY462772)    RKTC-------SNDT--SHTNEKCRCP---------------IYKVPTYFDYVP 126
L2.06c    (AY462781)    RPTCG-----SGKTA--TRAKDNCRCP---------------TYKVPTYFDYVP 118
L1.02b    (AY462731)    IKNSDGS-----RD---FTSEGKCGHND---------------NRVPTNLDYTP 116
L7.13a    (AY462745)    KKKSVDT-----VD---FTNDGKCGHTE---------------GTVPTNLDYVP 115
L3.02a    (AY462737)    IKSSPDF-----KS---FS-DHKCGHEE---------------SRVLTNLDYVP 115
L4.13a    (AY462738)    LKSSPDF-----KS---FS-DLKCGHYE---------------GAPPTYLDYVP 115
L8.06c    (AY462743)    IKKSDDT-----IV---FTDNRKCGHYE---------------GAPPTNLDYVP 117
L1.06L    (AY462746)    RKDLDGS-----DV---FTSQGYCGRKE---------------APVPTYLDYVP 114
L6.03c    (AY462732)    RKGLDGK-----II---FSDNGHCGRNE---------------RNVPTNLDYVP 118
L7.08h    (AY462733)    RKYSDGS-----SN---FSSEGKCGHKE---------------GTVPTYLDYVP 118
S5.12d    (AY462734)    RKYSDGS-----SN---FSSEGKCGHKE---------------GSPLTNLDYVP 113
L3.06d    (AY462735)    RKNSDGS-----RG---FTSQGKCGHYE---------------GAPPTNLDYVP 116
S3.07b    (AY462736)    RRKSDRT-----MD---LTSQGQCGRNENN-------------GYPLTNLDYVP 120
L7.08i    (AY462747)    RKKSDGT-----QL---FSSQGQCGHTE---------------GTVPTNLDYVP 116
L3.08c    (AY462729)    RKISGDT-----RT---FENAGKCRRHD---------------NKVPTNLDYVP 116
L1.12i    (AY462730)    RKISEDT-----RT---FENAGKCRRHD---------------NSVPTNLDYVP 114
L7.06b    (AY462748)    RKISGDT-----KV---FTSEGQCGHSE---------------TNVPTYLDYVP 116
L6.03d    (AY462739)    MQSEDNK-----QL---FS-YLKCGHNKK--------------NDPPTNLDYVP 117
L2.08g    (AY462740)    LQSEDNK-----QL---FL-YPKCGHNNK--------------NDLPTNLDYVP 116
S1.07g    (AY462741)    MQ-------------------------------------------------- 88
L1.12e    (AY462742)    MQSEDNK-----QL---FS-NPKCGHSNG--------------GAPLTNLDYVP 115
L3.12c    (AY462744)    IYKSDNL-----LN---FS-SDRCGHNNN--------------DGPLTNLDYVP 117
L3.08b    (AY462846)    MESRTNS-----YK---FS-NGQCGHYE---------------DAPPTNLDYVP 127
L6.06e    (AY462830)    TYKSDNF-----RT---FS-GYWCGH-K-E-------------GSVPTNLDYVP 121
S3.12o    (AY462829)    TYKSDNF-----RT---FS-GYWCGH-Y-E-------------GAPPTNLDYVP 111
L1.06m    (AY462828)    VYKLDGP-----K----FS-SDRCGHNY-N-------------GDPLTNLDYVP 111
L1.06c    (AY462834)    KYFPSNT-----TT---VS-SDKCGH-N-D-------------IDVPTNFDYVP 114
L8.08g    (AY462851)    MHSEDNK-----QL---FS-DYKCGH-Y-E-------------GSPLTNLDYVP 114
L6.03b    (AY462831)    RRKSDRT-----MD---FTSQGYCGHSE-T--------------NVPTNLDYVP 115
L5.06e    (AY462850)    KKPSSHE-----YS---FS-NGQCGHRD-E---------------NVPTYLDYVP 114
S3.12a    (AY462833)    KKKSGNT-----YN---FTTEGYCGRNE-G---------------APPTNLDYVP 111
S3.12b    (AY462832)    RRKSDRT-----MD---FTSQGYCGHSE-T--------------NVPTNLDYVP 119
PFD1235w  (AY462842)    IKSSVRD-----QT---FS-NDYCGHGE-H---------------EVLTNLDYVP 122
PF11_0521 (AY462836)    IKPSDKE-----HS---FS-SEYCGHYK-N-------------GDPLTNLDYVP 119
L5.06d    (AY462839)    KTANG-N-----TT---LW-NDNCGHHV-K-------------QDVPTNLDYVP 111
L1.12b    (AY462840)    IQSEGAA-----KS---FS-NEYCG-HE-Q-------------GSVPTNLDYVP 111
L2.06f    (AY462844)    KITDNGT-----TT---HS-YDKCGHHV-D-------------QDVPTNLDYVP 118
S3.12c    (AY462843)    IQSEKNT-----KS---FT-NPKCG-HD-E-------------NKVLTNLDYVP 126
L3.08g    (AY462821)    --TSDKY-----FR---H------------------------------------ 90
PF08_0140 (AY462835)    EYNSGKF-----FK---FS-NDQCGHNN-G-------------GAPPTNLDYVP 118
L1.12c    (AY462837)    KKSSGGL-----YV---FS-GGKCGRNE-T-------------NVP-TNLDYVP 123
S1.07e    (AY462824)    EKKDGGA-----YG---CT-HTSCHCA--G-------------GDVLTNFDYVP 122
L6.06b    (AY462728)    RH-TCGG-----GK---NPTHAKCQCVT---------------HDVPTYLDYVP 129
L1.08b    (AY462818)    RH-TCGG-----GK---NPTHAKCQCVT---------------HDVPTYLDYVP 133
L8.08b    (AY462826)    RE-KDSN-----GN---TCTVNKCKCVD---------------GDPPTNLDYVP 121
L1.08d    (AY462845)    KTITYGT-----TA---TSN-GKCGHDD---------------QDVQTYLDYVP 123
L3.12d    (AY462827)    IIKPDGT-----TK---DSNKGQCRCIT---------------GDPPTNLDYVP 119
S3.12i    (AY462823)    RPTCSND-----KA---DT-HEKCQCIG---------------GTVPTYFDYVP 119
L2.06e    (AY462773)    RATCSDE-----KHP--TPTDDKCQCINF---------------SVPTYFDYVP 122
L7.06a    (AY462775)    RHTCGSG-----K----TPTDDKCHCTNH---------------GVPTYFDYVP 119
S3.12k    (AY462774)    RNTCGGG-----QNR--TATNHDCQCINF---------------SVPTYFDYVP 123
L3.12b    (AY462776)    RKTCGIN-----DTW----TEGNCRCAAG---------------DVPTYFDYVP 125
S8.08b    (AY462777)    HATCGNN-----DSW----TEKKCRCAANI--------------RVPTYFDYVP 123
S3.12J    (AY462778)    RQTCN-G-----QNK----TQNNCRCRTN---------------DVPTYFDYVP 136
L5.12a    (AY462770)    RDACSGG-----T----TATNKKCRCATN---------------YVPTYFDYVP 121
L3.19a    (AY462771)    RDACSGG-----T----TLTQGKCRCPSH--------------KVPTYFDYVP 120
L2.06g    (AY462785)    RQTCSKG-----Q----GGTQGKCQCIDQ---------------TVPTYFDYVP 117
L6.06a    (AY462780)    RPTCGSG-----EKG--NATTGRCRCTTN---------------YVPTYFDYVP 120
S1.07f    (AY462787)    RKTCSMG-----Q----SHVNDKCTCANG---------------DVPTYLDYVP 118
L1.12J    (AY462788)    RKTCSMG-----Q----SHVNDKCTCANG---------------DVPTYFDYVP 125
L5.08e    (AY462752)    RQTCNDD----GTSS--HAIH-QCRCENKSG---------EPQDQVPTYFDYVP 117
S5.12c    (AY462755)    RPTCGSS----GETP--HVTPSQCRCINFS---------------VPTNFDYVP 112
S1.12c    (AY462754)    RNTCAG-----LTNW--TETHDKCHCINGD---------------VPTYFDYVP 110
L8.08i    (AY462750)    RATCSDSA--DKKGP--SVAKDHCRCDGDKPG------HDKPNIDPPTYFDYVP 124
S3.12n    (AY462779)    RATCSD-----ENNP--TQTQNNCQCIG---------------QAVPTYFDYVP 115
L7.11a    (AY462751)    HATCGR-----GRDA--TLTQNNCRCGDGKKPG---TNADNPNTDPPTYFDYVP 126
S3.12f    (AY462847)    HATCGR-----GRDA--TLTQGDCRCA---------------NADVPTYFDYVP 110
L5.08b    (AY462753)    RGTCGDN----EKTA--IQASHKCRCKDENGK--------NETNQVPTYFDYVP 120
S5.12a    (AY462825)    GQTCGGG----TTAT--IT---QCRC---------------ATNYVPTYFDYVP 110
L3.08d    (AY462820)    RKTCSND----QEDT--HE---KCRC---------------ATNDVPTFFDYVP 115
L7.06e    (AY462841)    K-TANG-----------NTTLWNYKCGH-------------DEGIVPTNLDYVP 125
L7.06f    (AY462807)    RGTCGGG-----NNA--KRAPSHCRCENKLGE---------PQDQVPTYFDYVP 134
L8.06g    (AY462808)    RGTCGGG-----NNA--KRAPSHCRCENKLGE---------PQDQVPTYFDYVP 121
L3.13a    (AY462809)    RGTCGSDE----KTA--TRASHKCRCENKIGE---------PQDQVPTYFDYVP 129
```

```
S8.08a    (AY462705)    RPTCSMN---GS--C--SQANKYCRCNGDQPG------NDKPNTDPPTYFDYVP 133
S3.12h    (AY462724)    RHTCNGQ----------NKTESNCRC---------------ITHDVPTYSNYVP 115
L6.11a    (AY462795)    RVTCNDN---GT--L--SQATKQCRCQKK---------DGANADQVPTYFDYVP 125
L8.06f    (AY462812)    RATCSER---NG-GC--SQANDKCRCPKTSD--------GKANDQVPTYFDYVP 125
L7.06c    (AY462720)    RPTCGGN---RQ-GP--SQAHDNCRCPN-------------GNDQVPTYFDYVP 129
L1.12h    (AY462704)    RPTCGSN---EN-TA--IQTSHKCRCKD---------------DQVPTYFDYVP 125
S7.08c    (AY462811)    HATCDSG---DNRGG--AQATKQCRCTKSS---------GANADQVPTYFDYVP 133
L7.12b    (AY462719)    RQTCN-----DDGTS--SNANHKCRCKNKKV--------TNETDQVPTYFDYVP 130
S3.07d    (AY462810)    RATCGG----DEKTG--TQASHKCRCKDKKG--------KNETDQVPTYFDYVP 133
MAL6P1.4  (AY462849)    RKTCGGD---NEKNS--TLASNKCRCKDEKG--------EHDTDQVPTYFDYVP 125
L2.06b    (AY462797)    RATCSDR-NG---SF--SQATKQCRCDG---------RNGTNADQVPTYLDYVP 133
L8.06e    (AY462803)    RATCSDSGDGK--GP--SQARNQCRCKD---------ENGNNTDQVPTYFDYVP 136
L5.06f    (AY462805)    RATCGDS---T--SP--SVARNKCRCD------------GKDAHQVPTYFDYVP 132
L1.08e    (AY462813)    RVTCSDK---Q--GE--SIANHKCRCPMT--------SDGKPNDQVPTYFDYVP 127
PF08_0103 (AY462819)    RPTCSG-------GE--SIAHNKCTC-----------INGDP----PTYFDYVP 126
L7.12c    (AY462814)    RPTCS--DRKG--SC--SQAKDNCRCD------------GSNTDQVPTYFDYVP 128
L7.08e    (AY462815)    RATCS--DLNG--SF--SQATKQCRCE------------GAN--VVPTYFDYVP 125
L3.08f    (AY462848)    RATCD--TGQG--P---SVANHYCRCDGR---------NGTNAHQVPTYFDYVP 130
S1.12a    (AY462798)    RPTCD--TGKG--P---SVAKDHCRCDGR---------NGTNADQVPTYFDYVP 131
L2.08e    (AY462718)    RATCND-TGQG--P---SQTHNKCRCEKKR-------GTSDNVSIVPTYFDYVP 133
L7.08k    (AY462804)    RATCGG-DSQG--P---SQARNQCRCEG-------------ANIVPTYFDYVP 125
L1.12d    (AY462717)    RKTCND-TGQG--P---SQTHNKCRCDG------------SNADQVPTYFDYVP 131
L5.06a    (AY462703)    RVTCSD---TK--GP--SVANHYCRCGDDKP------DDDKANVDPPTYFDYVP 143
L1.12f    (AY462789)    RKTACN---G---GK--SSTQGKCHCIAG---------------DVPTYFDYVP 123
S3.12e    (AY462714)    RQTCND---SG--TL--SHANHKCRCKKN--------DGTNETDQVPTYFDYVP 133
L3.12a    (AY462759)    RKTACGTR---------TGTQGRCRCAANI--------------DPPTYFDYVP 128
L3.12e    (AY462727)    RKTACAG----------TRTNDKCRCKGD---------------QVPTYFDYVP 119
L1.06a    (AY462757)    ANIEHDRT---------GVSHGRCGHEDD---------------NVPTNLDYVP 125
L1.06h    (AY462801)    RATCNSADGKY--Q---YQTQNQCRCGDDKKPKS---GKPGEVNIVPTYFDYVP 142
S1.07a                  RKTCG-TG---------TPTNKQCRCTTRV---------------VPT-FDYVP 120
S1.07b    (AY462767)    RKTCG-SAK--------SPTNKQCRCTTRV---------------VPTYFDYVP 122
L3.08h    (AY462769)    RPTCN-GGT--------SSTNTQCRCTIET---------------VPTYFDYVP 126
L1.06J    (AY462706)    RNTCG-EGT--------SPTQPDCRCIGAT---------------VPTYFDYVP 124
L2.08a    (AY462768)    RQTCG-SG---------EWTKDKCQCVTD----------------VPTYFDYVP 120
L2.12a    (AY462725)    RKTAC-AGT--------SQTDGRCRCVT-----------------DVPTYFDYAP 126
L1.12a    (AY462822)    KYLAG-RIT--------TVSNEKCGHVDD---------------DVPTNLDYVP 128
L2.08f    (AY462726)    IKTACGGGT--------TPTNKKCRCVST---------------DPPTYFDYVP 128
S1.12b    (AY462709)    RGTCGGDNEK---TG--TLTPSQCRCDDKPN------------TDPPTFLDYVP 132
L2.06h    (AY462710)    RATCGSSGE----TP--HVTPSRCRCSDNPN------------TDPPTYFDYVP 125
L8.06h    (AY462707)    RATCDSGDNEK--NT--IRTPSHCRCNDDQ---------------VPTYFDYVP 130
L5.06c    (AY462783)    RQTVCSGGKTP--TQ--G----KCRCID--------------FSVP-TYFDYVP 124
S3.07a    (AY462765)    ------------------------------------------------------
L7.08b    (AY462784)    RQTVCSEGTTP--TN--E----KCQCIN--------------FSVP-TYFDYVP 132
S7.08a                  ------------------------------------------------------
L1.12g    (AY462711)    RGTCGSGNNARR-AP--S----QCRCGNG-----------DVSIVP-TYFDYVP 134
L1.02a    (AY462762)    RGTCGSDEKNAT-RV--K---DKCRCSD--------------NQVP-TYFDYVP 129
L1.06g    (AY462806)    RNTCN-DDGTSS-RA--I---HKCRCDDKTSGGKAIKGSGDVNIVP-TYFDYVP 144
L2.08c    (AY462790)    GGTCVGHGKHST-LA--K---DKCRCEG-------------ANIVP-TYFDYVP 123
L5.12b    (AY462760)    HATCNGKERTEG-YC--RCNGDKPDGDN-------------PNTDPPTYFDYVP 134
L5.08c    (AY462749)    HATCSGSHRKGT-F---SQANHYCRCGNDKPGEDNPNTDP------PTYFDYVP 138
L8.08d    (AY462799)    RTTWS-DNRGG------AQANHYCGCNGD---------------------------- 116
L7.08f    (AY462817)    RPTCGDSESPS-------VAKDHCRCGNDQPGRDKSKAGNGDVNIVPTFFDYVP 131
S3.12m    (AY462782)    RPTCNGGKR----------TEGYCRCNDGKSSG--GKAGNGDVNIVPTYFDYVP 132
L6.02a    (AY462716)    R-NTCVGG-----SP--TQD--NCRCTTH---------------DVPTYFDYVP 124
S1.08b    (AY462793)    RGKACGGQ-----FE--AKN--NCQCING---------------DVPTYFDYVP 128
L6.02b    (AY462756)    R-RTCAGEQR---TE--TYD--KCHCING---------------DVPTNFDYVP 119
L8.08c    (AY462758)    RPT-CSTG-----TG--TQG--RCRCVS---------------TDPPTYFDYVP 111
S1.08a    (AY462786)    RQTACNGG-----SP--TKG--YCRCNGDQ------PNADNPNTDPPTYFDYVP 135
S5.12e    (AY462794)    RGT-CGTG-----TG--TQG--RCRCPNG-------------NNQVPTYFDYVP 123
L5.08a    (AY462791)    HATCDSGDNENTATL--ARN--QCRCS---------DKTSGANTDPPTYFDYVP 138
S1.07d    (AY462764)    RATCSDSDGK-SQYQ--TQN--QCRCQKEDGA--KAGKGSGNVSIVPTYFDYVP 139
S1.12d    (AY462792)    RATCSDS-DEKNATL--AKN--NCRCEG------------ANANQVPTYLDYVP 130
S3.12g    (AY462712)    HATCSDSGDEKNPSP--TDG--QCRCK---------------TNVVPTYFDYVP 133
L1.06b    (AY462761)    RNTCSNDK-----AG--TSG--KCRCND---------------NQVPTYFDYVP 120
L2.06L    (AY462763)    RKTPCAGK-----TA--TEAIHKCRCSDKS----KAIKAGGNVNIAPTYFDYVP 132
L5.06g    (AY462721)    RKTACAGG-----KG--TNR--QCHCIA---------------GEVPTYFDYVP 123
L2.13a    (AY462722)    R-DACSGG-----KG--TNG--QCHCIA---------------GEVPTYFDYVP 119
S1.07c    (AY462766)    RDACSGGK-----TP--TQG--KCRCGD------------GDVTIVPTYFDYVP 123
L6.03a    (AY462723)    KSGADGSITKSVMG--------QCRDVS---------------DVPTNFDYVP 123
L8.08e    (AY462796)    KMGADGSITESAMR--------QCRNVA---------------DVPTNFDYVP 119
L1.06k    (AY462838)    KDLNDKSGPRSTPS---------SYCGYS---------------DHTPLDDYIP 109
L5.08f    (AY462800)    RPTCNDTGRGGAQAKDK------CRCNDDKKPGKANDDVN----IVPTYFDYVP 132
L7.08g    (AY462802)    RGTCG-SGRDATRTPSQ------CRCEKKSGGKAGDVS------IVPTYFDYVP 129
```

```
L5.06b    (AY462708)    QYLRWFDEWA--- 140
S5.12f    (AY462715)    QYLRWFEEWAEDL 156
L2.06k    (AY462713)    QYLRWFDEW---- 144
L3.08a    (AY462772)    QFLRWFEEWAED- 138
L2.06c    (AY462781)    QYLRWFEEWAED- 130
L1.02b    (AY462731)    QFLRWFEEWAED- 128
L7.13a    (AY462745)    QFLRWFEEWAEDL 128
L3.02a    (AY462737)    QFLRWFEEWAED- 127
L4.13a    (AY462738)    QYLRWFEEWAED- 127
L8.06c    (AY462743)    QFLRWFEEWAED- 129
L1.06L    (AY462746)    QYLRWFDEWA--- 124
L6.03c    (AY462732)    QFLRWFEEWAED- 130
L7.08h    (AY462733)    QFLRWFEEWAED- 130
S5.12d    (AY462734)    QFLRWFEEWAEDL 126
L3.06d    (AY462735)    QFLRWFEEWAED- 128
S3.07b    (AY462736)    QFLRWFEEWAEDL 133
L7.08i    (AY462747)    QFLRWFEEWAED- 128
L3.08c    (AY462729)    QFLRWFEEWAED- 128
L1.12i    (AY462730)    QFLRWFDEWA--- 124
L7.06b    (AY462748)    QFLRWFEEWAED- 128
L6.03d    (AY462739)    QYLRWFEEWAED- 129
L2.08g    (AY462740)    QYLR--------- 120
S1.07g    (AY462741)    ----------
L1.12e    (AY462742)    QFLRWFEEW---- 124
L3.12c    (AY462744)    QFFRWFEEWAED- 129
L3.08b    (AY462846)    QFLRWFEEWAED- 139
L6.06e    (AY462830)    QYLRWFEEWA--- 131
S3.12o    (AY462829)    QFLRWFEEWAEDL 124
L1.06m    (AY462828)    QYLRWFDEWA--- 121
L1.06c    (AY462834)    QFLRWFEEWAET- 126
L8.08g    (AY462851)    QFLRWFEEWAED- 126
L6.03b    (AY462831)    QFLRWFEEWAED- 127
L5.06e    (AY462850)    QFLRWFDEWA--- 124
S3.12a    (AY462833)    QFLRWFEEWAED- 123
S3.12b    (AY462832)    QFLRWFEEWAEDL 132
PFD1235w  (AY462842)    QFLRWFEEWAED- 134
PF11_0521 (AY462836)    QFLRWFEEWAED- 131
L5.06d    (AY462839)    QFLRWFEEWA--- 121
L1.12b    (AY462840)    QYLRWFEEWAED- 123
L2.06f    (AY462844)    QHLRWFEEW---- 127
S3.12c    (AY462843)    QYLRWFEEWAEDL 139
L3.08g    (AY462821)    ----------
PF08_0140 (AY462835)    QIC---------- 127
L1.12c    (AY462837)    QYLRWFEEWAED- 135
S1.07e    (AY462824)    QYLRWFEEWAEDL 135
L6.06b    (AY462728)    QYLRWFEEWAED- 141
L1.08b    (AY462818)    QYLRWFEEWAED- 145
L8.08b    (AY462826)    QYLRWFEEWAED- 133
L1.08d    (AY462845)    QYLRWFEEWAED- 135
L3.12d    (AY462827)    QYLRWFEEWAED- 131
S3.12i    (AY462823)    QFLRWFEEWAEDL 132
L2.06e    (AY462773)    QYLRWFEEWAED- 134
L7.06a    (AY462775)    QYLRWFEEWAED- 131
S3.12k    (AY462774)    QYLRWFEEWAEDL 136
L3.12b    (AY462776)    QYLRWFEEWAED- 137
S8.08b    (AY462777)    QFLRWFEEWAEDL 136
S3.12J    (AY462778)    QYLRWFEEWAEDL 149
L5.12a    (AY462770)    QYLRWFEEWAEDL 134
L3.19a    (AY462771)    QYLRWFEEWAED- 132
L2.06g    (AY462785)    QYLRWF------- 123
L6.06a    (AY462780)    QYLRWFEEWAED- 132
S1.07f    (AY462787)    QFLRWFEEWA--- 128
L1.12J    (AY462788)    QYLRWFEEWA--- 135
L5.08e    (AY462752)    QYLRWFEEWAEDL 130
S5.12c    (AY462755)    QYLRWFEEWAEDL 125
S1.12c    (AY462754)    QYLRWFEEWAEDL 123
L8.08i    (AY462750)    QYLRWFEEWAED- 136
S3.12n    (AY462779)    QYLRWFEEWAEDL 128
L7.11a    (AY462751)    QYLRWFEEWAED- 138
S3.12f    (AY462847)    QYLRWFEEWAEDL 123
L5.08b    (AY462753)    QYLRWFEEWAEDL 133
S5.12a    (AY462825)    QYLRWFEEWAEDL 123
L3.08d    (AY462820)    QYLRWFE------ 122
L7.06e    (AY462841)    QYLRWFEEWQKI- 137
L7.06f    (AY462807)    QYLRWFEEWAED- 146
L8.06g    (AY462808)    QYLRWFEEWAED- 133
L3.13a    (AY462809)    QYLRWFEEWAE-- 140
```

```
S8.08a    (AY462705)    QYLRWFEEWAEDL 146
S3.12h    (AY462724)    QYLRWFEEWAEDL 128
L6.11a    (AY462795)    QYLRWFEEWAED- 137
L8.06f    (AY462812)    QYLRWFEEWAED- 137
L7.06c    (AY462720)    QYLRWFEEWAED- 141
L1.12h    (AY462704)    QYLRW-------- 130
S7.08c    (AY462811)    QYLRWFEEWAEDL 146
L7.12b    (AY462719)    QYLRWFEEWAED- 142
S3.07d    (AY462810)    QFLRWFEEWAEDL 146
MAL6P1.4  (AY462849)    QFLRWFEEWAED- 137
L2.06b    (AY462797)    QYLRWFEEWAED- 145
L8.06e    (AY462803)    QYLRWFEEWAEDL 149
L5.06f    (AY462805)    QFLRWFEEWAEDL 145
L1.08e    (AY462813)    QYLRWFEEWAED- 139
PF08_0103 (AY462819)    QYLRWFEEWAED- 138
L7.12c    (AY462814)    QFLRWFEEWAEDL 141
L7.08e    (AY462815)    QYLRWFEEWAED- 137
L3.08f    (AY462848)    QYLRWFEEWA--- 140
S1.12a    (AY462798)    QYLRWFEEWAEDL 144
L2.08e    (AY462718)    QYLRWFEEWAED- 145
L7.08k    (AY462804)    QYLRWFEEWAED- 137
L1.12d    (AY462717)    QFLRWFEEWAED- 143
L5.06a    (AY462703)    QFLRWFEEWA--- 153
L1.12f    (AY462789)    QFLRWFD------ 130
S3.12e    (AY462714)    QYLRWFEEW---- 142
L3.12a    (AY462759)    QFLRWFEEWAED- 140
L3.12e    (AY462727)    QYLRWFEEWAED- 131
L1.06a    (AY462757)    QFLRWFEEWAED- 137
L1.06h    (AY462801)    QFLRWFEEWAE-- 153
S1.07a                  QFLRWFEEWAEDL 133
S1.07b    (AY462767)    QFLRWFEEWAEDL 135
L3.08h    (AY462769)    QFPRWFEEWAED- 138
L1.06J    (AY462706)    QYLRWFDEWA--- 134
L2.08a    (AY462768)    QYLRWFEEWAED- 132
L2.12a    (AY462725)    QFVRWFEEWAED- 138
L1.12a    (AY462822)    QFLRWFEEWAED- 140
L2.08f    (AY462726)    QYLRWVEEWAED- 140
S1.12b    (AY462709)    QYLRWFEEWAEDL 145
L2.06h    (AY462710)    QFLRWFDEWA--- 135
L8.06h    (AY462707)    QYLRWFEEWAED- 142
L5.06c    (AY462783)    QFLRWFEEWA--- 134
S3.07a    (AY462765)    ------------
L7.08b    (AY462784)    QFLRWFEEWAED- 144
S7.08a                  ------------
L1.12g    (AY462711)    QYLRWFDEWA--- 144
L1.02a    (AY462762)    QYLRWFEEWAED- 141
L1.06g    (AY462806)    QYLRWFEEWAED- 156
L2.08c    (AY462790)    QFLRWFEEWA--- 133
L5.12b    (AY462760)    QFLRWFEEWAEDL 147
L5.08c    (AY462749)    QYLRWFEEWAEDL 151
L8.08d    (AY462799)    ------------
L7.08f    (AY462817)    QFLRWFEEWAED- 143
S3.12m    (AY462782)    QYLRWFEEWAEDL 145
L6.02a    (AY462716)    QFLRWFEEWAED- 136
S1.08b    (AY462793)    QYLRWFEEWAEDL 141
L6.02b    (AY462756)    QFLRWFEEWAED- 131
L8.08c    (AY462758)    QYLRWFEEWAED- 123
S1.08a    (AY462786)    QYLRWFEEWAEDL 148
S5.12e    (AY462794)    QYLRWFEEWAEDL 136
L5.08a    (AY462791)    QFLRWFEEWAEDL 151
S1.07d    (AY462764)    QFLRWFEEWAEDL 152
S1.12d    (AY462792)    QYLRWFEEWAEDL 143
S3.12g    (AY462712)    QYLRWFEEW---- 142
L1.06b    (AY462761)    QYLRWFEEWAED- 132
L2.06L    (AY462763)    QYLRWFDEWA--- 142
L5.06g    (AY462721)    QYLRWFEEWAEDL 136
L2.13a    (AY462722)    QYLRWFEEWAED- 131
S1.07c    (AY462766)    QFLRWFEEWAEDL 136
L6.03a    (AY462723)    QYLRWFEEWAED- 135
L8.08e    (AY462796)    QYLRWFEEWAED- 131
L1.06k    (AY462838)    QRLRWFEEWAEDL 122
L5.08f    (AY462800)    QYLRWFEDGQK-- 143
L7.08g    (AY462802)    QYLRWFEEWAED- 141
```

CIDR1α

```
L4.03b   (AY462639)    ---------LQDSIYWRTEKIKKCLENN-NGNRCKKKNKCKDDCDCFKRWVEHK-QQEWE 50
L7.08f   (AY462640)    -----VADMLQDSVEWKT-ELSKCINNNTNGNRCR--NGCNRDCKCYESWAKRK-EKEWG 52
L1.12a   (AY462635)    -----VADMLKDSIHWRTKKIKGCLKNGKK-K-C-GNQQCKVDCECFKRWVKQKKEQEWD 52
L3.08c   (AY462644)    -----VSDMLKDSVHWK-KKLQRCLENGTKIK-CR--NGCNSDCECFLQWVEQKKEKEWK 52
L7.08c   (AY462633)    -----VADMLKDSIHWRTKKIKGCLSNGAKIR-CNKKNKCNNDCDCFQKWVEQK-GKEWM 54
L3.12c   (AY462632)    -----VTDMLQDSISWRTEKIKKCLENNNGNR-CNKKNKCKDDCECFQRWVDKK-KTEWK 54
L4.03c   (AY462637)    -----VDDMLQDSIYWRTEKLSKCLQNGKTIK-CK--DKCKDDCECFERWVGKK-KTEWD 52
L3.12b   (AY462629)    ----------DSIHWRTKKLDKCINNSNESKACKNNNKCKDNCDCFEKWVKHK-QQEWD 49
L3.19a   (AY462628)    ----------DSIHWRTEKLDKCINNSNESKACKNNNKCKDDCDCFKRWVDQK-KKEWM 49
L4.03a   (AY462627)    -----VADMLHDSIHWRTEKLDKCINNSNESKACKNNNKCKDKCDCFQRWVKQK-KDEWK 55
L3.08a   (AY462631)    -----VSDMLNDSIHWRTKKLDKCINNSNESKACKNNNKCNKECGCFEKWVGHK-QQEWG 59
L7.08a   (AY462630)    -----VADMLHDSIHWRTKKIKSCINTTNESKACKNNNKCKTDCGCFQKWVQQK-KDEWD 55
L2.12a   (AY462636)    -----VADMLNDSIYWRTEKIKGCLENGTKTR-CKKNN-CKDNCGCFQKWVKQK-ETEWT 53
L3.08b   (AY462638)    -----VPDMLKDSIHWK-KKLEKCLKNG--SK-IKCTDRCKTPCDCFEKWVGQK-EKEWK 51
L3.12d   (AY462643)    -----VADMLKDSVEWR-NELGSCINKNKENT-CKTPKKCNKECTCFLKWVVKKKE-EWG 53
L7.08d   (AY462645)    -----VADMLQDSIKWR-DEHGNCINKDNDNT-CKN-KQCNSKCECFAKWVVKKKD-EWS 52
L6.11a   (AY462641)    -----VDDMLHDSIKWR-KELDNCL-KNENKQ-CIR-K-CNGKCDCFQRWVDQKKE-EWK 50
L3.12e   (AY462642)    -----VDDMLQDSIYWETQKLEKCL-KNEKKK-CGK-KICNGDCDCFQKWIGQKKEQEWT 53
L7.08b   (AY462634)    -----VPDMLKDSIHWK-KKLEKCL-KNGTKT-CGN-QKCKGNCDCFQRWIDKKKD-EWE 51

L4.03b   (AY462639)    KIVQHFNTQDISARGGNGNVVGFFS------------LSHDVLLEQVLD----KGVLLTS 94
L7.08f   (AY462640)    NIVKHFYKQDDIVE------VGFLAEI----------MKHDIVLEGVLQ----KKELLQI 92
L1.12a   (AY462635)    KIKEHFG--------------KQTDLG---------EWEPNDLLEQVLE----KGVLLTS 85
L3.08c   (AY462644)    PIKEHFGNQEAFKNKEENSASQMLGEE---------MNSADFVLDGVLKLEFYKDNSENN 103
L7.08c   (AY462633)    AIKEHFGNQEAFKNKGKNSASQMLGEE---------MSSPDFVLNYLLK----KDELLTS 101
L3.12c   (AY462632)    AIKEHFGNQEAFKKEGGNSASGMLGKE---------FESPDFVLQTLLK----KDLLLTS 101
L4.03c   (AY462637)    AIKKHFKTQDGFSNQGGMDSNVFLDRA---------FRSPDVVLEGVLK----KEVLLTS 99
L3.12b   (AY462629)    AIKQHFKKQKGFDSEGHNDIHSVLN----------LHMTPDFVLEGVLN----KDLLLKS 95
L3.19a   (AY462628)    AIKQHFRKQK------NIVIEDVF-----------MKLTHDDVLDSVLK----KDLLLKS 88
L4.03a   (AY462627)    PIKEHFRKQE------GIVLEQGP-----------IKLTHDAVLQTLLK----KDLLLKI 94
L3.08a   (AY462631)    QIKTHFYKQDG-----FDDFG-------------------HDFALNFLLK----KEELLEN 92
L7.08a   (AY462630)    AIKKHFYKQDI-----RGTVGNGNQGHNGGSGMLGTGLNHDFVLNYLLK----KDELLSS 106
L2.12a   (AY462636)    QIKDHFRKQK--------DMKDENR----------NDIDPGIILEFLLK----KDELLKS 91
L3.08b   (AY462638)    PIKQHFYKQD--------DIVKEVR---------LFKLTHDYVLEGVLK----LDVLLTS 90
L3.12d   (AY462643)    KIIDHFYKQ------------ENIQA---------GMHD--ITLAALLD----KDLLLEI 86
L7.08d   (AY462645)    NIKNHFNTQG-----------DIVQE---------TGCDPGVTLAALLE----EDELLKI 88
L6.11a   (AY462641)    NIKIHFGKQED--------MKEEIK----------GMDPG-IILEGVLN----IEDLFQN 87
L3.12e   (AY462642)    KIKEHFKTQKG--------FGEDVG----------QELPHYMILDGVLK----LEFSKEN 91
L7.08b   (AY462634)    NIKIHFAKQDF--------GKEGVFL---------GVFGSGYILEGVLE----KEELLKI 90

L4.03b   (AY462639)    LQEA------YGNAKEKEHIKKLLQETG----------VVG---GGE--HKTTIDKLL- 131
L7.08f   (AY462640)    IQDT------YGNSQETEHIKQLLNEEKKNQ-----VEAAD---GNDSQKKTTMDKLL- 136
L1.12a   (AY462635)    IKEG------YGNEKDIERIEALLKEEEDKN---EEEDEEA---GADNENKTTIDKLL- 131
L3.08c   (AY462644)    SAQDKQNSLNAEEAEELKRIRDIIEREENQDAAIAAGGSGI---GGANGKKTTIDKLL- 158
L7.08c   (AY462633)    LREG------YGKPEDIEHIRKMLDDEEE------ADG-GV---VGEN--KTTMDKLL- 141
L3.12c   (AY462632)    LQEA------YGNTEDIKHIKEMLDKEEA-------AVLDI---LGGGKDNTTIDKLL- 143
L4.03c   (AY462637)    LREG------YGNAEDIKHIEALLKEEEN----------V---VAVTENKNTIDKLL- 137
L3.12b   (AY462629)    LQEA------YGNAKDIKHIEELLEKEKKRE---EEEAEAG---VVGGKDNTTIDKLL- 141
L3.19a   (AY462628)    LREA------YGNEKDIDRIEKMLEQ-------------AG---VVGGEDNTTIDKLL- 125
L4.03a   (AY462627)    IQDV------HGDTDDIERIEALLDD----D---AAAVAAA---IASGEDNTTMDKLL- 136
L3.08a   (AY462631)    LQEA------YGKPEDIEHIKKLLNDEA------AAGALVV---DSGGENNTTMDKLL- 135
L7.08a   (AY462630)    IKEA------YGDTDDIKRIEELLQETG-------VGG---V---ASGGKDNTTMDKLL- 146
L2.12a   (AY462636)    LREA------YGNAKDIERIEALLNEDE-TK---SQAEDAG---ASGGKENTTMDKLL- 136
L3.08b   (AY462638)    IKGG------YGKPEDIKRIEALLKETG----------------VGGGKDNTTIDKLL- 126
L3.12d   (AY462643)    IEGTYG------NAEDIKHIKDLLDEE--ETA--VAAAIAV----GEN--NTTIDKLL- 128
L7.08d   (AY462645)    IEGTYG------KSKETEHIREMLQETGVANG--VASASGVSGTCGANGKNSTIDKLL- 138
L6.11a   (AY462641)    IKDTYG------DVKEIDHIKKLLKEE-------EAVAVLL----GGGENNTTIDKLL- 129
L3.12e   (AY462642)    TEEDKEN---NVSAKEIDLINKMLEED-------ETAAADV----ADNENNTTIDKLL- 136
L7.08b   (AY462634)    IEGTYG------KSKETEHIKKLLEEE-------TTVD-------ADNENNTTIDKLL- 128
```

DBLβ

```
L1.02b    (AY462659)    RRQHMCTSNLEYLQTKDG--PLNKRDGKLVNNSFLGDVLLSAKMDAEKIKELYKGQNNIK 58
MAL6P1.4  (AY462658)    RREHMCTSNLENLDVDSV--TENDK----ASHSLLGDVQLAAKTDAAEIIKRYKDQNNIQ 54
L1.11b    (AY462655)    -RQHMCTSNLENLDVG-S--VTKG--GKAIH-SLLGDVLLSAKMDADEIIKRYKHQNNIT 53
L1.11c    (AY462656)    --QHMCTSNLEHLETGQS--PLKNSDGKVVNNSFLGDVLLSAKMDAAEIINRYKNQNSIG 56
L2.06a    (AY462657)    -RQHMCTSNLEKLDVE-S--VTKN--DKASH-SLLGDVLLTAKMDADEIINRYKKQNNIE 53
L2.06b    (AY462660)    -RQHMCTSNLEKIDVKS---VTGNSN---VNDSFLGDILLAAKYEAENIKKLYVENNDRK 53
MAL6P1.316(AY462661)    -RQHMCTSNLEYLETDQG--PLKNSDGKFVNHSFLGDVLLAANHEAKKIKELYTKDNG-- 55
L5.06a    (AY462654)    RREHMCTSNLEYLINGNHQAILNVEKGKIN-HSFLGDVLLAAKKEAEFIKS---KVTNN- 55
L7.02a    (AY462653)    RREHMCTSNLEYLLKARGGQFGQVESGKCN-HSFLGDVLLAAKMEAEDIKS---RLNNN- 55
PF11_0008(AY462650)    --EHMCTSNLEYLLHVNKGPLLKVEPDKIN-HSFLGDVLLAAKYEAEFIKTNYTRLNGQ- 56
PF13_0003(AY462651)    -REHMCTSNLEYLLKGNSDQIMKVGNNKIN-HSFLGEVLLAAKYEAEFIKTNYTRLNGQ- 57
L1.02a    (AY462652)    RRQHMCTSNLEYLLHVNKGPLLKVEPDKIN-HSFLGDVLLASKFEGEYIKTNYKRLNGQ- 58
S3.07a    (AY462663)    RRQHMCTSNLENLDLSKEG---LSNSSIAS-NSLLGDVLLAAKFEADFIKSNYNKQKNPK 56
L1.08a    (AY462648)    RRQHMCTSNLEYLINGNHQAILNVENGKIN-HSFLGDVLLAAKYQAEDTIKDYQPKNDDQ 59
PF11_0521(AY462649)    -RQHMCTSNLEYLINGGHQAILNVKNGKIN-HSFLGDVLLAAKYQAQHTMKDYKSKNDKE 58
L7.12d    (AY462664)    RREHMCTSNLEHLNTGTK----GLSDGTLASHSLLGDVLLAAKYEAKNIKELY-EKNKDQ 55
S1.07a    (AY462665)    RRQHMCTSNLEHLNTGNK----GLSDGTLASHSLLGDVLLAAKYEANYIKQRYNDRTKAH 56
L7.12a    (AY462647)    RRQHMCTSNLEKLN------VDKVITNGKVNNSFFVEVLLAANKEAERTKNHYKKP---- 50
L7.12b    (AY462646)    ----MCTSNLEKLN------VDKAITNGKVNNSFFVEVLLAANKEAERTKNHYKKP---- 46
L8.13a    (AY462662)    RREHMCTSNLEKIHDK----FVTQNTNDHVNDTFLVDVLLAAKEEAEDIKKKYKEIKDKN 56
L1.11a                  ----------------------TDGNPKLVNNSFLGDVLLSAKYEAQRTKEDYEPVS--- 35

L1.02b    (AY462659)    NVTDP---KHQETICRALRYSFADLADIIRGRDMWDKDKGSTDMEKYFVPIFKEIREQLP 115
MAL6P1.4  (AY462658)    -LTDPIQQKDQEAMCRAIRYSFADLGDIIRGRDMWNKDSGSTEMEKHLISIFEKINEKLP 113
L1.11b    (AY462655)    GNIKQ---KDKEAMCRAVRYSFADLGDIIRGRDLWDLDDGSKKMEGHLKKIFGKIKQELP 110
L1.11c    (AY462656)    DSIDA---KHKESICRAVRYSFADLGDIIRGRDMWNKDSGSKDMEKYLVNIFDKIKKHP- 112
L2.06a    (AY462657)    DVTDP---NDQATVCRALRYSFADLGDIIRGRDLWE-NGEAKQLQKDLVTIFRHIHSSL- 108
L2.06b    (AY462660)    DLNDE---NDKATVCRAMKYSFADIGDIIRGRDIWDREPGMKHAKKHLKDVFDNIRKSLN 110
MAL6P1.316(AY462661)    -LNDL---KDKETVCRAMKYSFADLGDIIRGRDMWDNGTGMKHAKKHLKDVFDNIRKSLK 111
L5.06a    (AY462654)    --------DNGSAICRAMKYSFADIGDIIRGRDLWEHG-DQTKLQGHLQIIFGKIKENLD 106
L7.02a    (AY462653)    --------GNSSSICRAIKYSFADLGDIIRGKDLWEHK-DFKKLEEHLVKIFEKIKTELK 106
PF11_0008(AY462650)    --------NDNGAKCRAMKYSFADIGDIIRGKDLWGIQ-DFKDLQTKLVTIFGKIKEEIP 107
PF13_0003(AY462651)    --------NDNGAKCRAMKYSFADIGDIVRGRDLWEHN-DFKKLERDLVKIFGKIKEGIT 108
L1.02a    (AY462652)    --------NNNEDKCRAMKYSFADIGDIIRGRDLWERNGDMVKLQGHLETVFENIRKSLK 110
S3.07a    (AY462663)    AATDL---KDEEGICRAMKYSFADIGDITRGKDFWEKNGDAKRLQGHLKEIFGKIKEKLH 113
L1.08a    (AY462648)    G------------ICRAVRYSFADIGDIIKGTDLWDRDRGENKTQRNLETIFGKIKEGIT 107
PF11_0521(AY462649)    G------------ICRAIRYSFADIGDIIKGTDLWDKDGGEIKTQNHLVTIFDKIKAQLP 106
L7.12d    (AY462664)    --------SGHEVICRAVRYSFADIGDIIKGTDLWIDDVGEKKTQGNLVKIFEKIKEKLD 107
S1.07a    (AY462665)    GF------KDEETICRAIKYSFADIGDIIKGTDLWDQNNGEQTTQRKLREVFDKIKQKLP 110
L7.12a    (AY462647)    --------DEHPTACRAIRYSFADIGDIIRGRDLWERNRDMVKLETNLKKIFKNIKDKLP 102
L7.12b    (AY462646)    --------DEHPTACRAIRYSFADIGDIIRGRDLWERNRDMVKLETNLKKIFKNIKDKLP 98
L8.13a    (AY462662)    GLK-----DDQVTTCRAIKSSFADLGDIIRGRDLWDRDNGSTEMEKHLITIFGKIK-ELK 110
L1.11a                  ---------DEQSICRAVRYSFADLADIIRGRDMWDKDDGAQKMERILKSIFKNIYETIG 86

L1.02b    (AY462659)    EK-EQKKYS----NDGAYLDLRRDWWEANRHQVWRAMKC------------AIEKDNITK 158
MAL6P1.4  (AY462658)    EK-EQKKYS----NDGKYLDLRKDWWEANRYKVWKAMKC------------ATKNSKIP- 155
L1.11b    (AY462655)    QQ-IKEKYK---KDEDPYKQLREDWWEANRRQVWKAIKC------------ALKNGSFP- 153
L1.11c    (AY462656)    -E-IQGKYN---TDGPPYKKLREDWWEANRHQVWKAIKC------------ALKDGSFP- 154
L2.06a    (AY462657)    -N-GKGKYEGDYKKKPAYKQLREDWWEANRHQVWEAMKCHIG---------HLKDLSIDK 157
L2.06b    (AY462660)    AK-GISKY---DKDSPDYIKLREDWWEANRHQVWRAMKC------------ATKDINNNK 154
MAL6P1.316(AY462661)    NK-GNQKYNYDDKKLPPYKELREDWWEANRHQVWRAMKC------------AIKEATIDN 158
L5.06a    (AY462654)    DT-IKDKYAS--EDK-PYTQLRKDWWEANRKQIWEAMQCP----------KNGVHITCD- 151
L7.02a    (AY462653)    GT-FGKKYDS--DSNGKHTQLRADWWEANRAKVWEAMKCP----------KNG--IKCD- 150
PF11_0008(AY462650)    D--IKKKYSS--ENP-PYTTLREHWWEANRAKVWEAMQCP-T-----IPPVT---TSCDT 153
PF13_0003(AY462651)    DETTKKQYEK--DDT-DNKQLRCDWWEANRDQVWEAMQCKTT-----IPPVT---TSCDT 157
L1.02a    (AY462652)    GK-GNDKYND--DAP-KYLKLREDWWEANRAKVWEAMQCP-T-----IPPSRGGDIKCAE 160
S3.07a    (AY462663)    DESMKKIYEE--DKD-LYTKLREDWWEANRDQIWEAMQCP-T---SPSSPPRGNNTIC-- 164
L1.08a    (AY462648)    DETIRKKYD----SDPKHTKLRFDWWEANRDEIWKAMKCPTK----------PPVTTNCDT 154
PF11_0521(AY462649)    KD-IKGKYT----G-TKHLELRKDWWEANRDQVWKAMQCGND----------NPCSGESDH 151
L7.12d    (AY462664)    ET-SKNKYK---KSDDKHLDLRKDWWEANRDQIWEAMTCDLK-----SG-----SFPCS- 152
S1.07a    (AY462665)    QD-IRGKYPG--NGDPYHLKLRSDWWEANRHQVWKAMKCHIKDLKDKSGPQSTPSSYCGY 167
L7.12a    (AY462647)    GD-IKEKYKD--DDKGKYRTLREDWWEANRDQVWKAMQCGN-------------DNPCSG 146
L7.12b    (AY462646)    GD-IKEKYKD--DDKGKYRTLREDWWEANRKQIWEAMKCKTN----------GVDITCDS 145
L8.13a    (AY462662)    GE-DKYTNDT--DANPRYKKLRADWWTANRRQVWKAMICETP----------SGKNPCSG 157
L1.11a                  DK--KGKYTN---TDGKYLELREDWWDANRAKVWEAMKCAIKG-----LNVTSSDGKLSD 136

L1.02b    (AY462659)    -------CNGIPIEDYIPQRLRWMTEWAE- 180
MAL6P1.4  (AY462658)    -------CSGIPIEDYIPQRLRWMTEWAE- 177
L1.11b    (AY462655)    -----CKSDHTPLHDYIPQRLRWMTEWAE- 177
L1.11c    (AY462656)    -----CKSDHTPLHDYIPQRLRWMTEWAE- 178
L2.06a    (AY462657)    SKGHCGYSDHTPLDDYIPQRLRWMTEWAE- 186
L2.06b    (AY462660)    -------CNGIPIEDYIPQRLRWMTEWAE- 176
MAL6P1.316(AY462661)    -------CNGIPIEDYIPQRLRWMTEWAE- 180
L5.06a    (AY462654)    -------SG-VPVDDYIPQRLRWMTEWAE- 173
```

```
L7.02a    (AY462653)    -------NGETPLDDYVPQRLRWMTEWAE- 172
PF11_0008(AY462650)    -------TTVTPLVDYIPQRLRWMTEWAE- 175
PF13_0003(AY462651)    -------TTVTPLVDYIPQRLRWMTEWAE- 179
L1.02a    (AY462652)    -------SDV-PFDDYIPQRLRWMTEWAE- 181
S3.07a    (AY462663)    -------SDTTPYDDYVPQRLRWMTEWAE- 186
L1.08a    (AY462648)    -------TTVTPLVDYIPQRLRWMTEWAE- 176
PF11_0521(AY462649)    ----------TPLHDYIPQRLRWMTEWAE- 170
L7.12d    (AY462664)    -------DKTTPHEDYIPQRLRWMTEWAE- 174
S1.07a    (AY462665)    -------SDHTPLDDYIPQKLRWMTEWAE- 189
L7.12a    (AY462647)    --------TDVPLEDYIPQRLRWMTEWAE- 167
L7.12b    (AY462646)    --------DHTPLDDYIPQRLRWMTEWAE- 166
L8.13a    (AY462662)    --------TDVPLDDYIPQRLRWMTEWAE- 178
L1.11a                  HCG---YSDHTPLDDYIPQKLRWMTEWAE- 162
```

NTS-DBL1α

```
L2.12b/c                ------------------------------------------------------------
L2.12a/s (AY462669)     -MVTL--GGGGSSD----ADKYKNAQDAKHLFDIIGKDVYDEVH-KKNAD-YRGELQGRLS 52
L3.08a/c (AY462667)     -MSTL--GGG-------------TDKSAKHVLDEFGQQVYEQVK-NGEAKTYFDELHGDLS 44
L3.08l/s (AY462666)     -MSTL--GGG-------------TDKSAKHVLDEFGQQVYEQVK-NGEAKTYFDELHGDLS 44
L1.12c/c (AY462697)     -MASSTTY--GSV--------------KD-LFENIGKEVQELAK-NDAKQ-YRSQLKGDLS 41
L1.11b/s                ------------------------------------------------------------
L3.12c/s (AY462698)     -MEPKADA---SAPVT----ASTTYSSVKDLLEDIGKSIQKQAK-DAAEERSKEKLQGDVL 52
L2.06a/s (AY462668)     -MAPKAAV---S------APRYDQATSAKDFLDQIGQIVHTKVH-GAAKK-YYDYLHGDLP 49
L3.08c/s                ----------------------TDYSDAKDFLDKIGEDIYKVAK-NDADD-FREKLKGTLS 38
L2.08a/c (AY462673)     -MARGRRGGGSSKEKEDEPD-YTNVKDAKELLDKIGQTVH-KKV-HREDANYRGKLYGLLT 57
L3.06a/s (AY462674)     -MAAGRRGGGGTQEQK---DKYKYVNDAKDLLDQIGEKIQ-DIA-HKAALKYENELHGDLS 55
L1.12b/s (AY462670)     -MAR--PGSAGTQEQK---DKYKYVSDAKDLLDQIGEDIY-KKA-KNDANDFRDKLKGTLS 53
L8.08a/s                ------------------------------------------------------------
L7.08d/s (AY462672)     -MVK-QNGGGG--DGRDG---IDHAS-AKHLLDSIGEKVYKEKV-ESDRKTYFDDLHGTLS 52
L1.12a/s (AY462671)     -MVI--QITGG--EGR-----IEDAT-AKHLLDSIGKKVY-DKV-HSEVDAYRDKLKGTLS 48
L3.08J/s (AY462675)     -MAPQKPTAPDYNNVTN-----------AKDLFDEIGKHIN-PQV-NKEALTYQPELLGYLT 48
L1.11a/s (AY462700)     -MAPQKPTAPDYNNVNN-----------AKDLFDEVGK--YIEKKVRDAALERKGNLKGNLK 48
L7.12a/s                ---------------------------------NFTK--YIKRKASNNAKRHENVLKANLR 26
L3.08a/s (AY462701)     -MGPQR---------------ATQDEDAKHMFDRIGGTVQ-QQV-HTAADQYREKLKGHLS 43
MAL6P1.4 (AY462702)     -MGPP---GGSGGS--------TLDESVKDLFDRIGKEVH-DQV-EKEANQYKEALKGDLS 47
S1.07a/s (AY462695)     -MVKQVKPDGVIE-----------DTTAKHIFDRIGKIVYETV--EKDALPYENELHGLLT 47
S1.08c/s (AY462694)     -MVTQVNPGGNVEDATA-----------KNIFDRIGKKVYEEK-VKKVAEQYRSQLHGSLK 48
L1.12c/s (AY462696)     -MVKQVIVGGVIEDTPA-----------KNIFDRIGKIVYDK---HNEVDEYREKLKGTLS 47
L1.12d/s                ---------------------------------MIGEDVYDEV--HRKAVDYRNYLQGHLE 26
L1.11b/c                -MAPGITG---------------TKKTAKEVLDEIGETIQKGVH-RSAVDYIN-DLQGDLS 43
PF08_0140               -MVPPVRSPRAAGPAPAPASASTTYGSVKDLLEDIGKSIQKQAK-DAAEKRSNGELQGDLS 59
L8.19a/c (AY462685)     -MAPQS-GGGSPQD-----------ESAKHALDRIGEEVY-KEVKSEAEQRSKGELKAGVL 47
S1.08e/s                ------------------------------------------------------------
L7.11a/c (AY462686)     -MAP---GGRQGDGG-----EDIDHQSAKHLLDSIGKIVH-DQVKGEAEKRSNGDLKGSLT 51
L1.02b/c (AY462688)     -MAP---GGPQGGGTKDEDAKNMFDRIGKKVLDK---------VKEEAQT-YKDDLKGNLA 47
L7.11a/s (AY462693)     -MGPPAQGGTD--------------KGAKEVLDEFGQQVYNEKVKNEAET-YKEALTGQLS 45
L1.02c/c (AY462687)     -MARPGSGGGGSSQ------------DAKHVLDEIGQQVH-DQVEKEAKERSNGDLKGNLT 47
MAL7P1.50(AY462690)     -MASQ-SGGGSPQ-------------DAKHVLDEFGQQVH-KEVKKEAERRSKGELKGLLT 45
L1.02a/c (AY462689)     -MAPQGVSGGTQ------------DEDAKHVLDEFGEKVY-KEVKNESNG-FKDDLKGNLN 46
L3.12a/c (AY462681)     -MSTPGGPRGGGGSSGEDGIEHDK--DAKHLLDSIGKIVH-DEVKNGEAKTFKGELEGKLS 57
PFL0935c (AY462680)     ------------------------------------------------------------
L7.12b/c (AY462679)     -MVTPGGGQVGAAGSSGDAEKYKNATNVKDLLDMIGKDVHDKVK-GEAEQRSNGELKGLLS 59
L1.11a/c                -MAPQ--SGGTQ-------------KDAKNMFDRIGKRVHDEIV-EKEAKNYIDDLKGDLN 47
L3.08d/s                ---------------------------MFDRIGKIVHDKVKGEAETYF--DELKGDLQ 29
L3.08i/s                ------------------------------------------------------------
L3.08k/s (AY462682)     -MAASTTYSSA--------------TDAKHLLDMIGKDVHDQVKKEADAKNYIDDLKGNLQ 46
L2.06b/s                -MVT-GSGGSTQD------------EDAKHVLDKIGQQVYKEVK-NGGAEKYIEALKGNLQ 46
L1.12b/c (AY462676)     -MTP-SSAGTNGY------------SDAKDFLDKIGQQVYDKVK--EEAATYKDDLKGNLA 45
L1.12a/c (AY462677)     -MAPQSSGGGRGGGEED-----------AKDFLDKIGEKVYKEVK-GEAEKRSNGDLKAYVS 49
S1.07b/s (AY462678)     -MAPGNAGGGAGSVDKDGIEDTTA----KHLLDSIGKKVHDQVK-KEVEQRSNGDLKGFLT 55
L3.12a/s                -MAPIGDGG------------ADANKSAKEVLDEFGQKVHDEVK-NDA-KTYEGELKGFLS 46
L6.02a/s                ------------------------------------------------------------
L7.08c/s (AY462684)     -MAPQGSTGGGTQEDPIDET------SAKHVLDSIGKIVHDQVK-S-KSNGFKDELKGDLN 52
L7.08b/s (AY462692)     -MVAAAKGGGSSQD-------------AKEVLDRIGEKVHEEVK-NGDAKKYIEALRGNL- 45
L3.08h/s                ------------------------------------------------------GNLT 4
L7.08a/s (AY462691)     -MVTQGRQGGKDEEDPIEHN-----KDAKHLFDSIGEKVYREKVQS-DAKTYEGELKGNLS 54
L2.08a/s (AY462699)     -MAPGSTGTQDD-------------DAKNMFDRIGKEVH-DKVKN-DAKTYEGELKGNLA 44
S1.08b/s (AY462683)     -MGGGSGGGGSSQ---------EQDESVKHMFDRIGQQVYEQVKKDADAKNYIEKLKGDLN 51

L2.12b/c                ---IKIPQEPN---LLVP--CKLKCGYHTNVTKGH---GREYPCNDRWD-----IRFSDKY 45
L2.12a/s (AY462669)     SATYPGDKDSSGTKPSSP--CKLKCGYHTNVTKGH---GREYPCNDRWD-----IRFSDKY 103
L3.08a/c (AY462667)     EATYPGDENPNKTTPPNP--CLLQYDYNSNVTIGG---GREYPCKDRPE-----VRFSDEY 95
L3.08l/s (AY462666)     EATYPGDENPNKTTPPNP--CLLQYDYNSNVTIGG---GREYPCKDRPE-----VRFSDEY 95
L1.12c/c (AY462697)     QATYSRNSNGQ-ETPSDP--CELNHEYHTTVTGGF---DKNNPCKNRPN-----VRFSDIY 91
L1.11b/s                ----------------------LFMNMTLTSTGGF---DTNNPCANRLD-----VRFSDKY 31
L3.12c/s (AY462698)     RATIREGRMIQ-SGIAEL--CHLDYQWHTNVTSG-----KSNPCEGRAD-----VRFSDVI 100
L2.06a/s (AY462668)     RATYPKDENPEGSTENNP--CKLDYKYHTNVTIGG---DKEYPCKDRPD-----VRFSDTE 100
L3.08c/s                QATYPRDKYPKGTTPPNP--CELLYEYHTNVTKGG---DKEYPCANRSD-----IRFSDTK 89
L2.08a/c (AY462673)     QAIFSDI---TRVPTENP--CELDHRYHTNVSWG-----VINPCEHKSV-----ERFSEVS 103
L3.06a/s (AY462674)     KARFSDG---SVVKSNDP--CQLNYEFDTYVTST-----VIEPCEHKKG-----KRLSEVH 101
L1.12b/s (AY462670)     QATYPRDKYPEGTTPEDP--CDLDYNFHTNVTST-----VIDPCKHNSE-----ERFSDTQ 102
L8.08a/s                ----------------------CELDYQYHTNVTSN-----VIEPCKHNSE-----KRFSDTK 31
L7.08d/s (AY462672)     EATFSNG---DKVTNHDP--CNLNYKYDTTVTST-----EIEPCKHKSG-----KRLSEVH 98
L1.12a/s (AY462671)     KATYPKDKYPEGTTPKDP--CELEYKYHTNVTST-----EIDPCEHKKG-----KRLSEVH 97
L3.08J/s (AY462675)     QAIFS---DRSRVPTGNP--CQLDHTVHTNVTS-----NVIDPCKHNSE-----KRFSDTQ 94
L1.11a/s (AY462700)     SAKYR---EGYNIEHANTNICHLIHTHDTNVTEG---HGKEYPCANRSD-----IRFSDKQ 98
L7.12a/s                QAKFRHEFSAYRPNYGNP--CELDYRFHTNVWNRGAS--ERDPCYRRQP-----KNNSKLE 78
L3.08a/s (AY462701)     QATFREG---RMIESEKAELCKLNYKYHTNVTKG---RGREDPCLGRYP-----ERFFDTQ 93
MAL6P1.4 (AY462702)     SATFPTG---RRHEKPQSESCKLNYIYDTNVTSG---GGKENPCYGRQG-----VRFSDTK 97
```

```
S1.07a/s (AY462695)    NATFEKNPP-GKQTPARP--CKLNHEYHTNATN-----GRSYPCRKGTE-----KRFSEVS 95
S1.08c/s (AY462694)    NAIFEDAPD-GQQTPAGP--CQLKYKWHTNATN-----GKSYPCRTGTE-----ERFSQVH 96
L1.12c/s (AY462696)    QATFEGKP----IKVSVP--CGLEYQYHTNVTKG---HGREHPCRKGTE-----KRFSEVH 94
L1.12d/s               NTIFEKLPS-DKQALSDP--CDLNHEYHTTATEG---HNKENPCEKRSN-----VRFSDTK 76
L1.11b/c               KVHFKDG----EIDTHDI--CNLDHTKHTNV-----SWGVIHPCDNRLA-----NLFSEES 88
PF08_0140              KARFKDQ----KNTTTNP--CQLNYRNHSNVRS---SFENDNPCYGREE-----KRFSDSR 106
L8.19a/c (AY462685)    FATFSGEE---LAAFSDP--CGLIKNEGENLIR-----ARGHPCKKDTNG-NDVDRFSVKE 97
S1.08e/s               ------------------------------------------------------------
L7.11a/c (AY462686)    SVTLLGESVG----FSDP--CKLINDKVVKLID-----ARGDPCKKDTNG-NDVDRFSDKQ 100
L1.02b/c (AY462688)    SSSILGE----LAAFSDP--CNLIKNEGEKLLA-----LRDNPCGNGSGKGEDVSRFSKER 97
L7.11a/s (AY462693)    QVSINSETIT----FTDP--CTSEYDKHTTSAN-----GNTKPCGND-GKEDDSKRFSKNR 94
L1.02c/c (AY462687)    ISTIFDTETTG---TDDP--CSSDYTTRFDARGD--------PCKKDGTG-NDVERFSVKQ 94
MAL7P1.50(AY462690)    SAKLSGGEIAGT--TD-P--CSSDYTKHFE---ANSNR---YPC-GNT----NVDRFPDND 90
L1.02a/c (AY462689)    TASGSSGETLGSIDT-----CYLVNQYYTERLNGNSNR---YPC-GTRT--EEVKRFSDKQ 96
L3.12a/c (AY462681)    QAS-ILGESAGT---DDP--CTFK---YDELTGAARGE--RYPCKELSGKM-FENPFSDTL 106
PFL0935c (AY462680)    -----MGERASSNKT-----CTLVKEYYEHFNGDANSN--RYPCKELNGKM-GENRFSDTL 48
L7.12b/c (AY462679)    QAS-ILGETAFT---DKP--CNFDYDKLINGSGSGGNR---HPCKNLKG-STIEGRFSNTL 110
L1.11a/c               KANNSSDETIS---SLDP--CELQSEYTELINGSGSGVAARGDPCENLS-RKVEPRFSDTL 102
L3.08d/s               KAVTTIWQTV---DTDKP--CDLVEQYRSKANGGGGKGER—YPCKELGV-KVEVKRFSNSV 83
L3.08i/s               ------------------------------------------------------------
L3.08k/s (AY462682)    HAAKGSDETVG---TDDP--CTFKYDKLIGANDGNR-----HPCTNLKGN-ANEERFSDTL 96
L2.06b/s               -EAKGIGE---LASSLDT--CTLIKDKRHE-LLRARD----DPCTNLSG--KLEPRFSDTL 94
L1.12b/c (AY462676)    SSSILGVESASTA---DP--CNFEYHKLIG--ANDGNR--HPCESLSGKDAKKEERFSNTL 97
L1.12a/c (AY462677)    FASIFGEEKART---LDP--CNLKSEYTKLIEANRNR----YPCTNLSG--KVEPHFSNTL 99
S1.07b/s (AY462678)    STTIFGGERAIT---LDP--CNLKSEYTKLIEANSNR----YPCKNLKG-ITNEERFSNTL 106
L3.12a/s               STTIFGGVRDRT---LEP--CKLKSEYTKLIKDNSNR----YPCGNES---VSQKRFSVKQ 95
L6.02a/s               ------------------------EYYEHFNGDANSNR----HPCGN-RTGKEEVNRFSDTL 35
L7.08c/s (AY462684)    TANNSSGE---LASSIDP--CTLVEQYRSKANGGDTDKR--HPCRKDGTGKEEVNRFSDTL 106
L7.08b/s (AY462692)    TDSTILGET—VAFSDT----CDLVKQYISKVGGNSER----HPCGNGTGDAKKEERFSNTL 97
L3.08h/s               SSTFFGGETVSS---LHP--CGLDYTKRLK---GK-----RYPCANRQT-----VRFSDEY 47
L7.08a/s (AY462691)    FASIFDTETTG---TDDP--CGLDYIKRLN---GNNN---RHPCANRSP-----VRFSDEY 99
L2.08a/s (AY462699)    SSSIWK-ESAY---TTDT--CQLVYDYYTKRLNGK-----RYPCANRSP-----VRFSDES 89
S1.08b/s (AY462683)    TANGHSSETL---GTTDT--CTLVYKYYDDVNGGGAAPGERYPCGTGKE-----ERFSDTL 102

L2.12b/c               GGQCTNVKIHGNSKGGNG--TEVGACAPFRRLHLCDHHLSHMQADK---IMNKHNLLLEV 100
L2.12a/s (AY462669)    GGQCTNEKIHGNSKGENG--TEVGACAPFRRLHLCDHRLSHMQADK---IMNKHNLLLEV 158
L3.08a/c (AY462667)    GGQCTDSKIKGN-EDNKG-----GACAPFRRLFLCDQHLSHMKAEK---INNKHNLLLEV 146
L3.08l/s (AY462666)    GGQCTDSKIKGN-EDNKG-----GACAPFRRLFLCDQHLSHMKAEK---INNKHNLLLEV 146
L1.12c/c (AY462697)    GGQCTDSKIRGN-DT-----NNGGACAPFRRLFLCDHHLSHMQADQ---IDSKDNLLLEV 142
L1.11b/s               GGQCTDTKIHGN-EN-----NEFGACAPFRRLFLCDHHLSHMKVEK---INTKDNLLLEV 82
L3.12c/s (AY462698)    GGQCTNSKIKGN-EVKYG--KDIGACAPFRRLFLCDHHLSYMNAGK---TNTTDNLLLEV 154
L2.06a/s (AY462668)    GAQCDKSKIGGS-NS-----NKDGACAPYRRSSLCDHHLSYMNAGK---TNTTDNLLLEV 151
L3.08c/s               GAECDYRKIDGN-KG-----KTGGACAPYRRSSLCDHHLSYMNAGK---TNTTDNLLLEV 140
L2.08a/c (AY462673)    GGECDEKKIKGS---------NGGACAPFRRLHVCDKNLEQIKPHT---ITATHNLLVDV 151
L3.06a/s (AY462674)    SSECDRKKIKDSENN------TAGACAPFRRLHICDENLEQIKPHT---ITATHNLLAEV 152
L1.12b/s (AY462670)    GAECDNNKIRGSDKK-----SNGGACAPFRRLHLCDKNIQQIKTEN---ITT-HNLLLDV 153
L8.08a/s               GAECDKSKIRGSNGK-----SEG-ACAPYRRLHVCDRNLEQIDPAK---ITTTHNLLADV 82
L7.08d/s (AY462672)    RGECDNRKIKDSNVK-------EGACAPYRRLHVCDKNIQQIKAEQ---ITT-HNLLADV 147
L1.12a/s (AY462671)    SSECDNRKIRDCDKKN----NSVGACAPYRRLHVCDKNIQQIKTEN---ITT-HNLLLDV 149
L3.08J/s (AY462675)    GAECDRKKIKDNKG-------KEGACAPYRRLSLCDTNLEQIKTEN---ITT-HNLLVDV 143
L1.11a/s (AY462700)    GAECDKSKIKDGND-------EGGACAPYRRSHLCDQHLSHMKEDK---IDSKDNLLLEV 148
L7.12a/s               GAVCTNSKIKGNENKI----IDIGACAPYRRIHLCDYNLEHIHEGN---VLTTDDLLGNV 131
L3.08a/s (AY462701)    GSECATSKIEGNVGKKTNKGKSEGACAPYRRLHLCDQNLEHIDPDK---IESTHNLLVDV 150
MAL6P1.4 (AY462702)    GAKCYSYKIE-------DNDSSIGFCAPYRRLHLCVQNLEQIKPDQ---ITSTHNLLVDV 147
S1.07a/s (AY462695)    GGECDKNKIRGSKGDNE------GACAPYRRLNLCVRNLENISDFN---NINNDTLLADV 146
S1.08c/s (AY462694)    GGECDDSKIKGNKGSNEN---SEGACAPYRRLNLCVRNLENISALD---KINNDTLLADV 150
L1.12c/s (AY462696)    GGECANSKIKGNKGSKEN---SEGACAPYRRLHLCDYNLENINDYK---NINNDTLLVDV 148
L1.12d/s               GAECDKSKIKGNKDNKDIGGKSEGACAPYRRLHLCDHNLENISDFD---HINNDTLLADV 133
L1.11b/c               ASQCSTSRISGN-------ANNSGSCAPYRKLQLCDYNLERITDTN---TTNTNNLLVDV 138
PF08_0140              SGQCTYNRIKDSKEG----DNKVGACAPYRRLHICDHNLENINDYK---NINNHTLLVDV 159
L8.19a/c (AY462685)    QAEYDNKKMK----CS-N----GDACAPFRRLHLCNKNIQQIKTEN-IT---THNLLVDV 144
S1.08e/s               ----------------------------------------IKLIVKD---------NLLLEV 13
L7.11a/c (AY462686)    GAECNKSKIK----DS-DKKNKGGACAPYRRLHVCDKNMEKIATS--MT---KHDLLLDV 150
L1.02b/c (AY462688)    VDEYDEKKIKDN--KS-KGGNNEGECAPYRRLSLCNKNFPNMNSND-SS-KAKNDLLVDV 152
L7.11a/s (AY462693)    IAEYDEKKIR---------GSNDGACPPYRRLSLCNKNMVNMIPNN-NDGKAKHDLLADV 144
L1.02c/c (AY462687)    GAECGNSKIHGNS-KG-GTGTEVGACAPYRRLNLCNKNLENINKYD-NT-KAKHDLLAEV 150
MAL7P1.50(AY462690)    GAECDNSKIKGN--KG-KEDNSEGACAPYRRLSLCNKNFQNNNNDH-SS-NAKHDLLLDV 145
L1.02a/c (AY462689)    GAQCDKKKIK-------DSDSNGDACAPFRRLNLCNKNMVKMDTNN-NDGKATHTLLAKV 148
L3.12a/c (AY462681)    GGQCTNEKMRS-------GGK--GACAPYRRLHLCHHNLETINN---TTSTTSDTLLAEV 154
PFL0935c (AY462680)    GGQCTNKKIEG-------NKNNCGACAPYRRLHLCHHNLESID----TTSTTSDTLLAEV 97
L7.12b/c (AY462679)    GGQCTDSKMRS-------GGE--GACAPYRRLHLCHHNLESIETNNYDSDNAKHNLLAEV 161
L1.11a/c               GGQCTDSKMRS-------GGK--GACAPYRRLHLCSHNLESID----TTSMTTHKLLAEV 149
L3.08d/s               SGQCTNKKIEGNKYI---EGKDVGACAPYRRLHLCHHNLESIQTNNYNSSNAKHNLLAEV 140
L3.08i/s               ------------------------------LCSHNLETID-TKSTTSDT---LLLEV 23
L3.08k/s (AY462682)    GGQCTDSKIKGNKYNR-KTRKDCGACAPYRRLHLCHHNLETINNTTSTTSDT---LLAEV 152
L2.06b/s               GGQCTKEKIEGNKYIK-G--KDVGACAPYRRLHLCHHNLETINNTTSTTHD----LLAEL 147
L1.12b/c (AY462676)    GGQCTNKKMRSG---------GEGACAPYRRLHLCSHNLESIETNNYDSNNARHKLLAEV 148
```

```
L1.12a/c (AY462677)    GGQCTKEKISG-------STNTCGACAPYRRLHLCHHNLESIPTNNYNSSNARHNLLAEV 152
S1.07b/s (AY462678)    GGQCTKEKISG-------STNTCGACAPYRRLHLCDHNLETIETNNYDSNNARHNLLAEV 159
L3.12a/s               QAEYDNKKMKC-------SNG--GACAPYRRLHLCHHNLESIQTKNYNSSNARHNLLAEV 146
L6.02a/s               GGQCTDHRIKGNN-----RNVTGGACAPLRRLHLCHHNLESIQTNNYDSGKAMHTLLAEV 90
L7.08c/s (AY462684)    GGQCTDHRIKGND-----RNKTGGACAPLRRLHLCDYNLESINTDKIDSGNAKHDLLAEV 161
L7.08b/s (AY462692)    GGQCTDHRIKGND-----RNVTGGACAPFRRLHVCDRNLENIN--DYSKINNKDNLLLEV 150
L3.08h/s               GGQCPHNRIKDNE----TVDNNCGACAPYRRLHLCDYNLEKMGRT----STTKHDLLAEV 99
L7.08a/s (AY462691)    GGQCTHNRIKDNE----TVDNKCGACAPYRRLHLCDYNLEKMGTTK---SKARHNLLAEV 152
L2.08a/s (AY462699)    RSQCTYNRIKDNK----SEDNACGACAPFRRLSVCDYNLEKMGTKK---IDNTHKLLAEV 142
S1.08b/s (AY462683)    GGQCTYNRIKDSN----KNDNK-GACAPYRRLHLCDYNLESINNYN---SNARHKLLAEV 154

L2.12b/c               CYAAKYEGKSLVEKHKKFKEENSDFNIN--ICTVLARSFADI- 140
L2.12a/s (AY462669)    CYAAKYEGKSLVEKHKKFKEENSDFNIN--ICTVLARSFADI- 198
L3.08a/c (AY462667)    CLAAKYEGESLKGYHDKYNATYSDSRSQ--LCTVLARSFADI- 186
L3.08l/s (AY462666)    CLAAKYEGESLKGYHDKYNATYSDSR----------------- 172
L1.12c/c (AY462697)    SLAAQYEGKLLVERHRECKKTHEDFKTN--ICDVLARSFADI- 182
L1.11b/s               LLAAKYEGESIRN---EYDQKKDDYKLG--LCTALARSFADI- 119
L3.12c/s (AY462698)    CLAAQYEGESIIK---NYPQDHNNNEV---ICTALARSFADI- 190
L2.06a/s (AY462668)    CMAAKHEGASITRYHDQHKRTNP--DSK--ICTELARSFADI- 189
L3.08c/s               CLAALHEGQSITGRYPQHQETNEGTASQ--LCTVLARSFADI- 180
L2.08a/c (AY462673)    CMAAQFEGASIKGYYPKYQEKYKDTGS--TMCTMLARSFAD-- 190
L3.06a/s (AY462674)    CMAAQFEGASIKGYHPQYDVQYPGSGS--TMCTVLARSFADIG 193
L1.12b/s (AY462670)    CLAAKFEGQSITGYHPQYEVQYPSSGS--TICTALARSFADI- 193
L8.08a/s               CQAAKFEGQSIAGYYAQYEAQYPGSGS--TICTALARSFADI- 122
L7.08d/s (AY462672)    CMAAKFEGQSISGYHPKYEIQYPGSGSGFTLCTMLARS----- 185
L1.12a/s (AY462671)    CLAAKFEGQSITGYHPQYEVQYPSSGS--TMCTMLARSFADI- 189
L3.08J/s (AY462675)    CMAAKYEGTSLKGYHDQHQLTYPGYHS--QLCTELARSFADI- 183
L1.11a/s (AY462700)    SLAAQYEGQSIRVDHDKYKLDNDNSGS--KLCTELARSFADI- 188
L7.12a/s               LVMAKNEGASIVNSNAHNGVLN--------VCTVLARSFADI- 165
L3.08a/s (AY462701)    CLAAQYEGKSIR---TQYEQKKDDYKS--GLCTVLARSFADI- 187
MAL6P1.4 (AY462702)    LLAAKYEGQSITQDYPKYQATYDDSPS--KMCTMLARSFADI- 187
S1.07a/s (AY462695)    CLAALHEGDSIRSDHYKYKLTN--SSS--QICTMLARSFADIG 185
S1.08c/s (AY462694)    CLAALHEGAAIRGDHGKYQETNNDVNA--NICTMLARSFADIG 191
L1.12c/s (AY462696)    CLAALHEGASLQGYHDKYKETN--DSS--QLCTMLARSFADI- 186
L1.12d/s               CLAAKFEAESLEKYRDQYQLSNRDLHI--NICTVLARSFADI- 173
L1.11b/c               LLAAKYEGDSLSKYIKEHPEIIPNSN----ICTVLARSFADIG 177
PF08_0140              CLSAKHEGEMIANKLKEYDKSNYESR----ICTVLARSFADI- 197
L8.19a/c (AY462685)    CMAANYEAQSLIRDHPQYQEKYGDS----QICTVLARSFADI- 182
S1.08e/s               SLAAKYEGEYLRINHPQYQEKYGDS----QICTVLARSFADI- 51
L7.11a/c (AY462686)    CLAAKYEGDSLKHYSKKLNLTYTDSP--SQLCTELARSFADI- 190
L1.02b/c (AY462688)    CLAAKYEGESITLNYPKYEAIYEGSG--HTPCTMLARSFADI- 192
L7.11a/s (AY462693)    CYAAKHEGESLKNYHPQYKLTYGDS----QICTVLARSFADI- 182
L1.02c/c (AY462687)    CHAAKYEGASITLHYPQYQNKYDDSG--STMCTMLARSFADI- 190
MAL7P1.50(AY462690)    CMAANYEAQSLITYHDKHE--LTNVG--SQICTVLA------- 177
L1.02a/c (AY462689)    CYAAKYEGDSIKTHYPLYQHKYGDSD--SQICTVLARSFADI- 188
L3.12a/c (AY462681)    CMAAYYEGDLIKTHYTQHEQTNPDTK--SQLCTVLARSFADI- 194
PFL0935c (AY462680)    CYAAKFEGETLTTQHGQHQQTNPGTA--SQLCTVLARSFAD-- 136
L7.12b/c (AY462679)    CMAAKYEGDLIKTHHRQHQLTYPDSA--SELCTVLARSFADI- 201
L1.11a/c               CMAAKYEGDSITRYHPQHQETNPGTA--SQLCTVLARSFADI- 189
L3.08d/s               CYAAKYEGDSIKKYHDEHQGTN--KV--SNICTVLARSFADI- 178
L3.08i/s               CMAAKYEGQSITGYYPIYQTKYNDYG--SPICTVLARSFAD-- 62
L3.08k/s (AY462682)    CYAAKEEGASISGRYRQYVTKYKDYG--STMCTVLAR------ 191
L2.06b/s               CMAAKYEAESLEKYRDQYEAQYPGSG--STMCTVLARSFA--- 185
L1.12b/c (AY462676)    CYAAKYEAESLIHDHAQYRVKNTDFN--TTICTVLARSFADI- 188
L1.12a/c (AY462677)    CMAAKYEGDLIKTHYTKYKESNPGTA--SQLCTVLARSFADI- 192
S1.07b/s (AY462678)    CMAAKYEGDLIKTHYTPYQLTNEGTA--SQLCTVLARSFADI- 199
L3.12a/s               CMAAYYEGDLIKTRYTPYQHTNEGTA--SQLCTELARSFADI- 186
L6.02a/s               CMAAKYEGESIKVDYVKYRGNNPDFN--TTICTELARSFADI- 130
L7.08c/s (AY462684)    CMAAKYEGNSINTHYTIHKETNPGTA--SQLCTVLARSFADI- 201
L7.08b/s (AY462692)    CMAAYYEGQSIKDDHAHYQSKNSDFK--TNICTELARSFADI- 190
L3.08h/s               CMAAKYEGNSINTHYTPHQVTYSDS--AAELCTVLARSFADI- 139
L7.08a/s (AY462691)    CLAAKYEGESLKNYHAQYQAKNTDF--KTNICTELARS----- 188
L2.08a/s (AY462699)    CMAAKYEAESLEKYRDQYDAKYHDT--GFTICTALARSFAD-- 181
S1.08b/s (AY462683)    CYAAKHEGDLINTHYTPHQQKYKDTGTASQLCTVLARSFADIG 197
```

upsB-type 5'UTR-ATG

```
L2.06d/s (AY462590)     --TCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 50
L3.08g/s (AY462591)     --TCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATGTA-- 50
L2.08a/s (AY462592)     --TCATATATAATTTTACAAAATATATAAAACATAAGAAA------ATAATATATATA-- 50
S1.12e/s (AY462593)     TCTCATTTATAATTTTACAAAATATATAAAACATAAGAAA------ATAATATATATATA 54
L3.08b/s                --TCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 50
S1.07a/s (AY462594)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 51
S1.08c/s (AY462595)     -CTCATATATAATTTTACAAAATATATAAAACATAAGAAA------ATGATATATATA-- 51
PF08_0106(AY462596)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATAAA-- 51
S1.08g/s (AY462597)     CCTCATTTATAATTTTACAAAATATATAAAACATAA-AAA------ATAATATATGTA-- 51
PF08_0140(AY462598)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATAAA-- 51
S5.12a/s (AY462599)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 51
L3.06a/s (AY462600)     --TCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 50
S1.08d/s (AY462601)     ---------TAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATATA 45
S3.12a/s (AY462602)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATGTA-- 51
L2.06f/s (AY462603)     ------------TTTTACAAAATATATAAAACATAAGAAA------ATAATATATATA-- 40
S1.12c/s (AY462604)     TCTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 52
L7.11b/s (AY462605)     TCTCATATATAATTTTACAAAATATATAAAACATAAAAAA------GTAATATATAAA-- 52
S1.07c/s (AY462606)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATAAA-- 51
L3.08a/s (AY462607)     --TCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 50
S1.07b/s (AY462608)     -CTCATTTATAATTTTACAAAATATATAAAACATAAGAAA------ATAATATATATA-- 51
S3.12d/s (AY462609)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATGTA-- 51
L6.02b/s (AY462610)     TCTCATTTATAATTTTACAAAATATATAAAACATAAGAAA------ATAATATATATA-- 52
MAL7P1.50(AY462611)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAATAAAA-ATAATATATGTA-- 56
L6.02c/s (AY462612)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAAATAAAAATAATATATGTA-- 57
S5.12b/s (AY462613)     -CTCATTTATAATTTTACAAAATATATAAAACATAAGAAAAT------AATATATAAA-- 51
L1.12a/s (AY462614)     TCTCATTTATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATATA-- 52
S1.08b/s (AY462615)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAAAA-----TTATATATGTA-- 52
S3.12c/s (AY462616)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATGTA-- 51
L3.08f/s (AY462617)     --TCATTTATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATGTA-- 50
L3.08l/s (AY462618)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAAA-----TAATATATATATA 54
S3.12e/s (AY462619)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATATATA 53
S1.08f/s (AY462620)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATATA-- 51
S1.12a/s (AY462621)     -CTCATTTATAATTTTACAAAATATATAAAACATAAAAAAA------TAATATATGTA-- 51
L2.06e/s (AY462622)     -CTCATATATAATTTTACAAAATATATAAAACATAAGAAAA------TAATATATATA-- 51
L3.08e/s (AY462623)     --TCATTTATAATTTTACAAAATATATAAAACATAAAAAA------GTAATATATAAA-- 50
L1.11a/s (AY462624)     TCTCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 52
S1.07d/s (AY462625)     -CTCATATATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATGTA-- 51
L7.11a/s (AY462626)     TCTCATTTATAATTTTACAAAATATATAAAACATAAAAAA------ATAATATATATA-- 52
L3.12b/s                ------------------------------------------------------------
MAL6P1.316              -CTCATTTATAATTTTACAAAATATATAAAACATAAGAAA------ATAATATATGTA-- 51

L2.06d/s (AY462590)     --------ATTAAATATTTAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 99
L3.08g/s (AY462591)     --------ATTAAATATTTAAATAAAG-GAATACATAATATATATAATATTTTTCATAAA 101
L2.08a/s (AY462592)     --------ATTAAATATTTAAATAAAA-GAATACATAATATATATAATATTTTTCATAAA 101
S1.12e/s (AY462593)     --------ATTAAATATTTAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 103
L3.08b/s                --------ATTAAATACATAAAGAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 99
S1.07a/s (AY462594)     --------ATTAAATATTTAAATAAAA-GAATACATGATATATA--ATATTTTTCATAAA 100
S1.08c/s (AY462595)     --------ATTAAATATATAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 100
PF08_0106(AY462596)     --------ATTAAATATAAAAATAAAG-GAATACATGATATATA--ATATTTTTCATAAA 100
S1.08g/s (AY462597)     --------ATTAAATATTTAAATAAAG-GAATACATCAAATATA--ATATTTTTCATAAA 100
PF08_0140(AY462598)     --------ATTAAATATAAAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 100
S5.12a/s (AY462599)     --------ATTGAATAAGTAAATAAAGAGAATACATGAAATATA--ATATTTTTCATAAA 101
L3.06a/s (AY462600)     --------ATTAAATATAAAAATAAAG-GAATACATAATATATATAATATTTTTCATAAA 101
S1.08d/s (AY462601)     --------ATTAAATAGATAAATATTT-GAATACATGAAATATA--ATATTTTTCATAAA 94
S3.12a/s (AY462602)     --------ATTAAATATTTAAATAAAG-GAATACATGATATATA--ATATTTTTCATAAA 100
L2.06f/s (AY462603)     --------ATTAAATATAAAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 89
S1.12c/s (AY462604)     --------ATTAAATATATAAATAAAG-GAATACATGAAATATA--ATTTTTTTCATAAA 101
L7.11b/s (AY462605)     --------ATTAAATATAAAAATAAAG-GAATACATGAAATGTA--ATATTTTTCATAAA 101
S1.07c/s (AY462606)     --------ATTAAATATAAAAATAAAG-GAATACATCAAATATA--TTATTCTTTATTA- 99
L3.08a/s (AY462607)     --------ATTAAATATATAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 99
S1.07b/s (AY462608)     --------ATTAAATATTTAAATAAAA-GAATACATGATATATA--ATATTTTTCATAAA 100
S3.12d/s (AY462609)     --------ATTAAATATATAAATAAAA-GAATACATGATATATA--ATATTTTTCATAAA 100
L6.02b/s (AY462610)     --------ATTAAATATATAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 101
MAL7P1.50(AY462611)     --------ATTAAATATTTAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 105
L6.02c/s (AY462612)     --------ATTAAATATTTAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 106
S5.12b/s (AY462613)     --------ATTAAATATATAAATAAAA-GAATACATAATATATATAATATTTTTCATAAA 102
L1.12a/s (AY462614)     --------ATTAAATATATAAATAATG-GAATACATGAAATATA--ATATTTTTCATAAA 101
S1.08b/s (AY462615)     --------ATTAAATACATAAAGAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 101
S3.12c/s (AY462616)     --------ATTAAATACATAAATAAAG-GAATACATAATATATATAATATTTTTCATAAA 102
L3.08f/s (AY462617)     --------ATTAAATATATAAATAAAA-GAATACATAATATATATAATATTTTTCATAAA 101
L3.08l/s (AY462618)     TATATATAATTAAATAGATAAATAAAG-GAATGCATAAAATATA--ATATTTTTCATTAT 111
S3.12e/s (AY462619)     TATAT--AATTAAATATATAAATAAAG-GAATACATAATATATATAATATTTTTCATAAA 110
S1.08f/s (AY462620)     --------ATTAAATATATAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 100
S1.12a/s (AY462621)     --------ATTAAATATATAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 100
```

```
L2.06e/s (AY462622)    --------ATTAAATATAAAAATAAAA-GAATACATGAAATATA--ATATTTTTCATAAA 100
L3.08e/s (AY462623)    --------ATTAAATATAAAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 99
L1.11a/s (AY462624)    --------ATTAAATATTTAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 101
S1.07d/s (AY462625)    --------ATTAAATATATAAATAAAA-GTATACATGAAATATA--ATATTTTTCATAAA 100
L7.11a/s (AY462626)    --------ATTAAATATTTAAATAAAG-GAATACATGAAATATA--ATATTTTTCATAAA 101
L3.12b/s               ------------------------------------------------------------
MAL6P1.316             --------ATTAAATATATAAATAAAG-GAATACATAATATATATAAAATTTTTCATTAA 102

L2.06d/s (AY462590)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 147
L3.08g/s (AY462591)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 148
L2.08a/s (AY462592)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 148
S1.12e/s (AY462593)    AT-GTAATT-GTTGTTTTTTTTTTGTT---------AGAATA-TTTAAATTTATTATAA 151
L3.08b/s               AT-GTAATT-GTTGNTTTTTTTTTTTTTTTTTGGTAGAATA-TTTAAATTTATTATAA 156
S1.07a/s (AY462594)    AT-GTAATT-GTTGTTTTTTTTTGTT----------AGAATA-TTTAAATTTATTATAA 147
S1.08c/s (AY462595)    AT-GTAATT-GTTGTTTTTTTTGTT----------AGAATA-TTTAAATTTATTATAA 146
PF08_0106(AY462596)    AT-GTAATT-GTTGTTTTTTTTTTGTT---------AGAATA-TTTAAATTTATTATAA 148
S1.08g/s (AY462597)    AT-GTAATT-GTTGTTTTTTTTTTGTT----------AGAATA-TTTAAATTTATTATAA 147
PF08_0140(AY462598)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 147
S5.12a/s (AY462599)    ATAGTAATTAGTTGTTTTTTTTTTT-GTT--------AGAATAGTTTAAATTTATTATAA 152
L3.06a/s (AY462600)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATT-TAA 147
S1.08d/s (AY462601)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAC 142
S3.12a/s (AY462602)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTACTATAA 148
L2.06f/s (AY462603)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 136
S1.12c/s (AY462604)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 148
L7.11b/s (AY462605)    AT-GTAATTGTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 149
S1.07c/s (AY462606)    -T-ATAATT----------TTTTT--GTT--------AGAATA-TTTAAATTTACTATAA 137
L3.08a/s (AY462607)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------TGAATA-TTTAAATTTACTATAA 146
S1.07b/s (AY462608)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 148
S3.12d/s (AY462609)    AT-GTAATT-GTTGGTTTTTTTTT--GTT--------AGAATA-TTTAAATTTACTATAA 147
L6.02b/s (AY462610)    AT-GTAATT-GTTGTTTTTTTTTTTGTT--------AGAATA-TTTAAATTTATTATAA 150
MAL7P1.50(AY462611)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 152
L6.02c/s (AY462612)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 154
S5.12b/s (AY462613)    AT-GTAATT-GTTGTTTTTTTTT---GTT--------AGAATA-TTTAAATTTATTATAA 148
L1.12a/s (AY462614)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 149
S1.08b/s (AY462615)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 149
S3.12c/s (AY462616)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 149
L3.08f/s (AY462617)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 148
L3.08l/s (AY462618)    AT-GT----------TTTTTTTTTTTGTT--------AGAACA-TTTAAATTTATTATAA 152
S3.12e/s (AY462619)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 158
S1.08f/s (AY462620)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 148
S1.12a/s (AY462621)    AT-GTAATT-GTTGTTTTTTTTTT--GTT--------AGAATA-TTTAAATTTATTATAA 147
L2.06e/s (AY462622)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 148
L3.08e/s (AY462623)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 147
L1.11a/s (AY462624)    AT-GTAATT-GTTGTTTTTTTTTTTTGTT--------AGAATA-TTTAAATTTATTATAA 150
S1.07d/s (AY462625)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 148
L7.11a/s (AY462626)    AT-GTAATT-GTTGTTTTTTTTTTT-GTT--------AGAATA-TTTAAATTTATTATAA 149
L3.12b/s               ------------------------------------------------------------
MAL6P1.316             AT-----TT------TTTTTTTGTTTGTT--------AGAATA-TTTGAATATATTATAA 142

L2.06d/s (AY462590)    AAGTATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 198
L3.08g/s (AY462591)    AAGTATTAATATA-------TATTTTTTT--AAAAATATA----TAAAACTAATAATTAT 195
L2.08a/s (AY462592)    AAGTATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 199
S1.12e/s (AY462593)    AAGTATTAATATA-------AA-TTTTTTT-AAAAATATATATATAAAACTAATAATTAT 202
L3.08b/s               AAGTATTAATATA-------T--TTTTTTT-AAAAATATATATATAAAACTAATAATTAT 206
S1.07a/s (AY462594)    AAGTATTAATATA-------TATTTTTTTT-AAAAATATATATATAAAACTAATAAGTAT 199
S1.08c/s (AY462595)    AAGTATTAATATA-------TATTTTTTTT-AAAAATATATATATAAAACTAATAATTAT 198
PF08_0106(AY462596)    ATGTATTAATATA-------TA-TTTTTTT-AAAAATATATATATAAAACTAATAATTAT 199
S1.08g/s (AY462597)    AAATATTAATATA-------TA-TTTTTTT-AAAAATATATATATAAAACTAATAAGTAT 198
PF08_0140(AY462598)    ATGTATTAATATA-------TATTTTTTTT-AAAAATATATATATAAAACTAATAATTAT 199
S5.12a/s (AY462599)    ATGTATTAATATA-------AATTTTTTTT-TTTTAAAAATATATAAAACTAATAATTAT 204
L3.06a/s (AY462600)    AAGTATTAATATA-------TATTTTTTTA-AAAAATATATATATAAAACTAATAATTAT 199
S1.08d/s (AY462601)    AAGTATTAATATA-------AATTTTTTT-----AAAAATATATAAAACTAATAAGTAT 190
S3.12a/s (AY462602)    AAATATTAATATA-------TATTTTTT---AAAAATATATATATAAAACTAATAATAAT 198
L2.06f/s (AY462603)    AAGTATTAATATA-------AATTTTTTTA-AAAAATATATATATAAAACTAATAATTAT 188
S1.12c/s (AY462604)    AAATATTAATATA-------TATTGTTTT--AAAAATATATATATAAAACTAATAATTAT 199
L7.11b/s (AY462605)    AAATATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 200
S1.07c/s (AY462606)    AAATATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 188
L3.08a/s (AY462607)    AAATATTAATATA-------AATTTTTTTT-AAAAATATATATATAAAACTAATAAGTAT 198
S1.07b/s (AY462608)    AAATATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 199
S3.12d/s (AY462609)    GAATATTAATATA-------TATTTTTTTT-AAAAATATATATATAAAACTAATAATTAT 199
L6.02b/s (AY462610)    AAATATTAATATA----------TTTTTTT-AAAAATATATATATAAAACTAATAATTAT 200
MAL7P1.50(AY462611)    AAATATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 203
L6.02c/s (AY462612)    AAATATTAATATA-------AATTTTTTTT-AAAAATATATATATAAAACTAATAATTAT 206
S5.12b/s (AY462613)    AAATATTAATATA-------AATTTTTTTT-AAAAATATA----TAAAACTAATAAGTAT 196
L1.12a/s (AY462614)    AAGTATTAATATA-------TATTTTTTTT-AAAAATATATATATAAAACTAATAAGTAT 201
S1.08b/s (AY462615)    ATGTATTAATATA-------TTTTTTTTTT-AAAAATATATATATAAAACTAATAATTAT 201
S3.12c/s (AY462616)    AAGTATTAATATA-------TATTTTTTT--AAAAATATATATATAAAACTAATAATTAT 200
```

```
L3.08f/s (AY462617)    AAATATTAATATA-------AATTTTTTTTTTTTTAAAAATATATAAAACTAATAATTAT 201
L3.08l/s (AY462618)    AACTAAATATATAATTTTTTTTTTTCTTTTTTAAAAATATATATATAAAACTAATAATAAT 212
S3.12e/s (AY462619)    AAGTATTAATATA-----TATTTTTTTTTTTAAAAATATATATATAAAACTAATAATTAT 213
S1.08f/s (AY462620)    ACCTATTAATATA--------TATTTTTTTAAAAATATATATATAAAACTAATAATTAT 200
S1.12a/s (AY462621)    AAGTATTAATATA--------TATTTTTT-AAAAATATATATATAAAACTAATAATTAT 198
L2.06e/s (AY462622)    AAGTATTAATATA---------TATTTTTTAAAAATATATATATAAAACTAATAATTAT 199
L3.08e/s (AY462623)    AAGTATTAATATA-------TATTTTTTA--AAAATATATATATAAAACTAATAATTAT 198
L1.11a/s (AY462624)    AAGTATTAATATA-------TATTTTTTT--AAAA---ATAAGTAAAACTAATAATTAT 198
S1.07d/s (AY462625)    AAATATTAATATA-------AATTTTTTT--AAAA---ATATATAAAACTAATAAGTAT 196
L7.11a/s (AY462626)    AAGTATTAATATA-------TATTTTTTAAAAATATATATATATAAAACTAATAATTAT 202
L3.12b/s               ----TCTCATTTA-------TAATTTACA-------AAATATATAAAACTAATAATTAT 42
MAL6P1.316             AGCTATTTGAATA----------TTTGATT----AAAAAATATATAAAACTAGTTATGAT 188

L2.06d/s (AY462590)    TAT---TATATACATATTAAATATTATTTATA--TATATA------TAT-ATAT-TATAT 245
L3.08g/s (AY462591)    TAT---TATATACATATTAAATATTACTTAT---TA-ATA------TAT-ATAT-TATAT 240
L2.08a/s (AY462592)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTATATATATAT 248
S1.12e/s (AY462593)    TAT--ATACAAACATATGAAATATTGTTTTTT--TA-----ACATATATATATATAATAT 253
L3.08b/s               TAT---TATATACATATTAAATATTATTTATT---A-----ATATATATA---------- 245
S1.07a/s (AY462594)    TAT--ATACAAACATATGAAATATTATTTATT---A-----ATATATATA---------- 239
S1.08c/s (AY462595)    TAT---TATATACATATGAAATATTATTTATT--AATATATATATATATATATATTATAT 253
PF08_0106(AY462596)    TAT---TATATACATATTAAATATTATTTATA--AA------TATATATA---------- 238
S1.08g/s (AY462597)    TAT---TATATACATATGAAATATTATTTATT--AA------TATATATA---------- 237
PF08_0140(AY462598)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 244
S5.12a/s (AY462599)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 249
L3.06a/s (AY462600)    TAT---TATATACATATGAAATATTATTTATT--AATATA------TATTA----TATAT 244
S1.08d/s (AY462601)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 235
S3.12a/s (AY462602)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TAT---------- 236
L2.06f/s (AY462603)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 233
S1.12c/s (AY462604)    TAT--ATACAAACATATGAAATATTATTTATT--AATATA------TATTA----TATAT 245
L7.11b/s (AY462605)    TAT--ATACAAACATATGAAATATTATTTATT--AATATATATATATATAA----TATAT 252
S1.07c/s (AY462606)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 233
L3.08a/s (AY462607)    TAT---TATATACATATTAAATATTACTTATT--AATATA------TATTA----TATAT 243
S1.07b/s (AY462608)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TAT---------- 237
S3.12d/s (AY462609)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 244
L6.02b/s (AY462610)    TAT---TATATACATATGAAATATTATTTATT--AATATA------TATTA----TATAT 245
MAL7P1.50(AY462611)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA----TATAT 248
L6.02c/s (AY462612)    TAT---TATATACATATGAAATATTATTTATT--AATATA------TATTA---TATAT 252
S5.12b/s (AY462613)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATAT----TATAT 241
L1.12a/s (AY462614)    TAT---TATATACATATTAAATATCATTTATT--AATATATAT---TAT---------AT 244
S1.08b/s (AY462615)    TAT---TATATACATATTAAATATTATTTATT--AATATATATA--TATT--------AT 246
S3.12c/s (AY462616)    TAT---TATATACATATTAAATATTATTTATT--AATATATATA--TATAA------TAT 247
L3.08f/s (AY462617)    TAT---TATATACATATTAAATATTATTTATT--AATATATATA--TATA---------- 244
L3.08l/s (AY462618)    TACATACATATACATATTAAATATTATTTATT--AATATATATATAAATATATAT-ATAT 269
S3.12e/s (AY462619)    TAT--ATACAAACATATGAAATATTATTTATT----AATA------TATAT--------- 252
S1.08f/s (AY462620)    TAT--ATACAAACATATGAAATATTATTTATT--AATATA------TATAT---TATATA 247
S1.12a/s (AY462621)    TAT---TATATACATATTAAATATTGTTTTTTT-AACATA------TACATA--TATATA 246
L2.06e/s (AY462622)    TAT--ATACAAACATATGAAATATTATTTATT--AATATA------TATAT--------- 240
L3.08e/s (AY462623)    TAT---TATATACATATTAAATATTACTTATT--AATATA------TAT---------- 236
L1.11a/s (AY462624)    TAT---TATATACATATTAAATATTATTTCCT--AATATA------TATTGT-------- 239
S1.07d/s (AY462625)    TAT---TATATACATATTAAATATTATTTATT--AATATA------TATTA------TAT 239
L7.11a/s (AY462626)    TAT---TATATACATATTAAATATTACTTATT--AATATA------TATTA------TAT 245
L3.12b/s               TAT---TATATACATATTAAATATTATTTTTTT-AACATATACATATATTGTAATATTAT  98
MAL6P1.316             TAT--ATACATA-ATATAAATTTTTTTTTTTTTTTTGTACATATACATATATATATATA 245

L2.06d/s (AY462590)    ATATATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 294
L3.08g/s (AY462591)    ATATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 289
L2.08a/s (AY462592)    ATATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 297
S1.12e/s (AY462593)    ATATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATACA--AAT 302
L3.08b/s               --ATATTATAA-TATTACAAC-TAT-------TATAATTAATAT--ATATATATA--AAT 290
S1.07a/s (AY462594)    ---TATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 283
S1.08c/s (AY462595)    ATATATTATAA-TATTACTAC-TAT-------TATAATTA------ATATATACA--AAT 298
PF08_0106(AY462596)    ---TATTATAA-TATTACAAC-TAT-------TATAATTAATATATATATATATA--AAT 284
S1.08g/s (AY462597)    ---TATTATAA-TATTACTAC-TAT-------TATAATTAATATATATATATAAA--TAT 283
PF08_0140(AY462598)    ATATATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 293
S5.12a/s (AY462599)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTA------ATATATATA--AAT 292
L3.06a/s (AY462600)    ATAT--TATAA-TATTACAAC-TAT-------TATAATTAATAT--ATATATATA--AAT 291
S1.08d/s (AY462601)    A-ATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATACA--AAT 282
S3.12a/s (AY462602)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 283
L2.06f/s (AY462603)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTAATAT--AT----ATA--AAT 276
S1.12c/s (AY462604)    ATAT--TATAA-TATTACAAC-TAT-------TATAATTAATAT--ATATATATA--AAT 292
L7.11b/s (AY462605)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 299
S1.07c/s (AY462606)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 280
L3.08a/s (AY462607)    ATAT--TATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 290
S1.07b/s (AY462608)    --ATATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATATA--AAT 282
S3.12d/s (AY462609)    --ATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 289
L6.02b/s (AY462610)    --ATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 290
MAL7P1.50(AY462611)    ATATATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 297
```

```
L6.02c/s (AY462612)    ATATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 301
S5.12b/s (AY462613)    ATATATTATAA-TATTACAAC-TAT-------TATAATTAATAT--ATATATACA--AAT 290
L1.12a/s (AY462614)    ATATATTATAA-TATTACTAC-TAT-------TATAATTACTATATATATATACA--AAT 295
S1.08b/s (AY462615)    ATATATTATAA-TATTACTAC-TAT-------TATAATTA------ATATATATA--AAT 291
S3.12c/s (AY462616)    ATATATTATAA-TATTACAAC-TAT-------TATAATTACTATATATATATATA--AAT 298
L3.08f/s (AY462617)    --ATATTATAA-TATTACAAC-TAT-------TATAATTACTAT--ATATATACA--AAT 289
L3.08l/s (AY462618)    ATA-------------------------------------------------------- 272
S3.12e/s (AY462619)    ----ATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATATA--AAT 295
S1.08f/s (AY462620)    TA-TATTATAA-TATTAGAAC-TAT-------TATAATTACTAT--ATATATATA--AAT 295
S1.12a/s (AY462621)    TAATATTATAA-TATTACTAC-TAT-------TATAATTA--AT--ATATATATA--AAT 293
L2.06e/s (AY462622)    ----ATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 283
L3.08e/s (AY462623)    --ATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 281
L1.11a/s (AY462624)    -AATATTATAA-TATTACTAC-TAT-------TATAATTACTAT--ATATATACA--AAT 286
S1.07d/s (AY462625)    ATATATTATAA-TATTACTAC-TAT-------TATAATTACTATATATATATATA--AAT 290
L7.11a/s (AY462626)    ATATATTATAA-TATTACTAC-TAT-------TATAATTAATAT--ATATATATA--AAT 294
L3.12b/s               A-ATAGTACAACTATTAATATATATATATATATATATACAATATTTATATATATTGTAAT 158
MAL6P1.316             ATATATTATAA-CATTACAAA-TAT-------TATAATAAATATTTATATATATAAAAAT 298

L2.06d/s (AY462590)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 346
L3.08g/s (AY462591)    ATATATACAATAC-TTATATATACATAT----TTCAACAAA-A--A--CAATATTATTAT 339
L2.08a/s (AY462592)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 349
S1.12e/s (AY462593)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 354
L3.08b/s               ATATATATAATAC-TTATATATATATATATTCCAACACA-ATAC--TATTATTATTAT 348
S1.07a/s (AY462594)    ATATATATAATAC-CTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 337
S1.08c/s (AY462595)    ATATATATAATAC-TTATATATATATATAT--TTCAACA-A-ATA---TAATATCATTAT 350
PF08_0106(AY462596)    ATATATATAATAC-TTATATATATATAT----TTCAACA-A-ATA---TAATATCATTAT 336
S1.08g/s (AY462597)    ATATATATAATAC-TTATATATATATAT----TTCAACACA-ATAC--TATTATTATTAT 337
PF08_0140(AY462598)    ATATATATAATAC-TTATATATATATATATATTCCAACACA-ATAC--TATTATTATTAT 349
S5.12a/s (AY462599)    ATATATATAATAC-TTATATATATAT----ATTCCAACACA-ATACAATATTATTATTAT 346
L3.06a/s (AY462600)    ATATATATAATAC-TTATATATATATATATTCCAACACA-ATAC---AATATTATTAT 346
S1.08d/s (AY462601)    ATATATATAATAC-TTATATATATAT----ATTTCAACAAA-A-ACAATATTATTATTAT 336
S3.12a/s (AY462602)    ATATATATAATAC-TTATATATATAT----ATTCCAACA-A-ATACAATATTATCATTAT 336
L2.06f/s (AY462603)    ATATATATAATAC-TTATATATATAT----ATTCCAACACA-ATACAATATTATTATTAT 330
S1.12c/s (AY462604)    ATATATATAATAC-TTATATATATAT----ATTTCAACACA-ATAC--TATTATTATTAT 344
L7.11b/s (AY462605)    ATATATATAATAC-TTATATATATAT----ATTTCAACA-A-ATAT---AATATCATTAT 349
S1.07c/s (AY462606)    ATATATATAATAC-TTATATATATAT----ATTTCAACA-A-ATAT---AATATCATTAT 330
L3.08a/s (AY462607)    ATATATATAATAC-TTATATATATAT----ATTTCAACACA-ATAC--TATTATTATTAT 342
S1.07b/s (AY462608)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 336
S3.12d/s (AY462609)    ATATATATAATAC-TTATATATATATAT----TTCAACA-A-ATA---TAATATCATTAT 341
L6.02b/s (AY462610)    ATATATATAATAC-TTATATATATATATATATTCCAACACA-ATAC--TATTATTATTAT 348
MAL7P1.50(AY462611)    ATATATATAATAC-TTATATATATATATAT--TCCAACACA-ATAC--TATTATTATTAT 351
L6.02c/s (AY462612)    ATATATACAATAC-TTATATATATATATATATTCCAACACA-ATAC--TATTATTATTAT 357
S5.12b/s (AY462613)    ATATATACAATAC-TTATATATATATATATATTCCAACACA-ATAC--TATTATTATTAT 346
L1.12a/s (AY462614)    ATATATATAATAC-TTATATATATATAATTTTTCCAACACA-ATAC--TATTATTATTAT 351
S1.08b/s (AY462615)    ATATATATAATAC-TTATATATATATA----TTCCAACACA-ATAC--TATTATTATTAT 343
S3.12c/s (AY462616)    ATATATATAATAC-TTATATATATATA----TTTCAACACA-ATAC--TATTATTATTAT 350
L3.08f/s (AY462617)    ATATATATAATAC-TTATATATATATA----TTCCAACACA-ATAC--TATTATTATTAT 343
L3.08l/s (AY462618)    ---------------------------------------------------------- 273
S3.12e/s (AY462619)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 349
S1.08f/s (AY462620)    ATATATATAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 347
S1.12a/s (AY462621)    ATATATATAATAC-TTATATATATATAT----TTCAACACA-ATAC--TATTATTATTAT 345
L2.06e/s (AY462622)    ATATATACAATAC-TTATATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 337
L3.08e/s (AY462623)    ATATATATAATAC-TTATATATATATATATATTCCAACA-A-ATAT---AATATCATTAT 337
L1.11a/s (AY462624)    ATATATATATAA---TACTTATATATATATATTTCAACA-A-AAAC---AATATTATTAT 339
S1.07d/s (AY462625)    ATATATATAATAC-TTTTATATATATAT----TCCAACACA-ATAC--TATTATTATTAT 342
L7.11a/s (AY462626)    ATATATATAATAC-TTATATATATAT------TTCAACA-A-AAAC---AATATTATTAT 342
L3.12b/s               ACATAAATTATACCTTACATATATATATACA-TTCACAAAA-GTGT--TATTATTCTTAT 215
MAL6P1.316             ATATATATA-TATATAATATTTATATATATATCATAATAAATATATTAAATCTTACATAT 357

L2.06d/s (AY462590)    TCTACCATATCAC--AATACTCCCATAACATA-CATAC--------ATATATACCCCCAC 395
L3.08g/s (AY462591)    TCTACCATATCAC--TATACTCCCATAACATA-CATAT--------ATACATAC------ 382
L2.08a/s (AY462592)    TCTACCATATCAC--TATACTCCCATAACATA-CATAC--------ATACATACCCCCAC 398
S1.12e/s (AY462593)    TCTACCATATCAC--AATACTTCCATAACATA-CATACATACATATATACATACCCCTAC 411
L3.08b/s               TCTACCATATCAC--AATACTCCCATAACATA-C---------------------------- 375
S1.07a/s (AY462594)    TCTACCATATCAC--AATACTCCCATAACATA-CATACATAC------------------- 376
S1.08c/s (AY462595)    TCTACCATATCAC--TATACTCCCATAACATA-CATACAATCACCC---CACACCACACC 404
PF08_0106(AY462596)    TCTACCATATAAC--AATACTTCCATAACATA-CATACAATCACCC---CACACCACACC 390
S1.08g/s (AY462597)    TCTACCATATCAC--AATACTTCCATAACATA--ACATA---------------------- 372
PF08_0140(AY462598)    TCTACCATATCAC--AATACTCCCATAACATA-CATAC--------ATACATATATACAT 398
S5.12a/s (AY462599)    TCTACCATATCAC--AATACTCCCATAACATA-CATAT--------ATAC---------- 385
L3.06a/s (AY462600)    TCTACCATATCAC--AATACTCCCATAACATA-CGCA---------ATACGCCA------ 388
S1.08d/s (AY462601)    TCTACCATATTAC--AATACTCCCATAACATA-CGAA---------ATACGCCA------ 378
S3.12a/s (AY462602)    TCTACCATATCAC--AATACTCCCATAACATA-CGCA---------ATACGCCA------ 378
L2.06f/s (AY462603)    TCTACCATATTAC--AATACTCCCATAACATA-CATAC--------ATACATAC------ 373
S1.12c/s (AY462604)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCCA------ 386
L7.11b/s (AY462605)    TCTACCATATCAT--AATACTCCCATAACATA-CGCA---------ATACGCC------- 390
S1.07c/s (AY462606)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCCACCACCA 378
```

```
L3.08a/s (AY462607)    TCTACCATATCAC--AATACTCCCATAACATA-CGCA---------ATACGCCA------ 384
S1.07b/s (AY462608)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCC-ACCA-- 381
S3.12d/s (AY462609)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCC-ACCA-- 386
L6.02b/s (AY462610)    TCTACCATATCAC--AATACTCCCATAACATA-CGCA---------ATACGCC-ACCA-- 393
MAL7P1.50(AY462611)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCC-ACCA-- 396
L6.02c/s (AY462612)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGCC-ACCA-- 402
S5.12b/s (AY462613)    TCTACCATATCAC--TATACTCCCATAACATA-CATACATATATACATACCCCTACCAAA 403
L1.12a/s (AY462614)    TCTACCATATCAC--TATACTCCCATAACATA-CGCA---------ATACGC-CACCACC 398
S1.08b/s (AY462615)    TCTACCATATCAG--AATACTCCCATAACATAACATA---------ACATACATACCCCC 392
S3.12c/s (AY462616)    TCTACCATATCAC--AATACTCCCATAACATA-CGCA---------ATACGC-CACCACC 397
L3.08f/s (AY462617)    TCTACCATATCAC--TATACTCCC---------------------------------- 365
L3.08l/s (AY462618)    -CATCCAAATATTACAATACTCCCATAACATA--ACAT--------A-ACATACATAT-- 319
S3.12e/s (AY462619)    TCTACCATATCAC--AATACTCCCATAACATA-CATAT--------ATACATACATAC-- 396
S1.08f/s (AY462620)    TCTACCATATCAC--AATACTCCCATAACATA-CATAC--------ATACATACATA--- 393
S1.12a/s (AY462621)    TCTACCATATCAC--TATACTCCCATAACATA-CATAA--------ATATATATATAA-- 392
L2.06e/s (AY462622)    TCTACCATATTAC--AATACTCCCAAAACATA-CATAC--------ATACCCCTAC---- 382
L3.08e/s (AY462623)    TCTACCATATCAC--TATACTCCCATAACATAACATAA-------CATACAT-------- 380
L1.11a/s (AY462624)    TCTACCATATCAC--AATACTCCCATAACATA-CATA--------CATACCC-------- 380
S1.07d/s (AY462625)    TCTACCATATCAC--TATACTCCCATAACATA-CATACATATATACATACCCCCACGT-- 397
L7.11a/s (AY462626)    TCTACCATATCAC--TATACTCCCATAACATA-CATATATACATACATACCCCCACGT-- 397
L3.12b/s               TCTACCATATTAT--AATACTCCCATAACATAACATAA-------CATACATACCCCCAC 267
MAL6P1.316             ATATACATTTCAACAAATACATTCTTATTCTACTATATTACA----ATAATCTCATAACA 415

L2.06d/s (AY462590)    G------TACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAACCA-----C 442
L3.08g/s (AY462591)    ------------ACCAAA-----AGCAAACCACTTACGC-CAC-----TTACCA-----C 412
L2.08a/s (AY462592)    G------TACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAACCA-----C 445
S1.12e/s (AY462593)    CAAACACCTACCAAACACCTACCACTCCACCGCCCACACGAACCATGCAAACC------- 464
L3.08b/s               -----CCCTACCAAACACCTACCACTCCACCGCCCACACTTACCATGCAAACC------- 425
S1.07a/s (AY462594)    -----CCCTACCAAACACCTACCACTCCACCGCCCACACTTACCATGCAAACC------- 422
S1.08c/s (AY462595)    ACACCACCTACCAAACACCTACCAC-----CGCCCACACGAACCATGCAAACC------- 452
PF08_0106(AY462596)    ACACCACCTACCAACCACCTACCAC----------TTCAAGAACCATGCAAACC------- 434
S1.08g/s (AY462597)    -----ACATAC-ATACCCCTACCAC-----CGCCCACACGAACCATGCAAACCACGTATA 419
PF08_0140(AY462598)    ACCTA--CATACCCCTACCAAACACGTA-CCACGTATGA-CATAATGTAGTCTGGA-CGA 453
S5.12a/s (AY462599)    ------------CCCTACCAAACACCTA-CCACGTATGG-CATAATGTAGTCGGGA-AGA 428
L3.06a/s (AY462600)    --------CCACCACAACCACCTACCAAACCATGTATGC-CACGATATAAACCACG-TAT 436
S1.08d/s (AY462601)    --------CCACCACCGCCAA-CACGAA-CCATGCACGC-CACAAAATTGTATG------ 419
S3.12a/s (AY462602)    --------CCACCACCGCCCA-CACTTA-CCATGCAAAC-CCACAAATATATT------- 418
L2.06f/s (AY462603)    --------ATACCCCTACCAA-A---CA-CCACTTA----CCACGTATGCATGA------ 408
S1.12c/s (AY462604)    --------CCACCACTCCCCA--AACGAACCATGCAAAC-CCACAAATATATGT---ATG 430
L7.11b/s (AY462605)    ------------------------------------------------------------
S1.07c/s (AY462606)    --------CCGCCCACACGAACCATGCAAACAC-TAAAC-CACCACTTACCATGCA-AAC 425
L3.08a/s (AY462607)    --------CCACCACCGCCCA-CACGAA-CCATGCAAGA-CCCCACAAATATATGT-ATG 430
S1.07b/s (AY462608)    ------------CCACCGCCAA-CACGAA-CCATGCACGC-CACAAAAT---TGT---ATG 419
S3.12d/s (AY462609)    ------------CCACCGCCAA-CACGAA-CCATGCAAAC-CCACAAATATATGT---ATG 427
L6.02b/s (AY462610)    ------------CCACCGC-------------------------------------- 398
MAL7P1.50(AY462611)    ------------CCACTCCCCA-AACGAA-CCATGCAAGC-TCCAAAATATATGT---ATG 437
L6.02c/s (AY462612)    ------------CCACCGCCCA-CACTTA-CCACG---------------TATGT---ATG 429
S5.12b/s (AY462613)    CACCTACCACTCCACCGCCCA-CACGAA-CCATGCAAAC-CCTCACATATATTT---ATG 457
L1.12a/s (AY462614)    ACCG----CCAACACGAACCATG--CAAACCACGTATAC-CACGATATAAAC-C----AC 446
S1.08b/s (AY462615)    ACGT----ACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAAC-C----AC 442
S3.12c/s (AY462616)    ACCG----CCCACACTTACCATG--CAAACCACGTATAC-CACG-TATGTATGC----AT 445
L3.08f/s (AY462617)    ------------------------------------------------------------
L3.08l/s (AY462618)    --------AACCACGTATCT---ACCAAACCACTTGC---CACTATATATGCCA--CGAT 362
S3.12e/s (AY462619)    --------ATACAATCACCCCACACCACACCACTTGCGC-CACTTACCATGCAAA-CCAC 444
S1.08f/s (AY462620)    ------------CATACAATCA-CCCCACACCACAC---------CACTTACGCCA------ 425
S1.12a/s (AY462621)    --------GGTTATATACTAC-TGGTAAACTACACTG----ACCATGTATTCCA------ 431
L2.06e/s (AY462622)    ------------CAAACACC----ACCAAACCA-------------TGTATGCCA--CGAT 410
L3.08e/s (AY462623)    ------------ACACCAAAC-----CAAACCACTTACGC-CACTT-----ACCA-----C 411
L1.11a/s (AY462624)    ------------CTACCAAACACCACCAAACTATGTATGC-CACGATATAAACCA-----C 421
S1.07d/s (AY462625)    --------ACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAACCA-----C 441
L7.11a/s (AY462626)    --------ACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAACCA-----C 441
L3.12b/s               G------TACGTACCAAAACACCACCAAACCATGTATGC-CACGATATAAACCA-----C 313
MAL6P1.316             TAGA-TTATTATATATATATATTGATAAAACTATTATTATTATTAT-TATTTTA----TC 467

L2.06d/s (AY462590)    GTATGTATGCAT-------GTATGACATAATGTAGTGGTGGAGTT-AACAAAAATGGCGC 495
L3.08g/s (AY462591)    GTATGCATG----------ACATAATGTAGTCTGGACGAAGAAT---ATAAAAATGG--- 454
L2.08a/s (AY462592)    GTATGCATG---------------ACATAATGTAGTGGTGGAGTT-AGCAAAAATGGCGC 488
S1.12e/s (AY462593)    -CACAAATATATGT--ATGACATGACATAATGTAGTCCGGAGA--ATACAAAAATGGGGC 518
L3.08b/s               -CACAAATATATGT-ATGACATGACATAATGTAGTCACGAA----AAAC-AAAATGGCGC 480
S1.07a/s (AY462594)    -CACAAATATATGT--ATGACATGACATAATGTAGTCATGAAT---AACCAAAATGGTGA 477
S1.08c/s (AY462595)    -CTCACATATATGT--ATGACATGACATAATGTAGTGGGGGGAGTTAACCAAAATGGTGA 508
PF08_0106(AY462596)    -CACAAA-AT-TGT--ATGACATGACATAATGTAGTCAGGAAC---AAACGAAATGGCAG 485
S1.08g/s (AY462597)    CCACGTATGTATGC--ATG-TATGACATAATGTAGT---GCAC---CAACAAAATGGTGG 472
PF08_0140(AY462598)    AGAATAT-----------------------------------------AAAAATGGTTC 500
S5.12a/s (AY462599)    AGAAGAAGAA--------------------------------GAATACAAAAAATGGATC 487
L3.06a/s (AY462600)    GCATGACATAATGTAGTCACG----------------------AACAACCACAATGGCGG 498
S1.08d/s (AY462601)    ACATGAC-----------ATAATG--------TAGTCGCGAACCATAAACAAAATGG--G 460
```

```
S3.12a/s (AY462602)    -TATGAC-----------ATAATG--------TAGTGGTGG-----TGTTAAAATGGCGC 454
L2.06f/s (AY462603)    -CATAAT-----------GTAGTG--------CACCAATAA-----CG--AAAATGGCGC 442
S1.12c/s (AY462604)    ACATGAC-----------ATCATG--------TAGTGGTGGAGT-TAACAAAAATGGCTC 472
L7.11b/s (AY462605)    ------------------------------------------------------------
S1.07c/s (AY462606)    ACTAAACCACCAAACCATCAAACC---------CCACCAAATTGTATAACATCATGTTGT 484
L3.08a/s (AY462607)    ACATGAC-----------ATA--------------------------------ATGTAGT 462
S1.07b/s (AY462608)    ACATGACATA------ATGTAATGGTGG-TGTTAAAATGGCG----CCAGGTAATGCTGG 470
S3.12d/s (AY462609)    ACATGACATA------ATGTAGTGGTGG-AGTTAACAA-------------AAATGGTGC 469
L6.02b/s (AY462610)    ------------------------------------------------------------
MAL7P1.50(AY462611)    ACATGACATA------ATGTAATGGGGGGAATTAACCA-------------AAATGGCGT 480
L6.02c/s (AY462612)    ACATAATG---------------------------------------------------- 439
S5.12b/s (AY462613)    ACATAATGTA------GTGCACCAA-------TAACGA-------------AAATGGCGC 494
L1.12a/s (AY462614)    GTATG---TATGT----ATGTATGACATCATGTTGTCGG----------TACAATGGTTA 498
S1.08b/s (AY462615)    GTATG---TATG-------ACATGACATCATGTTGTCGG----------TACAATGGGGG 491
S3.12c/s (AY462616)    GTATGACATAATG----TAGTATGGAAGAAGAATA-------------CAAAAATGGCGG 501
L3.08f/s (AY462617)    ------------------------------------------------------------
L3.08l/s (AY462618)    ATATACCACGTAT----GCATGACATAATGTAGT-CACGAACGATAAACAAAAATG----- 415
S3.12e/s (AY462619)    GTATACCACGTAT----GTATGT-ATGACATAATGTAGTCCCGAACAAACAAAATGGCTC 501
S1.08f/s (AY462620)    -CTTACCACGTAT----GCATG-----ACATAATGTAGTCACGAACGACCACAATG----- 472
S1.12a/s (AY462621)    -CATATTGTATG-------ACATG-----ACATAATGTAGTCGATGAGTTAAAATG----- 475
L2.06e/s (AY462622)    ATAAACCACGTAT------GTATG-ACATCATGTAGTGGTGGAGTTAAAAAAAATG----- 461
L3.08e/s (AY462623)    GTATG---------------CATGACATAATGT---------------------------- 431
L1.11a/s (AY462624)    GTATG---------------CATGACATAATGTAGTCATGAAT---AACGAAAATGGCGC 465
S1.07d/s (AY462625)    GTATGTAT-----------GTATGACATCATGTTGTCG----------CAACCATGGGGC 482
L7.11a/s (AY462626)    GTATGCAT-----------GTATGACATCATGTAGTGGTGGAGT-TAACAAAAATGGGGC 491
L3.12b/s               GTATG---------------CATGACATAATGTAGTCGGGA-----AGAAAAAATGGCAG 354
MAL6P1.316             ATATCACAATACTCCCATAACATACATACATATATACATACATACATACATACCCC 527
```

upsC-type 5'UTR-ATG

```
L4.03a/c (AY462581)    -CACATACAATACGACTAAGAAACCAAATAACATCACAACAAACATATTGACTACCATTA 59
L7.11a/c (AY462582)    TCACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTA 60
L1.11c/c (AY462583)    --ACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTA 58
S1.12a/c (AY462584)    -CACATATAATACGACTAAGAAACAAAATAACATCATAACAAACATAGTGACTACCATTA 59
L1.11b/c (AY462585)    --CCATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGAATACCATTA 58
L7.12b/c (AY462586)    TCACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTA 60
PFD0615c (AY462587)    -CACATATAGTACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTA 59
L6.02a/c (AY462588)    -CACATATAATACGACTAAGAAACAAAATAACATCATAACAAACATAGTGACTACCATTA 59
L5.08a/c (AY462589)    -CACATATAATACGACTAAGAAACAAAATAACATCACAACAAACATAGTGACTACCATTA 59

L4.03a/c (AY462581)    CATGATATTACCACATAATTCATACCATTATATAATATTACTACATGGAACCGATAACCA 119
L7.11a/c (AY462582)    CATGATATTACCACATAATTCATACCATTATATAATATCACTACATGATAGTGATAACCA 120
L1.11c/c (AY462583)    CATGATATTACCACATAATATAAAGCATTATATAATAT---------ACACCAATACGTA 109
S1.12a/c (AY462584)    AATGGTATTACCACACAATTCATACCACTATATAATATTACTACATGATAGTGGTAACTA 119
L1.11b/c (AY462585)    CATGATATTACCACATAATTCATACCACTATATAATATTACTACATGATAGTGATAACCA 118
L7.12b/c (AY462586)    CATGATATTACCACATAATTCATACCATTATATAATATTACTACATGGTAGTGATAACCA 120
PFD0615c (AY462587)    CATGATATTACCACATAATTCATACCACTATATAATATTACTACATGATAGTGATAACCA 119
L6.02a/c (AY462588)    CATGATATTACCACATAATATAAAGCATTATATAATAT---------ACACCAATACGTA 110
L5.08a/c (AY462589)    CATGATATTACCACATAATTCATACCACTATATAATATTACGACATGGTAGTGATAACTA 119

L4.03a/c (AY462581)    CTATATCATATACACCACTACATATTAATA----CTATCGGAGATATTATGTGCACAAAT 175
L7.11a/c (AY462582)    TTATATCATATACACCACTACATAATAATA----GTAGCGGCGGTATAATGTACATGTAT 176
L1.11c/c (AY462583)    TTATG-----------ACTACATAGTAATA----GTAGCGACGGTATCATATACACGAAT 154
S1.12a/c (AY462584)    CTATATCATATACACCATTATATAGTAATA----GTAGCGGCGGTATCATGCACACGTAT 175
L1.11b/c (AY462585)    CTATATCATATACACCATTATATAGTAATA----GTAGTGGCGGTATAATGTACATGTAT 174
L7.12b/c (AY462586)    CTATATCATATACACCACTATATAGTAATA----ATAGCTGTGGTAATATGTACACGTAT 176
PFD0615c (AY462587)    TTATATCATATACACCACTATCTAATAATA----GTAGCGGAGGTATTATGTGCACAAAT 175
L6.02a/c (AY462588)    TTATG-----------ACTACATAGTAATA----GTAGCGGCGGTATCATGCACACGTAT 155
L5.08a/c (AY462589)    CTATATCATATACAACACTACTAACTATCACTACATAGTAACAGTAGTAGTTACAATCAA 179

L4.03a/c (AY462581)    ATATTATAATAATGGCACCCACAACCACGA-----CATCCTGGGAATATATATTTTCGTT 230
L7.11a/c (AY462582)    ATATTGTAACAATGGTAGCGACAACCACGG-----CATCATGGTAATATAGATTTTCGTT 231
L1.11c/c (AY462583)    ATATAGAAATATTAGTAGTCACAATCAAGAAGTAACATCATGGTAATATAGATTTTCATT 214
S1.12a/c (AY462584)    ATATTGTAACAGTGGTAGCTACAATCACTG-----CATCATGGTAATATAGATTTTCGTT 230
L1.11b/c (AY462585)    ATATTGTAATAGTGGTAGCTACAATCACTG-----CATCATG------------------ 211
L7.12b/c (AY462586)    ATATTGTAATAGTGGTAGCCAAAATCAGGA-----TATCATGGTAATGTAGATTTTCATT 231
PFD0615c (AY462587)    ATATTATAATAGTGGTAGCCACAACCACGG-----TATCATGGTAATATAGATTTTCATT 230
L6.02a/c (AY462588)    ATATTGTAACAGTGGTAGCGACAACCGCGG-----CATCATGGTAATATAGATTTTCATT 210
L5.08a/c (AY462589)    GAAGTAACATCATGGTAATGTATATTTCCG------TTCATTTCATTGTAATGTTTTGAT 233

L4.03a/c (AY462581)    TATATTTTCTTTATCATTTGTTGCCCAT-ACACTATTAATATGTATTTATGTTATAATGG 289
L7.11a/c (AY462582)    TATATCTTCCTTATTGTTTGTTGTGCAT-ACACTATTAATATACATTTATGTTATAATGG 290
L1.11c/c (AY462583)    CATATCTTCCTTATTGTTTGTTGTCCAT-ACACCATTAATATGTATTTATGTTATAATGG 273
S1.12a/c (AY462584)    TATATCTTCTTTATTGTTTGTTGTCCAT-ACACTATTAATATGTATTTATGTTATAATGG 289
L1.11b/c (AY462585)    ------------------------------------------------------------ 
L7.12b/c (AY462586)    CATATCTTCCTTATCGTTTGTTGTGCAT-ACACTATTAATATACATTTATGTTATAATGG 290
PFD0615c (AY462587)    CATATCTTCCTTATTGTTTGTTTTCCAT-ACACTATTAATATGTATTTATGTTATAATGG 289
L6.02a/c (AY462588)    CATATCTTCCTTATCGTTTGTTGTGTAT-ACACTATTAATATGTATTTATGTTATAATGG 269
L5.08a/c (AY462589)    GATAATGATGTTGTGTTTTGTTATAGATTGTGATAACAAGATATATGAATCCCATATTAG 293

L4.03a/c (AY462581)    AAAACTATGGTAACAATGTATGAATG-ACCCTCATAAATTAATAACACACACCTCAAAAC 348
L7.11a/c (AY462582)    TAGACTATGTTAACAATGTATGAATG-ACTATCGTAAATTAATAATAGATACATGAAAAC 349
L1.11c/c (AY462583)    TAGACTATGTTAACAATGTATGAATG-ATCATCGTAGATTAATAATAAATTCATGAAAAC 332
S1.12a/c (AY462584)    TAAACTATGTTAACAATGTATGAATG-ATCATCGTAGATTAATAATAGATGCATGGAAAC 348
L1.11b/c (AY462585)    ---ACTATGTTAACAATGTATGAATG-ACCATCATAAATTAATAACAGACGCATCAAAAC 267
L7.12b/c (AY462586)    TAGACTATGTTAACAATGTATGAATG-TCCATCATAAATTAATAATAGATGCATAAAAAC 349
PFD0615c (AY462587)    TAGACTATGTT---AATGTATGAATG-ACCATCATAAATTAATAACAGACGCATCAAAAC 345
L6.02a/c (AY462588)    TAGACTATGTTAACAATGTATGAATG-ACCATCGTAGATTAATAATAGATACATGAAAAC 328
L5.08a/c (AY462589)    --AGTAATAATGTGCATGGATGACAGTGGTAATATATACTAGCAGTGTATGAAAAAAAAT 351

L4.03a/c (AY462581)    A-ATGT------ATATGT------------GTGCATTTACAACA-TAATGTATTCCGGGG 388
L7.11a/c (AY462582)    T-GTGT------ATATGTAT--------GTGTGCATTTACAACA-TAATGTAGTCCGGAG 393
L1.11c/c (AY462583)    A-ATGTGTATGTATATGTAT--------GTGTGTATTTATGACA-TAATGTAATCGGAAA 382
S1.12a/c (AY462584)    C-GTGTATATGTATATGTATATATATATGTGTGCATTTATGACA-TAATGTAGTCGGGAA 406
L1.11b/c (AY462585)    A-ATGT------ATATGT------------GTGCATTTACAACA-TAATGTAGTCCGGAG 307
L7.12b/c (AY462586)    A-ATGT------ATATGTATATGTAT--GTGTGCATTTACAACA-TAATGTAGTCATGAA 399
PFD0615c (AY462587)    A-ATGT------ATATGT------------GTGCATTTATGACA-TAATGTAGTCGTGAA 385
L6.02a/c (AY462588)    A-ATGTGTATGTATATGTAT--------GTGTGCATTTATGACA-TCATGTAGTCACGAA 378
L5.08a/c (AY462589)    ATATGTATAA--CAATGTATATGTTTATGTATGCATTTATGACA-TAATGTAATCGGA-- 406
```

Legend to multiple sequence alignments

Multiple alignments (ClustalW) are shown for DBL1α, CIDR1α, DBLβ, NTS-DBL1α, and for 400bp of 5' UTR to ATG of upsB- and upsC-type. Sequence names indicate "timepoint.child.sequence": "L" for 4 months and "S" for 1 month longitudinal study (e.g. "S3.12b": 1 month study, time point 3 (=week 4), child 12, sequence b). Names of 5'UTR sequences are marked with "c" for central upsC-type and "s" for subtelomeric upsB-type upstream sequences. The sequences specifically mentioned in text are highlighted in yellow and green. The accession numbers of the sequences are shown in brackets. Sequences which were isolated from 3D7 culture strain are named according to plasmoDB annotations. The following sequences were isolated from both, children and 3D7: the 5' UTR sequence L3.08l/s of upsB-type (identical sequence in PF08_0103), and the DBL1α sequences L8.06f (PFL1970w), L7.12c (PFD0995c) and S3.12a (PF08_0141).

Supplement II: Study Forms of Malaria Case-Control Study

Study enrollment form – severe malaria case control study,
Modilon Hospital, Madang, PNG Feb-May 2003

Date (Day, Month, Year) | d | d | m | m | y | y | Bleed code

Identification **ID number**

Kristen nem	
Nem bilong papa	
Nem bilong mama	
Kolim nem bilong ples bilong yu (nem bilong **liklik ples**)	
Kolim nem bilong **as ples tru** na province bilong papa bilong yu	
Hamas christmas bilong yu	
Sex (F=female, M=male)	

History of disease *(ask child or parents)*

Yu bin sik tude o long las wik		Y/N
Sapos yes, wanem dei sik i kamap	(**0**=Tude, **1**=Asde, **2**=Asde bipo, **3**=Tripela dei i go pinis, **4**=Foapela dei i go pinis, **5**=Fivepela dei o moa i go pinis (hamas dei?))	
Wanem kain sik i kamap	Yu gat skin hot o skin col	Y/N
	Yu gat het pen *(ask child)*	Y/N
	Yu gat kus	Y/N
	Yu gat traut	Y/N
	Yu gat pekpek wara	Y/N
	Yu gat guria/ai tanim/hamas, wataim	
	Hap i dai (unconscious)	Y/N
	Arapela kain sik	Y/N
	Sapos yes, yu pin pilim olsem wanem	

Yu bin go long Helt Senta long las wik	Y/N
Yu bin kisim marasin o sut long las wik	Y/N
Sapos yes, hamas dei bipo	
Sapos yes, wanem nem bilong marasin na bilong wanem sik:	
Health book, date of last antimalarial treatment	

Severe malaria case-control stadi. Tok orait

Mi tok orait na long laik bilong mi yet pikinini bilong mi bai givim blut long dispela stadi "severe malaria case-control study" bilong PNGIMR. Ol i givim gutpela tok save long mi na long wanem kain wok bai ol i wokim wantaim dispela samples. Na mi save tu olsem sapos mi no laik pikinini bilong mi givim blut, em bai mi no mas sign.

Signed:

ID number | | Kristen nem | |

Clinical assessment:

MUAC	cm
Axillary temperature	˚C
Level of consciousness	
	*blantyre score **1-5***
Abnormal neurological features	Y/N
	***1** =hyperextension neck, **2** =limb rigidity, **3** =tooth grinding, **4** =upward deviation eyes, **5** =other, specify*
Prostration	Y/N
Convulsions	/24h
Respiratory rate/depth	/min /
	***1** =normal, **2** =deep, **3** =shallow, **4** =irregular, **5** =fast*
Chest indrawing/flaring nostril	Y/N / Y/N
color face / temp.skin limbs	/
	***1** =normal, **2** =pallor, **3** =other / **1** =normal temp, **2** =hot, **3** =cold, clammy skin*
Abnormal bleeding	Y/N
Hemoglobinuria	Y/N
	***9** =not done*
Other signs	

Laboratory parameters:

Parasites	P.f.	P.v. / P.m.	Per/200WBC
Hemoglobin value			g/dl
Lactate			mmol/L
Blood glucose			mmol/L

Comments	

CSF analysis result	
	1=done, positive for Meningitis, **2**=done, negative for Meningitis, **9**=not done
Any other disease suspected	
Any other disease (laboratory) confirmed	

Checklist for enrolment of children in Malaria Case-Control Study

Severe malaria criteria:

- Coma	**Blantyre coma score ≤ 3** **=> wait 30 min after generalized convulsion**
- Prostration	**Inability to sit without support or inability to feed...**
- Convulsions	**> 1/24h - generalized or focal, limbs, facial muscle, jerky deviation eye, excess salivation, irregular breathing patterns**
- Respiratory distress	**Chest in-drawing, grunting, irregular, abnormally deep respiration, flaring nostrils**
- Hyperlactatemia	**Lactate > 5 mmol/L**
- Hypoglycemia	**Glucose < 2.2 mmol/L**
- Severe anaemia	**Haemoglobin < 5 g/dl**
- Hemoglobinuria	(Make sure not by Glucose-6-phosphate dehydrogenase deficiency after primaquine treatment)
- Collapse	cold clammy skin (Systolic blood pressure [mm Hg] < 60 if younger than 5 years, or < 80 if older than 5 years)
- Abnormal bleeding	Spontaneous bleeding gums, nose, gastrointestinal tract....
- Jaundice	Yellow sclera, mucosal surface mouth

Blantyre Coma Score:

Response	Findings	Score
best motor response	localizes painful stimulus (pressure with blunt end of pencil on sternum/breast bone or supraorbital ridge)	2
	withdraws limb from painful stimulus (pressure with horizontal pencil on nail bed of finger or toe)	1
	no response or inappropriate response	0
best verbal response	cries inappropriately with painful stimulus or if verbal speaks	2
	moan or abnormal cry with painful stimulus	1
	no vocal response to painful stimulus	0
eye movement	watches or follows (**follow mother's face**)	1
	fails to watch or follow	0

Symptoms of acute, mild malaria:
- *P. falciparum* parasites
- Headache
- Fever, chills
- Muscle aches
- Diarrhoea, nausea

Age-matched children:
- less than 6 months old children: +/- 1 month
- older than 6 months: +/- 20% of age (e.g. 5 years old child: 4-6 years old child is age-matched)

Exclusion criteria:
- Anti-malarial treatment in the last 2 weeks
- Co-infection with any other disease
- Severe malnutrition (MUAC < 12 cm)

Supplement III: Transcriptional regulation of *P. falciparum var* genes

Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters

Till S. Voss, Mirjam Kaestli, Denise Vogel*, Selina Bopp and Hans-Peter Beck[1]

Department of Medical Parasitology and Infection Biology
Swiss Tropical Institute
Socinstrasse 57
4051 Basel, Switzerland

*Current address:
Institute of Molecular Biology
University of Zurich
Winterthurerstrasse 190
8057 Zurich, Switzerland

[1]Correspondence:
Hans-Peter.Beck@unibas.ch; Tel: +41-61-284-81-16; Fax: +41-61-271-86-54.

Published in *Molecular Microbiology* (2003) Jun;48(6):1593-1607.

Molecular Microbiology (2003) **48**(6), 1593–1607 doi:10.1046/j.1365-2958.2003.03528.x

Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters

Till S. Voss,† Mirjam Kaestli, Denise Vogel,‡ Selina Bopp and Hans-Peter Beck*
Swiss Tropical Institute, Department of Medical Parasitology and Infection Biology, Socinstrasse 57, 4051 Basel, Switzerland.

Summary

The *Plasmodium falciparum* virulence factor PfEMP1 is responsible for both antigenic variation and cytoadherence of infected erythrocytes in malaria. Approximately 50 *var* genes per parasite genome code for this highly polymorphic surface protein. We showed recently that chromosome-central and subtelomeric *var* genes are controlled by different promoters. Here, we report that transcriptional repression of *var* genes located in different chromosomal regions occurs by different mechanisms. Subtelomeric *var* gene transcription is repressed 4–8 h before that of chromosome-central *var* genes. Both repression events coincide with the shifted expression of two distinct nuclear proteins binding specifically to conserved sequence motifs, *SPE1* and *CPE*, present in the respective promoter. Furthermore, a reiterated and highly conserved subtelomeric *var* promoter element (*SPE2*) interacts with a nuclear factor exclusively expressed during S-phase. Promoter analysis by transient transfection suggested direct involvement of these interactions in *var* gene repression and silencing, and identified regions implicated in transcriptional activation of *var* genes.

Introduction

Cytoadherence of parasite-infected red blood cells (iRBCs) to host endothelial cells constitutes a major virulence determinant in *Plasmodium falciparum* malaria. This adhesion is mediated through a large parasite-encoded antigen, termed *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is deposited on the surface of infected erythrocytes during the advanced intraerythrocytic cycle (Leech *et al*., 1984; Biggs *et al*., 1991; Roberts *et al*., 1992). PfEMP1-mediated binding of iRBCs to endothelial cells occurs via host surface receptors such as ICAM-1, CD36 and CSA (Baruch *et al*., 1996; Gardner *et al*., 1996; Reeder *et al*., 1999). As a consequence, iRBCs sequester in microvasculatory capillaries of various organs and thus contribute to the severe morbidity and mortality associated with malaria tropica (Pongponratn *et al*., 1991; Berendt *et al*., 1994). Sequestration in the brain is involved in cerebral malaria with frequent fatal outcomes (Turner *et al*., 1994; Newbold *et al*., 1997), whereas binding to CSA and hyaluronic acid in the placenta poses a risk to both pregnant women and the foetus (Fried and Duffy, 1996; Beeson *et al*., 2000). PfEMP1 has also been shown to downregulate dendritic cell function (Urban *et al*., 1999) and to be involved in rosetting (Rowe *et al*., 1997; Chen *et al*., 1998a), another determinant of disease severity. Owing to its exposure on the surface of iRBCs, PfEMP1 is recognized by variant-specific antibodies, and these play a pivotal role in protection from clinical disease (Bull *et al*., 1998). However, antigenic variation of PfEMP1 allows the parasite to escape existing immune responses and, hence, to establish chronic infections.

PfEMP1 is encoded by the *var* gene family, which comprises ≈50 highly diverse genes per haploid genome (Baruch *et al*., 1995; Smith *et al*., 1995; Su *et al*., 1995; Gardner *et al*., 2002). The majority of *var* genes are located in the subtelomeric region of nearly all chromosomes, but are also found in central clusters on chromosomes 4, 6, 7, 8 and 12 (Rubio *et al*., 1996; Fischer *et al*., 1997; Hernandez-Rivas *et al*., 1997; Thompson *et al*., 1997; Gardner *et al*., 2002). We showed recently that, despite the tremendous sequence diversity observed in *var* gene coding regions, *var* gene promoters are highly conserved and exist in two distinct forms (Voss *et al*., 2000). With the completion of *P. falciparum* genome sequencing, a third type of *var* upstream sequence became apparent, and the three classes have been termed upsA, upsB and upsC (Gardner *et al*., 2002). The type of promoter associated with a particular *var* gene is

Accepted 19 February, 2003. *For correspondence. E-mail Hans-Peter.Beck@unibas.ch; Tel. (+41) 61 284 81 16; Fax (+41) 61 271 86 54. Present addresses: †The Walter and Eliza Hall Institute of Medical Research, Division of Infection and Immunity, 1G Royal Parade, Parkville, Victoria 3050, Australia. ‡Institute of Molecular Biology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

strongly correlated with its chromosomal location. *var* genes residing in internal domains are associated with the upsC (5B1-type) promoter. The most telomerically located *var* genes are almost exclusively controlled by a upsB (*var*17-type) promoter. Another subset of subtelomeric *var* genes is flanked by the upsA promoter. However, compared with the upsB type, these telomere-associated *var* genes are located towards the centromere but are transcribed towards the telomere. Episomally located upsB and upsC promoters were able to drive reporter gene expression (Deitsch *et al*., 1999; Voss *et al*., 2000), and various studies have shown that *var* genes can be transcribed from both locations (Fischer *et al*., 1997; Hernandez-Rivas *et al*., 1997; Scherf *et al*., 1998; Voss *et al*., 2000). Apparently, every *var* gene represents a single transcriptional unit capable of *in situ* activation. Nevertheless, *var* gene expression occurs in a mutually exclusive manner, i.e. only one *var* gene is actively transcribed while the remaining copies remain silenced (Chen *et al*., 1998b; Scherf *et al*., 1998). Switching in *var* gene expression takes place through *in situ* gene activation (Scherf *et al*., 1998). Mutually exclusive transcription among various expression sites has also been reported to participate in antigenic variation of the variant surface glycoproteins in *Trypanosoma brucei* (reviewed by Borst and Ulbert, 2001). However, the main mechanisms used for antigenic variation in protozoan and bacterial pathogens involve DNA rearrangements such as translocation of silenced genes into active expression sites by duplicative or reciprocal recombination (Borst and Greaves, 1987; Donelson, 1995).

Despite its importance in both parasite survival and virulence, the mechanisms involved in regulation of *var* gene expression remain largely unknown. It was shown by nuclear run-on analysis in monomorphic parasite populations expressing a single PfEMP1 variant that *var* gene transcription is controlled at the transcriptional level (Scherf *et al*., 1998). However, it is still puzzling how parasites activate a single *var* gene while keeping all other copies repressed or silenced. *var* gene promoters maintained on episomes were active irrespective of the state of the endogenous chromosomal promoter (Deitsch *et al*., 1999). In addition, Deitsch *et al*. (2001a) reported that silencing of episomal *var* gene promoter-driven transcription involved the co-operative interaction between the intron and 5′ flanking sequence of *var* genes. Complete silencing, however, was only achieved after plasmid transition through S-phase. Together, these results clearly suggest the involvement of epigenetic mechanisms in regulation of *var* gene transcription.

We were interested in identifying structural and functional elements in different *var* gene promoter types and to gain insight into regulation of *var* gene transcription. We show that repression of subtelomeric and chromosome-central *var* gene transcription during intraerythrocytic development occurs 4–8 h apart. This shifted repression sharply coincided with the expression of two distinct DNA-binding activities specifically interacting with sequence recognition motifs unique to either subtelomeric or central *var* gene promoters. Transient transfection experiments indicated a direct participation of these DNA-binding activities in *var* gene repression and also revealed regions with regulatory activities implicated in transcriptional activation. A third nuclear protein specifically expressed during S-phase interacted with a reiterated sequence motif in subtelomeric promoters. The likely involvement of this activity in silencing of subtelomeric *var* gene transcription is discussed.

Results

EMSA screening of subtelomeric and chromosome-central var gene *promoters*

To investigate *var* gene promoters for the presence of elements possibly involved in the regulation of *var* gene transcription, we tested a series of restriction fragments derived from the promoters of 3D7 *var* genes 4A3 (upsB) and 5B1 (upsC) (Voss *et al*., 2000) in gel retardation assays (Fig. 1). Initially, most of these fragments appeared to interact with nuclear proteins. However, it became evident that there was interference of a non-specific ssDNA-binding protein, which we identified as *P. falciparum* replication protein A (RPA) (Voss *et al*., 2002), with labelled single-stranded (ss) probes. Inclusion of unlabelled ss competitor DNA prevented the formation of these complexes. Thereafter, we detected three nuclear activities specifically binding to *var* gene promoter fragments 5B1s(4), 4A3s(1a) and 4A3s(8) (Figs 2 and 3). Cytosolic control extracts did not contain activities binding to any of these probes. Competition electrophoretic mobility shift assays (EMSAs) clearly revealed the specificity of these interactions. We observed two closely co-migrating complexes for 4A3s(8) and, in the case of 4A3s(1a) and 5B1s(4), even multiple shifted bands were apparent. In competition studies, the intensities of these additional bands decreased and increased concordant to the main complex, indicating partial dissociation of multisubunit protein complexes and/or partial proteolysis of the binding factors, rather than binding of different proteins to the same recognition sequence.

Specificity of var *promoter-binding activities*

To fine map the identified sequence elements, we carried out competition studies using overlapping double-stranded (ds) oligonucleotides. The complex formed with 5B1s(4) was not competed by 5B1s(4)-1, -2 and -4

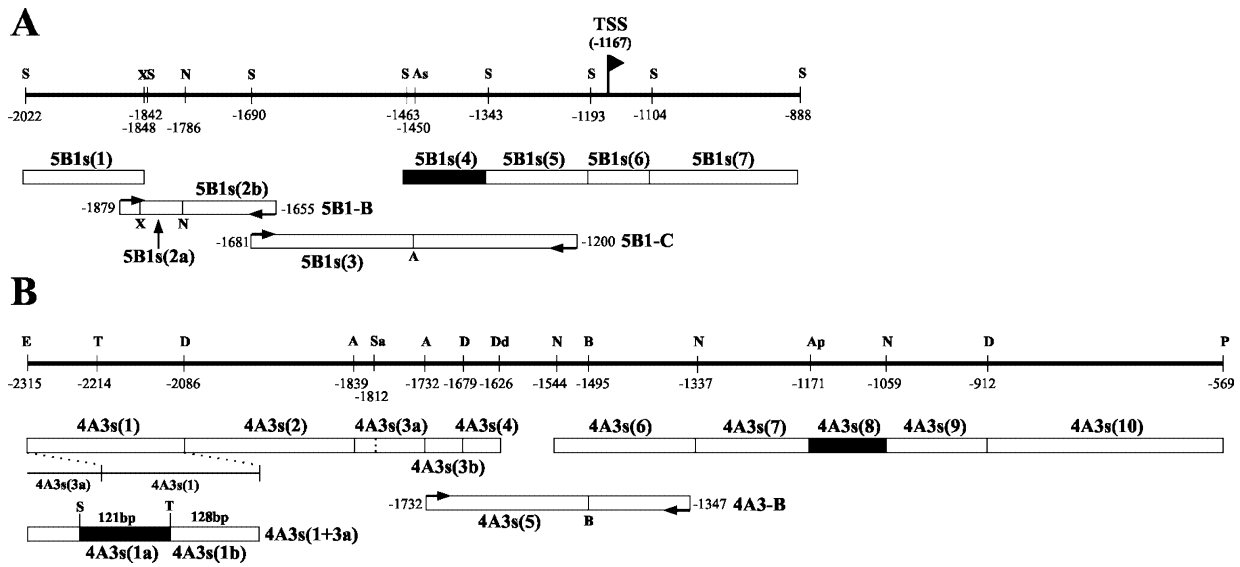
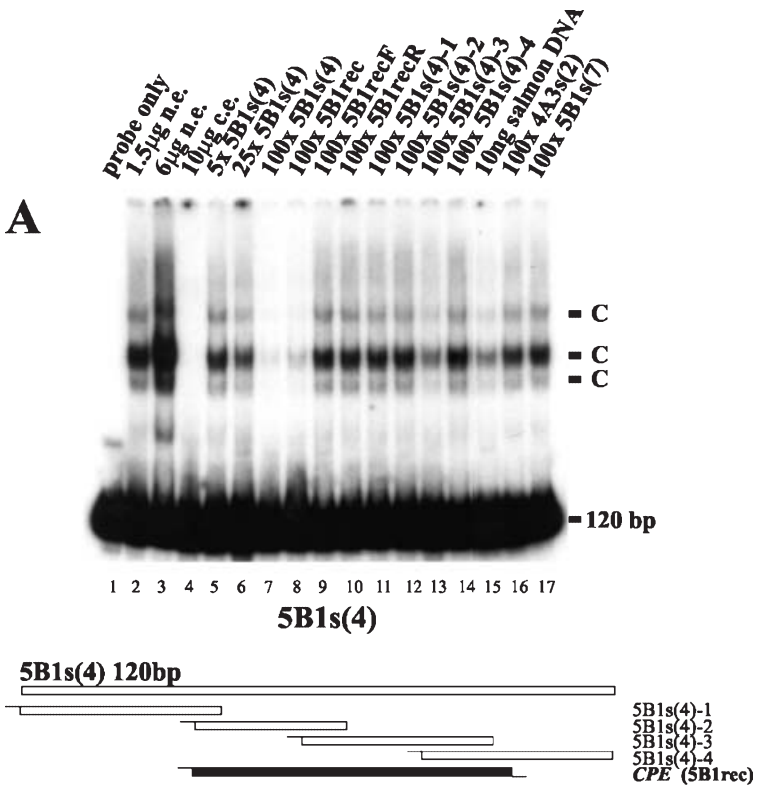


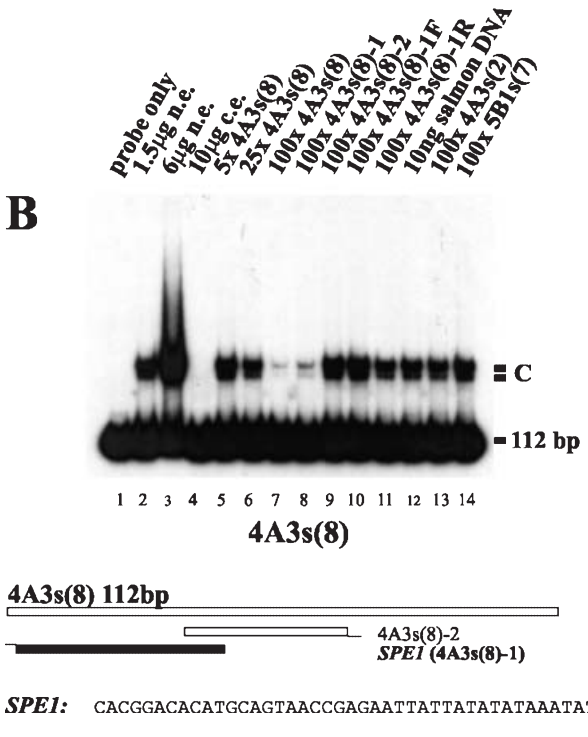
Fig. 1. *var* gene promoter fragments analysed in EMSAs. Fragments bound by nuclear proteins are highlighted in black.
A. Restriction map of the chromosome-central *var* gene 5B1 promoter (upsC) with positions relative to the initiation ATG. The flag depicts the position of the transcriptional start site (TSS) of a central *var* gene (*var7b*) (Deitsch *et al.*, 1999).
B. Restriction map of the subtelomeric *var* gene 4A3 promoter (upsB). Fragment 4A3s(1+3a) was an accidental hybrid molecule consisting of 4A3s(3a) ligated in reverse orientation to the 5′ end of 4A3s(1).
Horizontal arrows indicate PCR primers used to generate fragments not present in *var* promoter restriction libraries. A, *Ava*II; Ap, *Apa*LI; As, *Ase*I; B, *Bgl*II; Dd, *Dde*I; D, *Dra*I; E, *Eco*RV; N, *Nde*I; P, *Pac*I; S, *Ssp*I; Sa, *Sau*3AI; T, *Tsp*509I; X, *Xba*I.

(Fig. 2A). Efficient competition occurred only by a 100-fold molar excess of 5B1rec (69 bp). Neither the sense nor the antisense strand of 5B1rec had an effect on complex formation. The addition of 5B1s(4)-3 caused a reduction in signal intensity, indicating that a sequence motif contained within 5B1s(4)-3 alone is not sufficient for binding, but forms part of a recognition element encoded by 5B1rec, which we termed *CPE* (chromosome-central *var* gene promoter element). *CPE* is located at position −1426 to −1357 with respect to the initiation ATG of 5B1 and contains a 4 bp inverted repeat also present in 5B1s(4)-3. We identified *SPE1* (subtelomeric *var* gene promoter element 1) (Fig. 2B) in the promoter of 4A3 encompassing nucleotides −1171 to −1127 by competition of the nuclear activity binding to 4A3s(8) with 4A3s(8)-1. In contrast, neither 4A3s(8)-2 nor the sense and antisense strands of 4A3s(8)-1 interfered with binding.

The specific protein interaction involving 4A3s(1a) was efficiently competed by 4A3s(1b) (data not shown). As expected, a DNA–protein complex formed with 4A3s(1b) could be prevented by the addition of 4A3s(1a) (Fig. 3B). Although both probes were of similar size, the major complexes (C2, C3) formed with 4A3s(1b) showed a clearly reduced mobility compared with complex C1 formed with 4A3s(1a). These observations suggested an interaction of the same nuclear factor with identical or related sequence motifs in both probes, and the binding of multiple proteins to 4A3s(1b). To investigate this possibility in more detail, we performed EMSA competition studies using overlapping ds oligonucleotides covering the entire 4A3s(1) locus. 4A3s(1)-1, -2 and -3 were unable to inhibit complex formation, whereas competitors 4A3s(1)-1.2, -c1, -c2 and -c3 completely prevented binding to both probes (Fig. 3A–C). Sequence comparison revealed the presence of a conserved sequence motif (*SPE2*) in all four competitors consisting of a direct (T/G)GTGC(A/G) repeat spaced by four nucleotides (Fig. 3D). In 4A3s(1)-1, all but the last two nucleotides of this motif are present in ds conformation (see Table 1). Using Klenow enzyme, we generated the complete recognition motif in 4A3s(1)-1 by refilling the 5′ overhang. This procedure transformed 4A3s(1)-1, but not 4A3s(1)-2, into an efficient competitor (Fig. 3E). The introduction of two point mutations in either the first (repmotM1) or the second (repmotM2) repeat of the *SPE2* consensus sequence completely eliminated the ability to prevent protein binding (Fig. 3F). A ds oligonucleotide coding for a degenerate rep20 tandem repeat (Åslund *et al.*, 1985) also failed to compete (Fig. 3B). Together, these findings revealed that the *SPE2*–protein interaction occurred in a highly specific manner. The presence of three binding sites in the 4A3s(1b) probe greatly facilitated the binding of additional factors (complexes C2 and C3), and we hardly observed 4A3s(1b) complexes with an electrophoretic mobility corresponding to the binding of one protein (C1), despite the fact that the majority of radiolabelled 4A3s(1b) occurred as free probe. Strikingly,



CPE: TTATTTTATTTTTTCTATTTTTTTTTTTTTTTTTTCTTTT
G**ATGTTGTACAT**ATATATTATATATAATA



SPE1: CACGGACACATGCAGTAACCGAGAATTATTATATATAAATAT

Fig. 2. Competitive gel retardation assays and identification of *var* promoter elements *SPE1* and *CPE*. A. Fragment 5B1s(4) formed one major and two minor complexes after incubation with nuclear extracts (lanes 2 and 3) but not with cytoplasmic extracts (lane 4). Complex formation in the presence of various specific and non-specific competitors was analysed in lanes 5–17. A scheme depicting overlapping competitor DNAs covering 5B1s(4) and the nucleotide sequence of *CPE* is shown. The inverted 4 bp repeat separated by three nucleotides is highlighted in bold.
B. Fragment 4A3s(8) formed one major and one closely co-migrating minor complex after incubation with nuclear extracts (lanes 2 and 3) but not with cytoplasmic extracts (lane 4). Competition experiments were analysed in lanes 5–14. A scheme depicting overlapping competitor DNAs covering 4A3s(8) and the nucleotide sequence of *SPE1* is shown.
The molar excess of competitors is indicated above each lane. Sheared salmon sperm DNA was added at a 250-fold weight excess. ss oligonucleotide competitors are denoted by F (forward) or R (reverse). n.e., nuclear extract; c.e., cytosolic extract.

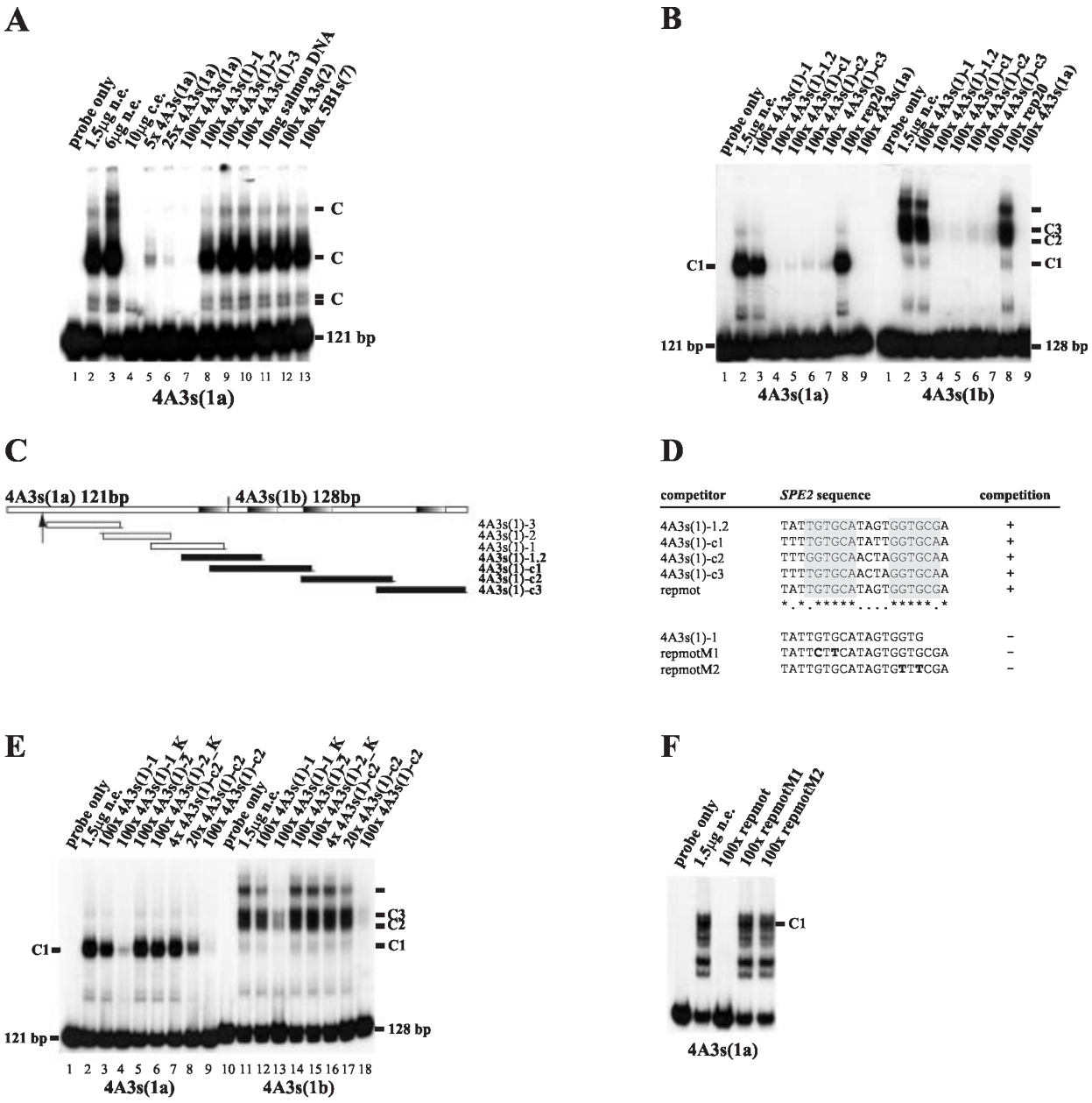


Fig. 3. Interaction of SPE2BP with *SPE2* sequence elements.
A. One major and three minor complexes were formed with fragment 4A3s(1a) after incubation with nuclear extracts (lanes 2 and 3) but not with cytoplasmic extracts (lane 4). In lanes 5–13, the effect of various competitors on binding was analysed.
B. Radiolabelled 4A3s(1a) and 4A3s(1b) were incubated with nuclear extracts (lanes 2). The complexes formed with 4A3s(1b) (C2 and C3) showed reduced mobility compared with C1 formed with 4A3s(1a). Various competitors were analysed in lanes 3–9. The rep20 competitor consisted of a degenerate rep20 tandem repeat. 4A3s(1a) was used as a control for competition experiments (lanes 9).
C. Competitors 4A3s(1)-1.2, -c1, -c2, -c3 (black) prevented complex formation with radiolabelled probes 4A3s(1a) and 4A3s(1b). Grey shaded rectangles indicate *SPE2* consensus sequences at their relative positions in 4A3s(1). The arrow marks the boundary to 4A3s(3a) in the hybrid molecule 4a3s(1a) (see Fig. 1).
D. The *SPE2* sequences present in various competitors are shaded. Point mutations introduced into repmotM1 and repmotM2 are highlighted in bold.
E. The effect of filling in the 5′ overhangs of 4A3s(1)-1 or 4A3s(1)-2 on *SPE2* binding. Klenow-treated ds oligonucleotides 4A3s(1)-1°K or 4A3s(1)-2°K in competition with 4A3s(1a) were analysed in lanes 4 and 6 respectively. Untreated competitors were used in samples 3 and 5. Lanes 12–15 show the effect of the Klenow-treated competitors on binding to radiolabelled 4A3s(1b). Reduction in complex formation through increasing amounts of 4A3s(1)-c2 was analysed in lanes 7–9 and 16–18.
F. The effect of point mutations in either the first (repmotM1) or the second (repmotM2) repeat present in the *SPE2* consensus sequence (repmot) on protein binding. The faster migrating complexes are most likely caused by proteolysis.

Table 1. PCR primers and oligonucleotides used in EMSA gel retardation assays.

Transfection constructs and PCR fragments	Oligonucleotide name	Used in the generation of	Oligonucleotide sequence
4A3 promoter sequence	4A3-28R	All pCAT4A3 derivatives	CTGTATTACATCAGTGCTTGCTATTTGTTTT**CCTAGG**GCGC
	4A3-2202F	pCAT4A3-2202	GCAT**AAGCTT**GCAACTAGGTGCAACATTTTAC
	4A3-1732F	pCAT4A3-1732 PCR 4A3-B	GCAT**AAGCTT**GGATCTAGTCTTTAGGGTTCCCAT
	4A3-1517F	pCAT4A3-1517	GCAT**AAGCTT**GGATCTATAGCTACTATATAAAGATCTG
	4A3-1373F	pCAT4A3-1373	GCAT**AAGCTT**GTTAAAGAACATATCTGTTCATCAAGGT
	4A3-1210F	pCAT4A3-1210	GCAT**AAGCTT**AAAAATCGAAATGGAAGATAC
	4A3-1114F	pCAT4A3-1114	GCAT**AAGCTT**GGGTTTAGGAATACGTATGCCTTTATG
	4A3-1066F	pCAT4A3-1066	GCAT**AAGCTT**GAGAGAATACATATGCTCTTG
	4A3-804F	pCAT4A3-804	GCAT**AAGCTT**CCTTATGCTACATGATATGTCATA
	4A3-515F	pCAT4A3-515	GCAT**AAGCTT**GAAACATGTATGTTTTTATATGTATGT
	4A3-2621F	pCAT4A3-Δ	TCCT**AAGCTT**ACTAATTTATGTCCTATAGGTACG
	4A3-1180R	pCAT4A3-Δ	CGAT**AAGCTT**TTTTGGGTATCTTCCATTTCG
	4A3-1347R	PCR 4A3-B	CAATTTCTTGTATAGACAAGTAGTTTCC
	4A3-*Hpa*I_R	pVLUC4A3/-Δ	CAGTGCTTGCTATTTGTTTT**CAATTG**CG
5B1 promoter sequence	5B1EABR	All pCAT5B1 derivatives	CTATTG**TTCGAA**ATACTTCGCGTATAATCTCATTATTACACGTACG
	5B1-1879F	pCAT5B1-1879 PCR 5B1-B	GCAT**AAGCTT**GCATCCTATATGTATATTATATACATTCTCTC
	5B1-1681F	pCAT5B1-1681 PCR 5B1-C	GCAT**AAGCTT**CATAATTTCATCATTATTAAAGTAGAGAAA
	5B1-1508F	pCAT5B1-1508	GCAT**AAGCTT**GAAGTATGTATACAAAATAGATG
	5B1-1381F	pCAT5B1-1381	GCAT**AAGCTT**GTACATATATATTATATATAATAC
	5B1-1325F	pCAT5B1-1325	GCAT**AAGCTT**ATATAACAAAAAAAAATTAATATG
	5B1-1228F	pCAT5B1-1228	GCAT**AAGCTT**CATAGAAATGTGGTAGATAATATAGATAGA
	5B1-1086F	pCAT5B1-1086	GCAT**AAGCTT**TATATATGTTCTTCGATTATAATTCTTTT
	5B1-2522F	pCAT5B1-Δ	GCAT**AAGCTT**CCTTATGCTACATGATATGTCATA
	5B1-1483R	pCAT5B1-Δ	CGAT**AAGCTT**CATCTATTTTGTATACATAATTC
	5B1-1655R	PCR 5B1-B	GTATTAAAGTAGTAATAATTTCATCTC
	5B1-1200R	PCR 5B1-C	GTATCTTTACACCATCTATTATATCTATC
	5B1-*Hpa*I_R	pVLUC5B1/-Δ	CACATACATATATCCATATACA**CAATTG**GC
	T7-*Kpn*I	pVLUC5B1/-Δ pVLUC4A3/-Δ	GG**GGTACC**TAATACGACTCACTATAGGG

EMSA competitors and probes	Oligonucleotide name	Competitor	Oligonucleotide sequence
	5B1s(4)-1F	5B1s(4)-1	ATGTTTTTTTTATTAATATAATAATCCTTTTTTTTATGTTATTTTA
	5B1s(4)-1R		AAAAAAATAATTATATTATTAGGAAAAAAATACAATAAAAT
	5B1s(4)-2F	5B1s(4)-2	ATGTTATTTTATTTTTTCTATTTTTTTTTTTTTT
	5B1s(4)-2R		AATAAAATAAAAAAGATAAAAAAAAAAAAAA
	5B1s(4)-3F	5B1s(4)-3	ATGTTTTTTTTTTTTTTTTTCTTTTGATGTTGTACATATATAT
	5B1s(4)-3R		AAAAAAAAAAAAAAAAGAAAACTACAACATGTATATATA
	5B1s(4)-4F	5B1s(4)-4	ATGTACATATATATTATATATAATACATTATATATTATAAT
	5B1s(4)-4R		ATGTATATATAATATATATTATGTAATATATAATATTA
	5B1recF	5B1rec	TGTTATTTTATTTTTTCTATTTTTTTTTTTTTTTTTTCTTTTGATGT TGTACATATATATT
	5B1recR		ATATATAATAAATAAAATAAAAAAGATAAAAAAAAAAAAAAAAAAG AAAACTACAACATGTATATATAATATATATTATGTA
	4A3s(1)-1F	4A3s(1)-1	CACATTTTTTTTGGTGCGACTTTATTGTGCATAGTGGTG
	4A3s(1)-1R		GTGTAAAAAAAACCACGCTGAAATAACACGTATCACCACGCT
	4A3s(1)-1.2F	4A3s(1)-1.2	CGACTTTATTGTGCATAGTGGTGCGAATTTATACTTTGGTGCA
	4A3s(1)-1.2R		AGTTGCACCAAAGTATAAATTCGCACCACTATGCACAATAAAG TCG
	4A3s(1)-2F	4A3s(1)-2	TAGTGATACCACACATGTGGGAAGACCACACATTTTTTT
	4A3s(1)-2R		ACTATGGTGTGTACACCCTTCTGGTGTGTAAAAAAA
	4A3s(1)-3F	4A3s(1)-3	ATCACACATACGTGGTAATACCACATATATAGTGATACCAC
	4A3s(1)-3R		TAGTGTGTATGCACCATTATGGTGTATATATCACTATGGTGTG
	4A3s(1)-c1F	4A3s(1)-c1	CATAGTGGTGCGAATTTATACTTTGGTGCAACTAGGTGCAACAT TTTACTTTTG
	4A3s(1)-c1R		GTATCACCACGCTTAAATATGAAACCACGTTGATCCACGTTG TAAAATGAAAACACGT
	4A3s(1)-c2F	4A3s(1)-c2	CTTTTGTGCAACTAGGTGCAACTTGATAAACACTGCGATGGAG GTGC

Table 1. *cont.*

EMSA competitors and probes	Oligonucleotide name	Competitor	Oligonucleotide sequence
	4A3s(1)-c2R		AA GAAAACACGTTGATCCACGTTGAACTATTTGTGACGCTAC CTCCACGTTGTA
	4A3s(1)-c3F	4A3s(1)-c3	GGTGCAACATTTTACTTTTGTGCATATTGGTGCAACATTT TACTTTGGTGCA
	4A3s(1)-c3R		CCACGTTGTAAAATGAAAACACGTATAACCACGTTGTAAAAT GAAACCACGTTGA
	repmotF	repmot	CTTTATTGTGCATAGTGGTGCGAATTTATA
	repmotR		GAAATAACACGTATCACCACGCTTAAATAT
	repmotM1F	repmotM1	CTTTATT***C***T***T***CATAGTGGTGCGAATTTATA
	repmotM1R		GAAATAA***G***A***A***GTATCACCACGCTTAAATAT
	repmotM2F	repmotM2	CTTTATTGTGCATAGTG***TT T***CGAATTTATA
	repmotM2R		GAAATAACACGTATCAC***A***AA***A***GCTTAAATAT
	4A3s(8)-1F	4A3s(8)-1	TGCACGGACACATGCAGTAACCGAGAATTATTATATATAAATAT
	4A3s(8)-1R		GTGCCTGTGTACGTCATTGGCTCTTAATAATATATATTTATA
	4A3s(8)-2F	4A3s(8)-2	ATATAAATATATATATGTATATTTTGGGTTTAGGAAT
	4A3s(8)-2R		TATATTTATATATATACATATAAAACCCAAATCCTTATGC

Forward (F) oligonucleotides are indicated 5′ → 3′, reverse (R) oligonucleotides 3′ → 5′. Restriction enzyme recognition sites are highlighted in bold. Point mutations introduced into repmotM1 and remotM2 are italicized and in bold.

whereas a 20-fold molar excess of 4A3s(1)-c2 significantly reduced complex formation with 4A3s(1a) (which harbours only one binding site) no effect was observed with 4A3s(1b) (Fig. 3E). We therefore conclude that binding of this factor to *SPE2* elements occurs in a co-operative manner.

To test for cross-reactivity, we used oligonucleotides 5B1rec, 4A3s(8)-1 and repmot, harbouring the recognition motifs *CPE*, *SPE1* and *SPE2*, respectively, as competitors. Figure 4 shows that the three DNA-binding factors represent distinct activities and that central and subtelomeric *var* gene promoters interact with different nuclear proteins. Multiple sequence alignments of *var* gene 5′ flanking sequences retrieved from the NCBI Malaria Genetics and Genomics section (http://www.ncbi.nlm.nih.gov/Malaria/plasmodiumbl.html) showed that the *CPE* element was highly conserved among all seven upsC sequences investigated. Variation was observed in six nucleotide positions and in the length of the poly dT stretches only, whereas the 4 bp inverted repeat was perfectly conserved. Analysis of 12 upsB promoters revealed a perfect conservation of the *SPE1* recognition motif except for two nucleotide positions. *SPE2* elements were present in all upsB sequences analysed and occurred in arrays of 5–18 repeats. The position of *SPE1*, *SPE2* and *CPE* with respect to the ATG start codon of *var* genes is also conserved (data not shown).

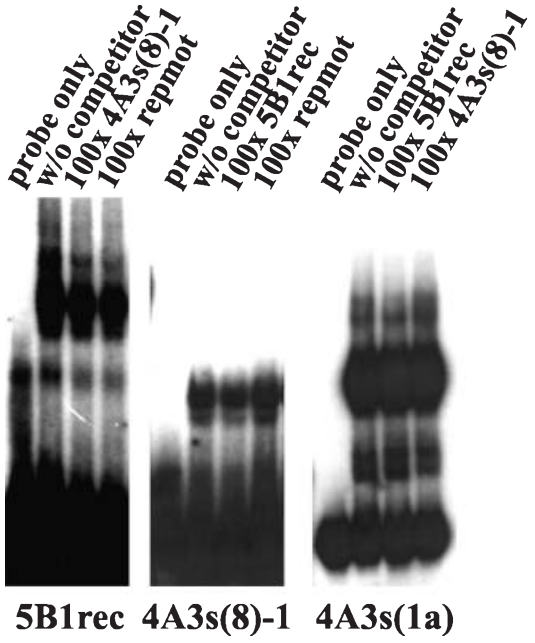


Fig. 4. Cross-competition EMSAs. EMSA showing the absence of cross-competition between *CPE*, *SPE1* and *SPE2* and their cognate binding factors. n.e., nuclear extract.

Stage-specific expression of CPE-*,* SPE1- *and* SPE2-*binding proteins*

We investigated the expression of the *CPE*-, *SPE1*- and *SPE2*-binding activities (CPEBP, SPE1BP and SPE2BP) across the intraerythrocytic parasite cell cycle (Fig. 5). To standardize this comparative analysis, we used stage-specific protein preparations derived from an equal number of nuclei. SPE1BP first appeared in parasites 16–26 h post invasion (h.p.i.) and increased to maximal levels in parasites older than 24 h.p.i. CPEBP was not detected until 24–34 h.p.i. and was most pronounced in schizonts. SPE2BP, however, was exclusively expressed in late-stage parasites older than 34 h.p.i. The expression of SPE1BP and CPEBP in trophozoites and schizonts suggested a possible involve-

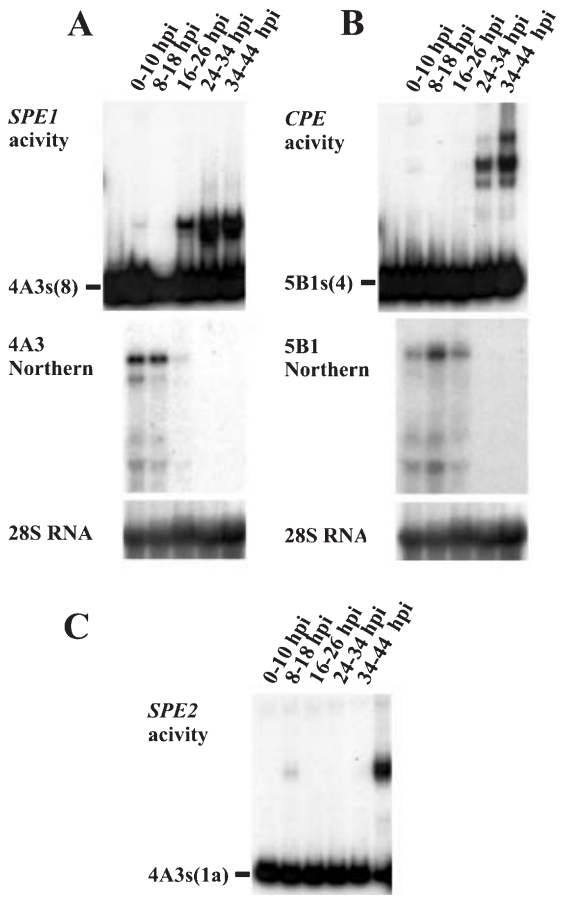


Fig. 5. Stage-specific expression of SPE1BP, CPEBP and SPE2BP in relation to *var* gene transcription. Intraerythrocytic stages are indicated above each lane in hours post invasion (h.p.i.).
A. Top: *SPE1*-binding activity in nuclear extracts derived from synchronized parasite cultures. Middle: upsB-type *var* gene transcription monitored by Northern analysis of total RNA obtained from the same stages. The filter was probed with a radiolabelled PCR fragment encompassing nucleotides –460 to –1 of the 4A3 5′ UTR. Bottom: stage-specific 28S Northern as a control for equal RNA loading.
B. Top: *CPE*-binding activity in nuclear extracts derived from synchronized parasite cultures. Middle: stage-specific Northern analysis of upsC-type *var* gene transcription. The filter was probed with a radiolabelled PCR fragment encompassing nucleotides –514 to – 237 of the 5B1 5′ UTR.
C. Stage specificity of the *SPE2*-binding activity.

ment in *var* gene repression as *var* gene transcription was shown to occur in parasites up to 18–27 h.p.i. only (Kyes *et al*., 2000). We therefore compared stage-specific transcription of upsB- and upsC-type *var* genes with expression of SPE1BP and CPEBP. This was realized by Northern analysis using total RNA isolated from the same stages as for isolation of nuclear proteins. Probes from the conserved 5′ untranslated regions (UTR) of *var* genes 4A3 or 5B1, identifying upsB- and upsC-type *var* transcripts, respectively, clearly showed that transcription of subtelomeric *var* genes is turned off earlier than that of central *var* genes. Concordantly, the shifted cessation of subtelomeric and central *var* gene transcription exactly correlated with the delayed occurrence of CPEBP over SPE1BP, indicating their involvement in repression of *var* gene transcription.

var *gene promoter activity analysis by transient transfection*

We showed recently that both types of *var* gene promoters were active in transient transfection assays (Voss *et al*., 2000). As an alternative approach to identify functional *var* gene promoter regions, we generated nested 5′ deletions in the 5B1 and 4A3 promoters. These fragments were used to replace the full-length promoters in pCAT5B1 and pCAT4A3 respectively. The data presented in Fig. 6A reveal two regions in the 5B1 promoter dominantly affecting promoter activity. Deletion of nucleotides –2522 to –1879 reduced activity to 26%, indicating the presence of *cis*-acting elements involved in the activation of chromosome-central *var* gene transcription. Further nucleotide removal to –1508 had no dramatic effect. Deletion of the sequence from –1508 to –1325, harbouring the entire *CPE* element, caused an additional drop in promoter activity to barely detectable levels (6% compared with pCAT5B1). This region may therefore comprise additional positive regulatory elements and the core promoter. Deletion to –1228 and –1086 completely abolished promoter activity.

In the 4A3 promoter, deletion of the most 5′ 419 bp had no effect on overall promoter activity (Fig. 6B). However, removal of nucleotides –2202 to –1732 (containing the *SPE2* motifs) resulted in a 2.2-fold increase. Further deletion to –1210 again caused a stepwise reduction to the value observed with the full-length constructs. 5′ removal of the *SPE1* motif in pCAT4A3-1114 and pCAT4A3-1066 caused no effect. The large standard deviation in the pCAT4A3-804 sample resulted from an extraordinarily high value obtained in one experiment, which most probably represented an artifact. The average relative activity for this construct obtained in the other three experiments, however, was 28%. Further deletion to –514 completely abolished promoter activity. We therefore conclude that the region encompassing –1066 to –514 contains functional *cis*-acting elements involved in activation and basal transcription of subtelomeric *var* genes.

To investigate any possible effect of *CPE* or *SPE1* on *var* gene repression, we transiently transfected synchronized parasites with pVLUC5B1-Δ and pVLUC4A3-Δ carrying internal deletions encompassing *CPE* and *SPE1* respectively. We used the luciferase gene for these experiments as the long half-life of CAT in eukaryotic cells (50 h) (Thompson *et al*., 1991) makes this reporter unsuitable for the analysis of stage-specific pro-

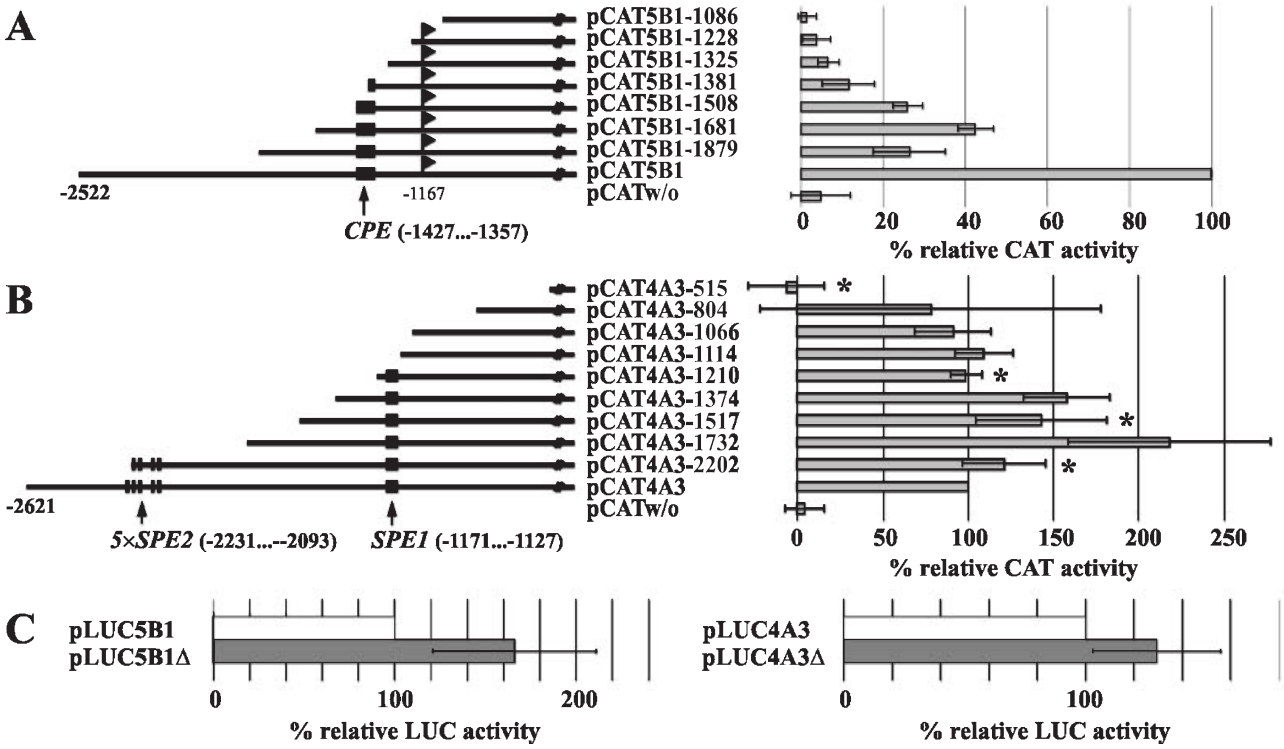


Fig. 6. Analysis of 5B1 and 4A3 *var* gene promoter activities by transient transfection. For transfections using CAT constructs, pCAM5/3 was used as a positive control (Crabb and Cowman, 1996; data not shown). pCATw/o, lacking a promoter, was used as a negative control (Voss *et al.*, 2000). The length of the promoter in each construct is depicted on the left. Activity values are indicated as a percentage of relative CAT or luciferase activity compared with the activity obtained with parental constructs. Experiments have been repeated four times.

A. Nested 5′ deletion analysis of the 5B1 promoter (upsC). The location of *CPE* is indicated by a filled rectangle. The putative TSS is depicted by a flag.

B. 5′ nested 4A3 promoter (upsB) deletion analysis. The positions of *SPE1* and the repeated *SPE2* elements are indicated by rectangles. Experiments have been repeated four times, and CAT activity has been measured three times in each experiment. Constructs marked by an asterisk have been transfected three times only.

C. Effect of internal promoter deletions on promoter activity. In pVLUC5B1-Δ, nucleotides −1483 to −1325 were deleted from the 5B1 promoter including *CPE* (–1427 to −1357). Nucleotides −1180 to −1073 [including *SPE1* (–1171 to −1127)] were deleted from the 4A3 promoter in pVLUC4A3-Δ. Control transfections without plasmid DNA yielded activity values between −14 and 0.

moter activity. Deletion of nucleotides −1483 to −1325 (pVLUC5B1-Δ) resulted in a 1.65-fold average increase in luciferase activity compared with that obtained with the parental plasmid pVLUC5B1. Deletion of *SPE1* in pVLUC4A3-Δ led to a 1.3–fold increase compared with pVLUC4A3 (Fig. 6C). These results indicate an involvement of the interactions of CPEBP and SPE1BP with their cognate regulatory binding motifs in stage-specific repression of chromosome-central and subtelomeric *var* gene expression respectively.

Discussion

Plasmodium falciparum var gene expression is controlled by different yet conserved promoters, and these structural differences correlate with chromosomal location and orientation of *var* genes (Voss *et al.*, 2000; Gardner *et al.*, 2002). Recent findings suggest that *cis*-acting elements mediating *var* gene expression and silencing must be conserved within every *var* gene promoter and that regulation of *var* gene expression is subject to higher level silencing mechanisms.

We identified three highly conserved *var* gene promoter elements interacting with distinct DNA-binding proteins. The *SPE1* and *SPE2* motifs are specifically associated with subtelomeric upsB promoters, and *CPE* is a characteristic sequence element in chromosome-internal promoters (upsC). Complex formation with these elements occurred in a highly sequence-specific manner and involved three different nuclear activities. Genome-wide BLAST analysis (http://www.ncbi.nlm.nih.gov/Malaria) revealed that these elements are exclusively associated with *var* genes. *SPE1*, *SPE2* and *CPE* have no significant similarities to eukaryotic transcription factor binding sites reported in the TRANSFAC database (http://www.transfac.gbf.de). However, the organization of

the *SPE2* motif as two direct 6 bp repeats spaced by four nucleotides is reminiscent of nuclear hormone receptor (NHR) response elements (reviewed by Aranda and Pascual, 2001). Whether this is of any significance for the nature of the *SPE2*-binding protein remains to be elucidated, most of all because NHRs have as yet only been described in metazoans.

var genes are transcribed in parasites up to 18–27 h.p.i. (Kyes *et al*., 2000). Here, we show for the first time that subtelomeric *var* messages were observed in parasites up to 18 h.p.i. only, whereas chromosome-internal *var* transcripts were still detected in parasites 16–26 h.p.i. This was in perfect correlation with the occurrence of SPE1BP 18 h.p.i. and the 4–8 h delayed expression of CPEBP. These results showed that the structural differences in *var* gene promoters reflect functional and regulatory differences in transcriptional repression of *var* genes located in different chromosomal domains. They further suggested that *SPE1* and its cognate binding factor SPE1BP are directly involved in repression of subtelomeric *var* genes, whereas chromosome-internal *var* gene transcription is repressed 4–8 h later by the interaction of CPEBP with *CPE*.

In transient transfection experiments, internal deletion of stretches harbouring *SPE1* or *CPE* from the subtelomeric or chromosome-internal promoter led to an increase in luciferase activity compared with the wild-type promoters (1.65-fold for *CPE* and 1.3-fold for *SPE1*). These results clearly hint at an involvement of these regulatory elements in *var* gene repression. However, the absence of a pronounced difference in activity between wild-type and mutant promoters indicates that more complex processes are involved. On one hand, and as suggested by others (Scherf *et al*., 1998; Deitsch *et al*., 1999), epigenetic mechanisms may be involved in the regulation of *var* gene expression. When working with episomes, such effects may impede the investigation of promoter regulatory processes, as the chromatin configuration of episomal and chromosomal DNA may differ substantially. For example, it has been shown that episomal *var* gene promoters are desilenced irrespective of the transcriptional state of the endogenous promoter (Deitsch *et al*., 1999). Similar findings have been obtained for the *P. falciparum* gbp130 promoter, in which the loss of developmental restriction of the episomal promoter was attributed to observed differences in chromatin structure between episomal and chromosomal copies (Horrocks and Lanzer, 1999). Furthermore, even if replicated plasmids were properly assembled into chromatin during S-phase, rapid loss of episomes as a result of inefficient segregation (van Dijk *et al*., 1997; O'Donnell *et al*., 2001) and continuous uptake of naked plasmid DNA into parasite nuclei during intracellular growth (Deitsch *et al*., 2001b) would mask the repressive effects of *SPE1* or *CPE*. On the other hand, additional *trans*-acting factors may participate in *var* repression and, thus, total inhibition of repression would only be observed after preventing these interactions as well. Studies using transgenic parasites carrying mutations in *SPE1* and *CPE* in chromosomal *var* promoters are definitely needed to explore further their effect on *var* gene repression.

We identified several regions involved in activation of *var* gene transcription in both promoters. 5′ nested removal of a fragment containing the *CPE* motif from the 5B1 promoter resulted in a decrease in activity to barely detectable levels, suggesting the presence of positive *cis*-acting element(s) and the core promoter within the deleted region. It is therefore not surprising that no repressive effect was observed upon nested deletion of *CPE*. A second region displaying strong activating properties is positioned further upstream (−2522 to −1879) as deletion of this stretch retained only 26% activity. 5′ deletion to −1228 and −1086 (including the putative TSS at −1167) led to a complete abolishment of promoter activity and suggests that the TSS site mapped for *var7b* (Deitsch *et al*., 1999) is conserved among chromosome-central *var* genes. Deletion analysis of the subtelomeric 4A3 promoter revealed a functional region with activation potential between nucleotides −1066 and −504. Our results also indicate that the TSS of subtelomeric *var* genes is located between nucleotides −804 and −514. We were not able to detect any DNA–protein interactions in the fragments participating in transcriptional activation of *var* genes. This does not imply, however, that such interactions do not exist. Factors involved in transcriptional activation of a single *var* gene are likely to be of low abundance, aggravating their detection in standard gel retardation experiments. Furthermore, we cannot exclude the possibility that the binding conditions used in our experiments were inappropriate for stable complex formation involving other sequences and nuclear proteins.

The interaction of SPE2BP with *SPE2* occurred in a highly sequence-specific manner. By sequence comparison, we found that *SPE2* elements occur in arrays of 5–18 repeats in upsB promoters, and these arrays are located close to the rep20 repeat region. We showed that the presence of multiple reiterated *SPE2* motifs facilitated co-operative protein binding. Degenerate rep20 tandem repeats did not interfere with *SPE2* binding, confirming that *SPE2* arrays represent a defined subtelomeric *var* gene promoter element providing multiple binding sites for a distinct nuclear protein.

We hypothesize that the function of SPE2BP may be to participate in transcriptional silencing of subtelomeric *var* genes. (i) We observed a 2.2-fold increase in promoter activity upon deletion of the *SPE2* repeat array. (ii) It is known that assembly of silent chromatin requires transi-

tion through S-phase in other organisms (Miller and Nasmyth, 1984; Firestein *et al*., 2000). It was also convincingly shown that silencing of episomal transcription mediated by a *var* gene promoter in *P. falciparum* required plasmid passage through S-phase (Deitsch *et al*., 2001a). SPE2BP was specifically expressed during S-phase and mitosis in parasites 34–44 h.p.i. and, thus, this activity may be involved in establishing the silenced state of subtelomeric *var* genes in newly developing merozoites. (iii) Perinuclear localization of telomeric clusters facilitates telomeric silencing in *Saccharomyces cerevisiae* (Maillet *et al*., 1996; Andrulis *et al*., 1998), and high concentrations of proteins essential for silencing are found in this nuclear compartment (Gotta *et al*., 1996; Maillet *et al*., 1996). As in yeast, clusters of *P. falciparum* telomeres locate to the nuclear periphery (Freitas-Junior *et al*., 2000). Orthologues of most of the yeast proteins involved in telomeric gene silencing (e.g. RAP1, SIR2-4, MLP, Ku-complex) have been identified in the *P. falciparum* genome (Scherf *et al*., 2001), raising the possibility that this parasite uses similar mechanisms to silence genes close to chromosomal ends. In such a model, the *SPE2* elements could act as silencers interacting with regulatory proteins to initiate nucleation and assembly of silenced chromatin spreading throughout the nearby *var* gene. Co-operative binding of the regulatory factors to reiterated *SPE2* motifs could prove essential in efficient co-repressor recruitment and may be important to establish readily and inherit reliably the silent state of subtelomeric *var* genes. A recent study reported that *var* gene silencing involves co-operative interactions between the *var* intron and 5′ flanking region (Deitsch *et al*., 2001a). These experiments have only been done with a chromosome-central *var* promoter, and it remains to be tested whether this is also true for subtelomeric *var* promoters. Considering the structural and functional differences in the conserved *var* promoters and a probable different nuclear location of subtelomeric and central *var* genes, it is conceivable that silencing of subtelomeric *var* genes is mediated by different mechanisms.

Our results shed new light on transcriptional regulation of the *var* gene family and inevitably raise the question about the biological significance of the structural and functional differences in *var* gene promoters. The observed differences might only represent an evolutionary result of the spreading of *var* genes throughout the parasite genome where different chromosomal and nuclear domains may impose different overall regulatory mechanisms. But these findings might also reflect distinct functional *var* gene subsets. Variant-specific antibodies play an important role in protection from clinical disease (Bull *et al*., 1998) and, therefore, antigenic variation is a prerequisite for survival in the human host. We hypothesize that, in naive hosts, expression of chromosome-central *var* genes may prove beneficial for the parasite. Chromosome-central *var* genes reside in stable genomic domains (Pologe and Ravetch, 1988; Lanzer *et al*., 1993) where only little sequence variation over time is expected to occur. Similarly, upsC-type *var* genes may be quite stable over time because of their inverse orientation and centromere-proximal location compared with upsB-type *var* genes. These genes may code for PfEMP1 variants that confer stable and high-affinity binding of iRBCs to selected host receptors, thus efficiently preventing spleen-dependent killing. In contrast, subtelomeric *var* genes located in highly recombinogenic regions and subject to frequent ectopic recombination events (Freitas-Junior *et al*., 2000) may have increased antigenic variability, possibly coupled with decreasing binding affinity. In hyperimmune individuals, expression of subtelomeric *var* genes would allow successive exposure of highly variable PfEMP1 molecules to escape existing anti-PfEMP1 responses. Based on agglutination frequencies of field isolates, Bull *et al.* (2000) similarly proposed the existence of a PfEMP1 subset functionally selected for optimal cytoadherence properties in young children that might play an important role in causing severe disease. In older children, immune selection would be more important, displacing optimally cytoadherent variants by those with novel epitopes. Monitoring of *var* gene expression in the field as well as advances in our understanding of PfEMP1–host receptor interactions will be important in testing this hypothesis.

In conclusion, we show for the first time that transcriptional repression of *var* genes located in different chromosomal domains uses different mechanisms. Interactions of different nuclear proteins with distinct promoter elements unique to either subtelomeric or chromosome-internal *var* gene promoters are involved in these processes. Our findings would support programmed *var* gene regulation and open the way for the identification and functional characterization of molecular components engaged in this process. This will be of great value in our understanding of antigenic variation and PfEMP1-mediated virulence.

Experimental procedures

Parasite cultures

Plasmodium falciparum 3D7 parasites were cultured in 100 mm and 150 mm Petri dishes at 5% haematocrit as described previously (Trager and Jensen, 1978) in RPMI medium supplemented with 0.5% albumax (Gibco BRL). Growth synchronization was achieved by sorbitol lysis (Lambros and Vanderberg, 1979).

Parasite nuclear extracts

Parasites were released from RBCs by saponin lysis. Nuclear

proteins were extracted into high-salt buffer from isolated parasite nuclei as described previously (Voss *et al*., 2002).

Subcloning of var *gene 4A3 and 5B1 5′ flanking regions*

The 4A3 5′ flanking region encompassing nucleotides −2312 to −569 was excised from pCAT4A3 (Voss *et al*., 2000) with *Eco*RV and *Pac*I. After digestion of the purified promoter fragment with *Apa*LI, *Ava*II, *Dde*I, *Dra*I and *Nde*I, subfragments were treated with Klenow enzyme to polish 5′ overhangs. Blunt-ended restriction fragments were cloned into pGEM3-Zf(+) (Promega) in a cycle-restriction ligation as described previously (Push *et al*., 1997). The *Sfc*I–*Hin*dIII fragment of the 5B1 promoter encompassing nucleotides −2142 to −248 was excised from pCAT5B1 (Voss *et al*., 2000) and digested further with *Ssp*I. Restriction fragments were cloned as above. After transformation into *Escherichia coli*, individual clones were identified by sequencing. Using this approach, we obtained plasmid clones carrying subfragments 4A3s(1+3a), (2), (3a), (3b), (4)–(10) and 5B1s(1), (4)–(7) (Fig. 1).

Probes and competitor DNA used in EMSAs

Promoter fragments subcloned into pGEM3-Zf(+) were excised with *Bam*HI and *Sac*I and agarose gel purified. Insert 4A3s(1+3a) was digested further with *Tsp509*I and *Sau*3AI to obtain fragments 4A3s(1a) and 4A3s(1b). Fragment 4A3s(5) was obtained by *Bgl*II digestion of the 4A3-B polymerase chain reaction (PCR) product amplified with primers 4A3-1732F and 4A3-1347R (Table 1). PCR fragments 5B1-B and 5B1-C were amplified with primer pairs 5B1-1879F/5B1-1655R and 5B1-1681F/5B1-1200R respectively. 5B1-B was digested with *Xba*I and *Nde*I to yield fragments 5B1s(2a) and 5B1s(2b); 5B1C was digested with *Ase*I to generate 5B1s(3) (Fig. 1). Fragments were radiolabelled with Klenow enzyme by incubating 1 pmol of DNA at 30°C for 20 min in 1 × React2 buffer (Gibco BRL) in the presence of 10 μCi of [α^{32}-P]-dCTP and 50 μM each dATP/dGTP/dTTP. Labelled fragments were excised from 5% nondenaturing polyacrylamide gels and eluted overnight into buffer E (0.1% SDS, 500 mM NH_4 acetate, 1 mM EDTA), followed by precipitation and resuspension in 1 × React3 buffer (Gibco BRL).

Double-stranded (ds) oligonucleotides used as competitors or probes (Table 1) were obtained by annealing equimolar amounts of complementary oligonucleotides in 1′React3 at 95°C for 5 min followed by slow cooling to room temperature. Annealing was confirmed by 20% PAGE analysis. ds oligonucleotide probes 4A3s(8)-1 and 5B1rec were labelled as above and purified using Sephadex G-25 spin columns (Amersham).

EMSA

Crude nuclear proteins (1.5–3 μg) were incubated with 5 fmol of radiolabelled probe in 20 μl of 1 × EMSA buffer [20 mM Hepes, pH 7.8, 60 mM KCl, 0.5 mM EDTA, 2 mM dithiothreitol (DTT), 2 mM $MgCl_2$, 25 μM $ZnCl_2$, 0.1% Triton X-100, 10% glycerol] containing 2 μg pg poly-(dI–dC) as non-specific competitor DNA for 20 min at room temperature. An aliquot of 200 fmol of single-stranded (ss) oligonucleotide 5B1motF (5′-AGAAATGTGGTAGATAATATAGATAGAAAG-3′) was included in every reaction to prevent the formation of nonspecific ssDNA–protein complexes (Voss *et al*., 2002). Binding reactions were analysed on 5% or 6% polyacrylamide gels in 0.5% TBE. For competition experiments, labelled probes were added 10 min after incubation of protein and competitor DNA.

Transfection constructs

Transfection constructs used in this study were derivatives of pCAT4A3 and pCAT5B1. In these plasmids, *var* gene 5′ flanking regions and the *P. falciparum* calmodulin gene 3′ region control the expression of the chloramphenicol acetyltransferase (*cat*) gene (Voss *et al*., 2000). pCAT4A3 was digested with *Hin*dIII and *Bam*HI, and pCAT5B1 was digested with *Hin*dIII to eliminate the *var* gene promoter. 5′ nested deletions of the 4A3 promoter were established by PCR using the primer 4A3-28R in conjunction with the respective 4A3 forward primers (Table 1). 4A3-28R is positioned at −1 with respect to the initiation ATG and harbours a *Bam*HI site; all forward primers carry a *Hin*dIII recognition sequence. PCR products were cloned into the *Bam*HI–*Hin*dIII backbone of pCAT4A3. 5′ nested deletions of the 5B1 promoter were obtained by PCR using 5B1EABR together with the forward 5B1 primers (Table 1). 5B1EABR is located at position −209 with respect to the initiation ATG of *var* gene 5B1 and includes a wild-type *Hin*dIII site. A *Hin*dIII site is also present in all 5B1 forward primers. PCR products were cloned into the *Hin*dIII backbone of pCAT5B1 in sense orientation.

pVLH/int (kindly provided by K. Deitsch), carrying the *var7b* promoter (Deitsch *et al*., 2001a), was used to generate luciferase constructs. The *var* intron present in pVLH/int was excised with *Bam*HI and *Spe*I, and the plasmid was religated (pVLUC) after polishing with Klenow enzyme. The 5B1 promoter, amplified from pCAT5B1 with primers T7-*Kpn*I_F and 5B1-*Hpa*I_R (−108), was ligated into *Kpn*I–*Hpa*I-digested pVLUC, thus replacing the *var7b* promoter (pVLUC5B1). Likewise, the 4A3 promoter [primers T7-*Kpn*I_F and 4A3-*Hpa*I_R (−1)] was introduced to generate pVLUC4A3. To obtain an internal 114 bp deletion in the 4A3 promoter (encompassing *SPE1*), we cloned the *Hin*dIII-digested PCR fragment generated with primers 4A3-2621F and 4A3-1180R (nucleotides −2621 to −1180) into the *Hin*dIII site of pCAT4A3-1066 (pCAT4A3-Δ). The PCR fragment amplified with primers 5B1-2522F and 5B1-1483R was digested with *Hin*dIII and cloned into the upstream *Hin*dIII site of pCAT5B1-1325 to generate pCAT5B1-Δ, carrying an internal deletion of 158 bp (including *CPE*). The 5B1-Δ and 4A3-Δ promoters were amplified as above and cloned into *Kpn*I–*Hpa*I-digested pVLUC to obtain pVLUC5B1-Δ and pVLUC4A3-Δ respectively.

Isolation of parasite total RNA and Northern analysis

Parasite total RNA was isolated and stored as described previously (Kyes *et al*., 2000). For Northern analysis, equal amounts of RNA extracted from synchronized parasite cul-

tures were electrophoretically separated on a 0.8% agarose gel (5 mM GTC) and vacuum transferred to Hybond-XL membranes (Amersham). Subtelomeric *var* transcripts were detected with a 460 bp probe (–460 to –1 of the 4A3 5′ flanking region) amplified with oligonucleotides 17F (5′-GTT TATATATTTTGTAAAATTATAA/TATGAG-3′) and 4A3-28R. Chromosome-central *var* gene transcription was monitored with a 277 bp PCR fragment (–514 to –237) generated with primers 5B1-514F (5′-GCATAAGCTTCCATCACATATAGTAC GACTAAGAAACA-3′) and 5B1EABR. PCR fragments were gel purified and radiolabelled with [α^{32}-P]-dATP and Klenow polymerase using primers 4A3-28R or 5B1EABR respectively. Hybridization was performed at 42°C in UltraHyb buffer (Ambion). The 28S RNA probe was amplified from 3D7 genomic DNA with primers Pf28SF (5′-GTGATGAGAT TGAAGTCAGACG-3′) and Pf28SR (5′-AGTTCAACGAAC CTCTTCTCC-3′).

Parasite transfection and reporter assays

Transient transfection of cultured 3D7 *P. falciparum* ring-stage parasites and CAT assays were performed as described previously (Voss *et al*., 2000). For luciferase assays, saponin-released parasites were lysed in reporter lysis buffer (Promega) at time points according to the occurrence of the respective DNA-binding activities, i.e. 20 h.p.i. for parasites transfected with 4A3 promoter constructs, and 30 h.p.i. for parasites transfected with 5B1 promoter constructs (Fig. 5). The lysate was freeze–thawed once. Luciferase activity in parasite lysates was determined using the luciferase assay system (Promega) and the Autoluminat LB 953 luminometer.

Sequence analysis

Sequence data for *P. falciparum* chromosomes were obtained from The Sanger Centre (http://www.sanger.-ac.uk/Projects/P_falciparum), from The Institute for Genomic Research (http://www.tigr.org), from the Naval Medical Research Centre and from The Stanford DNA Sequence and Technology Centre (http://www-sequence.stanford.edu/group/malaria). Sequencing of *P. falciparum* chromosomes was accomplished as part of the Malaria Genome Project with support from The Wellcome Trust, the Burroughs Wellcome Fund and the US Department of Defense.

Acknowledgements

We are grateful to Charles Thompson and Nancy Hynes for helpful comments on the manuscript, and to Kirk Deitsch for providing pVLH/int. We also thank Patrick Müller for making available the luminometer. This work was supported by the Swiss National Science Foundation Grant 031-059 064.99. T. S. V. received a scholarship from The Boehringer Ingelheim Fonds, Germany. Sequence data for *P. falciparum* chromosomes were obtained from The Sanger Centre (http://www.sanger.-ac.uk/Projects/P_falciparum), from the Institute for Genomic Research (http://www.tigr.org), from the Naval Medical Research Centre and from the Stanford DNA Sequence and Technology Centre (http://www-sequence.stanford.edu/group/malaria). Sequencing of *P. falciparum* chromosomes was accomplished as part of the Malaria Genome Project with support from the Wellcome Trust, the Burroughs Wellcome Fund and the US Department of Defense.

References

Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394:** 592–595.

Aranda, A., and Pascual, A. (2001) Nuclear hormone receptors and gene expression. *Physiol Rev* **81:** 1269–1304.

Åslund, L., Franzen, L., Westin, G., Persson, T., Wigzell, H., and Pettersson, U. (1985) Highly reiterated non-coding sequence in the genome of *Plasmodium falciparum* is composed of 21 base-pair tandem repeats. *J Mol Biol* **185:** 509–516.

Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X.H., Ma, X.C., Feldman, M., *et al.* (1995) Cloning the *Plasmodium falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82:** 77–87.

Baruch, D.I., Gormley, J.A., Ma, C., Howard, R.J., and Pasloske, B.L. (1996) *Plasmodium falciparum* erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci USA* **93:** 3497–3502.

Beeson, J.G., Rogerson, S.J., Cooke, B.M., Reeder, J.C., Chai, W.G., Lawson, A.M., *et al.* (2000) Adhesion of *Plasmodium falciparum*-infected erythrocytes to hyaluronic acid in placental malaria. *Nature Med* **6:** 86–90.

Berendt, A.R., Ferguson, D.J.P., Gardner, J., Turner, G., Rowe, A., Mccormick, C., *et al.* (1994) Molecular mechanisms of sequestration in malaria. *Parasitology* **108:** S19–S28.

Biggs, B.A., Gooze, L., Wycherley, K., Wollish, W., Southwell, B., Leech, J.H., and Brown, G.V. (1991) Antigenic variation in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **88:** 9171–9174.

Borst, P., and Greaves, D.R. (1987) Programmed gene rearrangements altering gene expression. *Science* **235:** 658–667.

Borst, P., and Ulbert, S. (2001) Control of VSG gene expression sites. *Mol Biochem Parasitol* **114:** 17–27.

Bull, P.C., Lowe, B.S., Kortok, M., Molyneux, C.S., Newbold, C.I., and Marsh, K. (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Med* **4:** 358–360.

Bull, P.C., Kortok, M., Kai, O., Ndungu, F., Ross, A., Lowe, B.S., *et al.* (2000) *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis* **182:** 252–259.

Chen, Q., Barragan, A., Fernandez, V., Sundstrom, A., Schlichtherle, M., Sahlen, A., *et al.* (1998a) Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite *P. falciparum*. *J Exp Med* **187:** 15–23.

Chen, Q.J., Fernandez, V., Sundstrom, A., Schlichtherle, M., Datta, S., Hagblom, P., and Wahlgren, M. (1998b) Developmental selection of *var* gene expression in *Plasmodium falciparum*. *Nature* **394:** 392–395.

Crabb, B.S., and Cowman, A.F. (1996) Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **93:** 7289–7294.

Deitsch, K.W., del Pinal, A., and Wellems, T.E. (1999) Intracluster recombination and *var* transcription switches in the antigenic variation of *Plasmodium falciparum*. *Mol Biochem Parasitol* **101:** 107–116.

Deitsch, K.W., Calderwood, M.S., and Wellems, T.E. (2001a) Malaria – cooperative silencing elements in *var* genes. *Nature* **412:** 875–876.

Deitsch, K.W., Driskill, C.L., and Wellems, T.E. (2001b) Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res* **29:** 850–853.

van Dijk, M.R., Vinkenoog, R., Ramesar, J., Vervenne, R.A.W., Waters, A.P., and Janse, C.J. (1997) Replication, expression and segregation of plasmid-borne DNA in genetically transformed malaria parasites. *Mol Biochem Parasitol* **86:** 155–162.

Donelson, J.E. (1995) Mechanisms of antigenic variation in *Borrelia hermsii* and African trypanosomes. *J Biol Chem* **270:** 7783–7786.

Firestein, R., Cui, X.M., Huie, P., and Cleary, M.L. (2000) Set domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of *Drosophila* Su (var) 3-9. *Mol Cell Biol* **20:** 4900–4909.

Fischer, K., Horrocks, P., Preuss, M., Wiesner, J., Wunsch, S., Camargo, A.A., and Lanzer, M. (1997) Expression of *var* genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. *Mol Cell Biol* **17:** 3679–3686.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., *et al.* (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407:** 1018–1022.

Fried, M., and Duffy, P.E. (1996) Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science* **272:** 1502–1504.

Gardner, J.P., Pinches, R.A., Roberts, D.J., and Newbold, C.I. (1996) Variant antigens and endothelial receptor adhesion in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **93:** 3503–3508.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419:** 498–511.

Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., and Gasser, S.M. (1996) The clustering of telomeres and colocalization with RAP1, SIR3, and SIR4 proteins in wild-type *Saccharomyces cerevisiae*. *J Cell Biol* **134:** 1349–1363.

Hernandez-Rivas, R., Mattei, D., Sterkers, Y., Peterson, D.S., Wellems, T.E., and Scherf, A. (1997) Expressed *var* genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol Cell Biol* **17:** 604–611.

Horrocks, P., and Lanzer, M. (1999) Differences in nucleosome organisation over episomally located plasmids coincides with aberrant promoter activity in *P. falciparum*. *Parasitol Int* **48:** 55–61.

Kyes, S., Pinches, R., and Newbold, C. (2000) A simple RNA analysis method shows *var* and *rif* multigene family expression patterns in *Plasmodium falciparum*. *Mol Biochem Parasitol* **105:** 311–315.

Lambros, C., and Vanderberg, J.P. (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* **65:** 418–420.

Lanzer, M., DeBruin, D., and Ravetch, J.V. (1993) Transcriptional differences in polymorphic and conserved domains of a complete cloned *Plasmodium falciparum* chromosome. *Nature* **361:** 654–657.

Leech, J.H., Barnwell, J.W., Miller, L.H., and Howard, R.J. (1984) Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum* infected erythrocytes. *J Exp Med* **159:** 1567–1575.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E., and Gasser, S.M. (1996) Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and SIR protein concentration in silencer-mediated repression. *Genes Dev* **10:** 1796–1811.

Miller, A.M., and Nasmyth, K.A. (1984) Role of DNA replication in the repression of silent mating type loci in yeast. *Nature* **312:** 247–251.

Newbold, C., Warn, P., Black, G., Berendt, A., Craig, A., Snow, B., *et al.* (1997) Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. *Am J Trop Med Hyg* **57:** 389–398.

O'Donnell, R.A., Preiser, P.R., Williamson, D.H., Moore, P.W., Cowman, A.F., and Crabb, B.S. (2001) An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected *Plasmodium falciparum* parasites. *Nucleic Acids Res* **29:** 716–724.

Pologe, L.G., and Ravetch, J.V. (1988) Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* **55:** 869–874.

Pongponratn, E., Riganti, M., Punpoowong, B., and Aikawa, M. (1991) Microvascular sequestration of parasitised erythrocytes in human *falciparum* malaria: a pathological study. *Am J Trop Med Hyg* **44:** 168–175.

Push, C., Schmitt, H., and Blin, N. (1997) Increased cloning efficiency by cycle restriction-ligation (CRL). *Techn Tips Online* **1 (**10): T40071.

Reeder, J.C., Cowman, A.F., Davern, K.M., Beeson, J.G., Thompson, J.K., Rogerson, S.J., and Brown, G.V. (1999) The adhesion of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate A is mediated by *P. falciparum* erythrocyte membrane protein 1. *Proc Natl Acad Sci USA* **96:** 5198–5202.

Roberts, D.J., Craig, A.G., Berendt, A.R., Pinches, R., Nash, G., Marsh, K., and Newbold, C.I. (1992) Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357:** 689–692.

Rowe, J.A., Moulds, J.M., Newbold, C.I., and Miller, L.H. (1997) *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* **388:** 292–295.

Rubio, J.P., Thompson, J.K., and Cowman, A.F. (1996) The *var* genes of *Plasmodium falciparum* are located in the

subtelomeric region of most chromosomes. *EMBO J* **15:** 4069–4077.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., *et al.* (1998) Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J* **17:** 5418–5426.

Scherf, A., Figueiredo, L.M., and Freitas-Junior, L.H. (2001) *Plasmodium* telomeres: a pathogen's perspective. *Curr Opin Microbiol* **4:** 409–414.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., *et al.* (1995) Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82:** 101–110.

Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., *et al.* (1995) The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82:** 89–100.

Thompson, J.F., Hayes, L.S., and Lloyd, D.B. (1991) Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* **103:** 171–177.

Thompson, J.K., Rubio, J.P., Caruana, S., Brockman, A., Wickham, M.E., and Cowman, A.F. (1997) The chromosomal organization of the *Plasmodium falciparum var* gene family is conserved. *Mol Biochem Parasitol* **87:** 49–60.

Trager, W., and Jensen, J.B. (1978) Cultivation of malarial parasites. *Nature* **273:** 621–622.

Turner, G.D.H., Morrison, H., Jones, M., Davis, T.M.E., Looareesuwan, S., Buley, I.D., *et al.* (1994) An immunohistochemical study of the pathology of fatal malaria – evidence for widespread endothelial activation and a potential role for intercellular-adhesion molecule-1 in cerebral sequestration. *Am J Pathol* **145:** 1057–1069.

Urban, B.C., Ferguson, D.J.P., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M., and Roberts, D.J. (1999) *Plasmodium falciparum*-infected erythrocytes modulate the maturation of dendritic cells. *Nature* **400:** 73–77.

Voss, T.S., Thompson, J.K., Waterkeyn, J., Felger, I., Weiss, N., Cowman, A.F., and Beck, H.P. (2000) Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5 ''flanking sequences. *Mol Biochem Parasitol* **107:** 103–115.

Voss, T.S., Mini, T., Jenoe, P., and Beck, H.P. (2002) *Plasmodium falciparum* possesses a cell-cycle regulated short-type replication protein A large subunit encoded by an unusual transcript. *J Biol Chem* **277:** 17493–17501.

Supplement IV: Pictures of study site and children of Papua New Guinea

Pictures of study site and of study participants

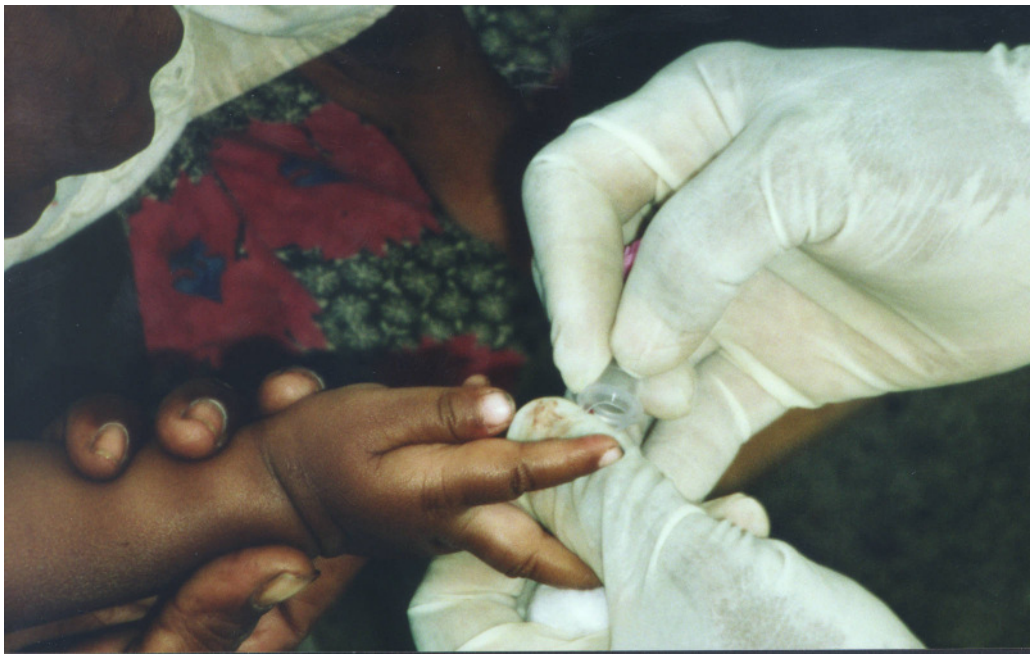
Madang Province at the North Coast of Papua New Guinea (Photo Mirjam Kaestli, 2003)

Maiwara school, study site of longitudinal study
(Photo Mirjam Kaestli, 2001)

Children of Maiwara school, participating in the longitudinal study (Photo Mirjam Kaestli, 2001)

Children of Maiwara school area (Photo Mirjam Kaestli, 2001)

Typical house in the Maiwara school area
(Photo Mirjam Kästli, 2001)



Blood collection of asymptomatic controls (Photo Mirjam Kaestli, 2001)

Children of Deguir and Yagaum in the catchment area of the case-control study
(Photo Mirjam Kästli, 2003)

References of Introduction and Discussion

1. Breman JG, Alilio MS, Mills A. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. Am J Trop Med Hyg. 2004;71:1-15.

2. Mackintosh CL, Beeson JG, Marsh K. Clinical features and pathogenesis of severe malaria. Trends Parasitol. 2004;20:597-603.

3. Severe *falciparum* malaria. World Health Organization, Communicable Diseases Cluster. Trans R Soc Trop Med Hyg. 2000;94 Suppl 1:S1-90.

4. Adams S, Brown H, Turner G. Breaking down the blood-brain barrier: signaling a path to cerebral malaria? Trends Parasitol. 2002;18:360-366.

5. English M, Waruiru C, Amukoye E et al. Deep breathing in children with severe malaria: indicator of metabolic acidosis and poor outcome. Am J Trop Med Hyg. 1996;55:521-524.

6. Clark IA, Rockett KA, Cowden WB. Proposed link between cytokines, nitric oxide and human cerebral malaria. Parasitol Today. 1991;7:205-207.

7. Schofield L, Hackett F. Signal transduction in host cells by a glycosylphosphatidylinositol toxin of malaria parasites. J Exp Med. 1993;177:145-153.

8. Berendt AR, Ferguson DJ, Newbold CI. Sequestration in *Plasmodium falciparum* malaria: Sticky cells and sticky problems. Parasitol Today. 1990;6:247-254.

9. Handunnetti SM, David PH, Perera KL, Mendis KN. Uninfected erythrocytes form "rosettes" around *Plasmodium falciparum* infected erythrocytes. Am J Trop Med Hyg. 1989;40:115-118.

10. Pain A, Ferguson DJ, Kai O et al. Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci U S A. 2001;98:1805-1810.

11. Turner GD, Morrison H, Jones M et al. An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. Am J Pathol. 1994;145:1057-1069.

12. Newbold C, Warn P, Black G et al. Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. Am J Trop Med Hyg. 1997;57:389-398.

13. Brown H, Hien TT, Day N et al. Evidence of blood-brain barrier dysfunction in human cerebral malaria. Neuropathol Appl Neurobiol. 1999;25:331-340.

14. Pino P, Vouldoukis I, Kolb JP et al. *Plasmodium falciparum*--infected erythrocyte adhesion induces caspase activation and apoptosis in human endothelial cells. J Infect Dis. 2003;187:1283-1290.

15. Chotivanich K, Sritabal J, Udomsangpetch R et al. Platelet-induced autoagglutination of *Plasmodium falciparum*-infected red blood cells and disease severity in Thailand. J Infect Dis. 2004;189:1052-1055.

16. Serghides L, Smith TG, Patel SN, Kain KC. CD36 and malaria: friends or foes? Trends Parasitol. 2003;19:461-469.

17. McGilvray ID, Serghides L, Kapus A, Rotstein OD, Kain KC. Nonopsonic monocyte/macrophage phagocytosis of *Plasmodium falciparum*-parasitized erythrocytes: a role for CD36 in malarial clearance. Blood. 2000;96:3231-3240.

18. Patel SN, Serghides L, Smith TG et al. CD36 mediates the phagocytosis of *Plasmodium falciparum*-infected erythrocytes by rodent macrophages. J Infect Dis. 2004;189:204-213.

19. Urban BC, Ferguson DJ, Pain A et al. *Plasmodium falciparum*-infected erythrocytes modulate the maturation of dendritic cells. Nature. 1999;400:73-77.

20. Sherman IW, Eda S, Winograd E. Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. Microbes Infect. 2003;5:897-909.

21. Fried M, Duffy PE. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. Science. 1996;272:1502-1504.

22. Rowe A, Obeiro J, Newbold CI, Marsh K. *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. Infect Immun. 1995;63:2323-2326.

23. Rowe JA, Shafi J, Kai OK, Marsh K, Raza A. Nonimmune IgM, but not IgG binds to the surface of *Plasmodium falciparum*-infected erythrocytes and correlates with rosetting and severe malaria. Am J Trop Med Hyg. 2002;66:692-699.

24. Rowe JA, Rogerson SJ, Raza A et al. Mapping of the region of complement receptor (CR) 1 required for *Plasmodium falciparum* rosetting and demonstration of the importance of CR1 in rosetting in field isolates. J Immunol. 2000;165:6341-6346.

25. Rowe JA, Obiero J, Marsh K, Raza A. Short report: Positive correlation between rosetting and parasitemia in *Plasmodium falciparum* clinical isolates. Am J Trop Med Hyg. 2002;66:458-460.

26. Leech JH, Barnwell JW, Miller LH, Howard RJ. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. J Exp Med. 1984;159:1567-1575.

27. Waller KL, Cooke BM, Nunomura W, Mohandas N, Coppel RL. Mapping the binding domains involved in the interaction between the *Plasmodium falciparum* knob-associated histidine-rich protein (KAHRP) and the cytoadherence ligand P. *falciparum* erythrocyte membrane protein 1 (PfEMP1). J Biol Chem. 1999;274:23808-23813.

28. Waterkeyn JG, Wickham ME, Davern KM et al. Targeted mutagenesis of *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) disrupts cytoadherence of malaria-infected red blood cells. EMBO J. 2000;19:2813-2823.

29. Kriek N, Tilley L, Horrocks P et al. Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes. Mol Microbiol. 2003;50:1215-1227.

30. Gardner MJ, Hall N, Fung E et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002;419:498-511.

31. Baruch DI, Pasloske BL, Singh HB et al. Cloning the P. *falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell. 1995;82:77-87.

32. Su XZ, Heatwole VM, Wertheimer SP et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell. 1995;82:89-100.

33. Smith JD, Chitnis CE, Craig AG et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell. 1995;82:101-110.

34. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH. Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. Mol Biochem Parasitol. 2000;110:293-310.

35. Peterson DS, Miller LH, Wellems TE. Isolation of multiple sequences from the *Plasmodium falciparum* genome that encode conserved domains homologous to those in erythrocyte-binding proteins. Proc Natl Acad Sci U S A. 1995;92:7100-7104.

36. Chen Q, Barragan A, Fernandez V et al. Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. *falciparum*. J Exp Med. 1998;187:15-23.

37. Smith JD, Kyes S, Craig AG et al. Analysis of adhesive domains from the A4VAR *Plasmodium falciparum* erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol Biochem Parasitol. 1998;97:133-148.

38. Smith JD, Craig AG, Kriek N et al. Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci U S A. 2000;97:1766-1771.

39. Buffet PA, Gamain B, Scheidig C et al. *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. Proc Natl Acad Sci U S A. 1999;96:12743-12748.

40. Degen R, Weiss N, Beck HP. *Plasmodium falciparum*: cloned and expressed CIDR domains of PfEMP1 bind to chondroitin sulfate A. Exp Parasitol. 2000;95:113-121.

41. Yipp BG, Robbins SM, Resek ME et al. Src-family kinase signaling modulates the adhesion of *Plasmodium falciparum* on human microvascular endothelium under flow. Blood. 2003;101:2850-2857.

42. Gamain B, Smith JD, Miller LH, Baruch DI. Modifications in the CD36 binding domain of the *Plasmodium falciparum* variant antigen are responsible for the inability of chondroitin sulfate A adherent parasites to bind CD36. Blood. 2001;97:3268-3274.

43. Rowe JA, Kyes SA. The role of *Plasmodium falciparum var* genes in malaria in pregnancy. Mol Microbiol. 2004;53:1011-1019.

44. Voss TS, Thompson JK, Waterkeyn J et al. Genomic distribution and functional characterisation of two distinct and conserved *Plasmodium falciparum var* gene 5' flanking sequences. Mol Biochem Parasitol. 2000;107:103-115.

45. Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. Malar J. 2003;2:27.

46. Smith JD, Gamain B, Baruch DI, Kyes S. Decoding the language of *var* genes and *Plasmodium falciparum* sequestration. Trends Parasitol. 2001;17:538-545.

47. Freitas-Junior LH, Bottius E, Pirrit LA et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. *falciparum*. Nature. 2000;407:1018-1022.

48. Salanti A, Staalsoe T, Lavstsen T et al. Selective upregulation of a single distinctly structured *var* gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. Mol Microbiol. 2003;49:179-191.

49. Salanti A, Dahlback M, Turner L et al. Evidence for the Involvement of VAR2CSA in Pregnancy-associated Malaria. J Exp Med. 2004;200:1197-1203.

50. Artavanis-Tsakonas K, Tongren JE, Riley EM. The war between the malaria parasite and the immune system: immunity, immunoregulation and immunopathology. Clin Exp Immunol. 2003;133:145-152.

51. Urban BC, Roberts DJ. Malaria, monocytes, macrophages and myeloid dendritic cells: sticking of infected erythrocytes switches off host cells. Curr Opin Immunol. 2002;14:458-465.

52. Donati D, Zhang LP, Chen Q et al. Identification of a polyclonal B-cell activator in *Plasmodium falciparum*. Infect Immun. 2004;72:5412-5418.

53. Kyes S, Horrocks P, Newbold C. Antigenic variation at the infected red cell surface in malaria. Annu Rev Microbiol. 2001;55:673-707.

54. Scherf A, Hernandez-Rivas R, Buffet P et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J. 1998;17:5418-5426.

55. Deitsch KW, Calderwood MS, Wellems TE. Malaria. Cooperative silencing elements in *var* genes. Nature. 2001;412:875-876.

56. Fernandez V, Chen Q, Sundstrom A et al. Mosaic-like transcription of *var* genes in single *Plasmodium falciparum* parasites. Mol Biochem Parasitol. 2002;121:195-203.

57. Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI. A study of *var* gene transcription in vitro using universal *var* gene primers. Mol Biochem Parasitol. 2000;105:13-23.

58. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP. Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters. Mol Microbiol. 2003;48:1593-1607.

59. Calderwood MS, Gannoun-Zaki L, Wellems TE, Deitsch KW. *Plasmodium falciparum var* genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. J Biol Chem. 2003;278:34125-34132.

60. Galinski MR, Corredor V. Variant antigen expression in malaria infections: posttranscriptional gene silencing, virulence and severe pathology. Mol Biochem Parasitol. 2004;134:17-25.

61. Barnwell JW, Howard RJ, Miller LH. Influence of the spleen on the expression of surface antigens on parasitized erythrocytes. Ciba Found Symp. 1983;94:117-136.

62. Fischer K, Chavchich M, Huestis R et al. Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of *Plasmodium* chabaudi, a malaria species that undergoes antigenic variation in the laboratory mouse. Mol Microbiol. 2003;48:1209-1223.

63. Demar M, Legrand E, Hommel D, Esterre P, Carme B. *Plasmodium falciparum* malaria in splenectomized patients: two case reports in French Guiana and a literature review. Am J Trop Med Hyg. 2004;71:290-293.

64. Abdel-Latif MS, Dietz K, Issifou S, Kremsner PG, Klinkert MQ. Antibodies to *Plasmodium falciparum* rifin proteins are associated with rapid parasite clearance and asymptomatic infections. Infect Immun. 2003;71:6229-6233.

65. Winograd E, Sherman IW. Malaria infection induces a conformational change in erythrocyte band 3 protein. Mol Biochem Parasitol. 2004;138:83-87.

66. Saul A. The role of variant surface antigens on malaria-infected red blood cells. Parasitol Today. 1999;15:455-457.

67. Biggs BA, Anders RF, Dillon HE et al. Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. J Immunol. 1992;149:2047-2054.

68. Bull PC, Lowe BS, Kortok M et al. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med. 1998;4:358-360.

69. Bull PC, Kortok M, Kai O et al. *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J Infect Dis. 2000;182:252-259.

70. Bian Z, Wang G, Tian X, Fan J. Expression of *Plasmodium falciparum*-infected erythrocyte membrane protein from cerebral malaria patients. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 1999;17:359-362.

71. Kirchgatter K, Portillo HA. Association of severe noncerebral *Plasmodium falciparum* malaria in Brazil with expressed PfEMP1 DBL1 alpha sequences lacking cysteine residues. Mol Med. 2002;8:16-23.

72. Jensen AT, Magistrado P, Sharp S et al. *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A *var* genes. J Exp Med. 2004;199:1179-1190.

73. Roberts DJ, Craig AG, Berendt AR et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature. 1992;357:689-692.

74. Peters J, Fowler E, Gatton M et al. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci U S A. 2002;99:10689-10694.

75. Chen Q, Pettersson F, Vogt AM et al. Immunization with PfEMP1-DBL1alpha generates antibodies that disrupt rosettes and protect against the sequestration of *Plasmodium falciparum*-infected erythrocytes. Vaccine. 2004;22:2701-2712.

76. Allen SJ, O'Donnell A, Alexander ND, Clegg JB. Severe malaria in children in Papua New Guinea. QJM. 1996;89:779-788.

77. Muller I, Bockarie M, Alpers M, Smith T. The epidemiology of malaria in Papua New Guinea. Trends Parasitol. 2003;19:253-259.

78. Smith T, Genton B, Baea K et al. Prospective risk of morbidity in relation to malaria infection in an area of high endemicity of multiple species of *Plasmodium*. Am J Trop Med Hyg. 2001;64:262-267.

79. Mgone CS, Koki G, Paniu MM et al. Occurrence of the erythrocyte band 3 (AE1) gene deletion in relation to malaria endemicity in Papua New Guinea. Trans R Soc Trop Med Hyg. 1996;90:228-231.

80. Oppenheimer SJ, Higgs DR, Weatherall DJ, Barker J, Spark RA. Alpha thalassaemia in Papua New Guinea. Lancet. 1984;1:424-426.

81. Cockburn IA, Mackinnon MJ, O'Donnell A et al. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. Proc Natl Acad Sci U S A. 2004;101:272-277.

82. Patel SS, King CL, Mgone CS, Kazura JW, Zimmerman PA. Glycophorin C (Gerbich antigen blood group) and band 3 polymorphisms in two malaria holoendemic regions of Papua New Guinea. Am J Hematol. 2004;75:1-5.

83. al Yaman F, Genton B, Mokela D et al. Human cerebral malaria: lack of significant association between erythrocyte rosetting and disease severity. Trans R Soc Trop Med Hyg. 1995;89:55-58.

84. Ofori MF, Staalsoe T, Bam V et al. Expression of variant surface antigens by *Plasmodium falciparum* parasites in the peripheral blood of clinically immune pregnant women indicates ongoing placental infection. Infect Immun. 2003;71:1584-1586.

85. Tuikue Ndam NG, Fievet N, Bertin G et al. Variable Adhesion Abilities and Overlapping Antigenic Properties in Placental *Plasmodium falciparum* Isolates. J Infect Dis. 2004;190:2001-2009.

86. Beeson JG, Brown GV, Molyneux ME et al. *Plasmodium falciparum* isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties. J Infect Dis. 1999;180:464-472.

87. Fernandez V, Treutiger CJ, Nash GB, Wahlgren M. Multiple adhesive phenotypes linked to rosetting binding of erythrocytes in *Plasmodium falciparum* malaria. Infect Immun. 1998;66:2969-2975.

88. Dyer M, Day K. Expression of *Plasmodium falciparum* trimeric G proteins and their involvement in switching to sexual development. Mol Biochem Parasitol. 2000;110:437-448.

89. Bull PC, Lowe BS, Kaleli N et al. *Plasmodium falciparum* infections are associated with agglutinating antibodies to parasite-infected erythrocyte surface antigens among healthy Kenyan children. J Infect Dis. 2002;185:1688-1691.

90. Recker M, Nee S, Bull PC et al. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature. 2004;429:555-558.

91. Antia R, Koella J. Theoretical immunology: parasitic turncoat. Nature. 2004;429:511-513.

92. Robinson BA, Welch TL, Smith JD. Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol. 2003;47:1265-1278.

93. Gatton ML, Cheng Q. Investigating antigenic variation and other parasite-host interactions in *Plasmodium falciparum* infections in naive hosts. Parasitology. 2004;128:367-376.

94. Gatton ML, Peters JM, Fowler EV, Cheng Q. Switching rates of *Plasmodium falciparum var* genes: faster than we thought? Trends Parasitol. 2003;19:202-208.

95. Horrocks P, Pinches R, Christodoulou Z, Kyes SA, Newbold CI. Variable *var* transition rates underlie antigenic variation in malaria. Proc Natl Acad Sci U S A. 2004;101:11129-11134.

96. Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C. Immunity to non-cerebral severe malaria is acquired after one or two infections. Nat Med. 1999;5:340-343.

97. Gratepanche S, Gamain B, Smith JD et al. Induction of crossreactive antibodies against the *Plasmodium falciparum* variant protein. Proc Natl Acad Sci U S A. 2003;100:13007-13012.

98. Taylor HM, Kyes SA, Newbold CI. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. Mol Biochem Parasitol. 2000;110:391-397.

99. Kraemer SM, Smith JD. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum var* gene family. Mol Microbiol. 2003;50:1527-1538.

100. Cubillos M, Espejo F, Purmova J, Martinez JC, Patarroyo ME. Alpha helix shortening in 1522 MSP-1 conserved peptide analogs is associated with immunogenicity and protection against P. *falciparum* malaria. Proteins. 2003;50:400-409.

101. Fried M, Wendler JP, Mutabingwa TK, Duffy PE. Mass spectrometric analysis of *Plasmodium falciparum* erythrocyte membrane protein-1 variants expressed by placental malaria parasites. Proteomics. 2004;4:1086-1093.

102. Reeder JC, Cowman AF, Davern KM et al. The adhesion of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate A is mediated by P. *falciparum* erythrocyte membrane protein 1. Proc Natl Acad Sci U S A. 1999;96:5198-5202.

Curriculum vitae

- Name: Mirjam Kaestli
- Address (work): Swiss Tropical Institute, Socinstrasse 57, P.O. Box, 4002 Basel, Switzerland
  Email: mirjam.kaestli@unibas.ch
  Phone: 0041-61-284 81 20, Fax: 0041-61-284 81 01

- Born: 21 Nov 1975
- Place of Birth: St. Gallen, Switzerland
- Citizen of: St. Margrethen (SG), Switzerland
- Languages:

German:	Mother tongue
English:	Very good knowledge
French:	Intermediate knowledge
Spanish and Pidgin:	Basic understanding

EDUCATION

- 2000 - Jan 05: **Ph.D. in Microbiology**, Swiss Tropical Institute and Biozentrum, University of Basel, Switzerland,

  Subject: "*Plasmodium falciparum var gene expression dynamics and its relevance in malaria disease in children from Papua New Guinea"*
  Supervisor: Prof. Dr. H.-P. Beck

- 1995 - 2000: **Diploma in Molecular Biology**, Biozentrum, University of Basel,

  Diploma thesis in Molecular Parasitology, Swiss Tropical Institute, University of Basel, Supervisor: Prof. Dr. H.-P. Beck
  Subject: "*The expression of the Plasmodium falciparum var gene family in naturally infected blood samples*"

- 1990 - 1995: General qualification for university entrance, Kantonsschule St. Gallen, Switzerland, Matura Typus B (Latin)

WORK EXPERIENCE

- **2001 and 2003 4 and 3 months field study at Institute of Medical Research, Madang, Papua New Guinea**
  Coordination and conduct of field project for molecular expression studies in malaria infected children

- **June 2003 stay at Ifakara Health Research and Development Centre, Ifakara, Tanzania**
  Training of local graduate student in advanced lab techniques and organizational support for field project

- Experience in the following lab techniques:

  Molecular biology
  General molecular biology technologies incl. cloning and sequencing, quantitative real-time PCR, magnetic bead-based isolation of gene specific mRNA, Northern- and Dot Blot analysis, reverse-transcription PCR, RFLP analysis

  Cell culture
  Plasmodium falciparum, CHO cells, static binding assays

  Bioinformatics
  Application of various programs for sequence analysis and database search (online, PC and Mac) (e.g. DNAstar, BioEdit, Mega, Arlequin, Phylip, ncbi Blast, plasmo database)

  Biostatistics
  Basic knowledge in biostatistical analysis (Stata version 8)

TRAINEESHIPS

- July 98 - Aug 98 Laboratory practical training at microbiology unit of clinical laboratory at Victoria Hospital, Seychelles:
  Learning of laboratory techniques for diagnosis of bacterial infections

- Jan 95 - June 95 Training to Swiss Red Cross assistant nurse and working in nursing home, St. Gallen, Switzerland

PUBLICATIONS

- **Kaestli M**, Cockburn I, Cortes A, Baea K, Rowe A, Beck HP. Differential expression of *Plasmodium falciparum var* gene subgroups is associated with virulence in a malaria case-control study of children in Papua New Guinea. submitted to Blood (January 2005)

- **Kaestli M**, Cortes A, Lagog M, Ott M, Beck HP. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. J Infect Dis. 2004 May 15;189(10):1942-51.

- Voss TS, **Kaestli M**, Vogel D, Bopp S, Beck HP. Identification of nuclear proteins that interact differentially with *Plasmodium falciparum var* gene promoters. Mol Microbiol. 2003 Jun; 48(6):1593-607.

PRESENTATIONS OF PhD PROJECT

- July 2004 — Oral presentation at "IX European Multicolloquium of Parasitology", EMOP IX and MEEGID VII, Valencia, Spain
- April 2004 — Oral presentation at "The Scripps Research Institute", San Diego, USA
- Nov 2003 — Oral presentation at “Functional Genomics Board Meeting”, Novartis Institutes for Biomedical Research, Basel, Switzerland
- Oct 2003 — Oral presentation at “PhD student meeting”, Swiss Society of Tropical Medicine and Parasitology SSTMP, Münchenwiler, Switzerland
- Sep 2001 — Poster presentation at “Malaria meeting of the British Society of Parasitology”, Leeds, UK

SCHOLARSHIP

- June 2004 — Winner of European Federation of Parasitologists (EFP) Scholarship for EMOP IX

During my studies I attended lectures and courses of the following lecturers:

A. Wiemken, A. Zumstein, C. Daubenberger, C. Lengeler, C. Thompson, G. Pluschke, G. Schatz, G. Schwarz, H. Huber, H. Riezman, H. Rudin, H. Sigel, HC. Im Hof, HP. Beck, HP. Hauri, I. Felger, I. Sick, J. Engel, J. Seelig, JG. Nicholls, L. Tauscher, M. Hall, M. Jungen, M. Primig, M. Spieß, M. Tanner, MA. Rüegg, N. Weiß, P. Jenoe, P. Matthias, P. Philippsen, P. Vounatsou, R. Brun, R. Gisler, T. Bickle, T. Schirmer, T. Smith, U. Jenal, U. Séquin, UA. Meyer, W. Keller, W. Rudin, W. Gehring