Cloning of the Repertoire of Individual *Plasmodium falciparum* var Genes Using Transformation Associated Recombination (TAR)

Annette Gaida1,2,*, Marion M. Becker3,*, Christoph D. Schmid1,2, Tobias Bühlmann1,2, Edward J. Louis3,*, Hans-Peter Beck1,2

1 Department for Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland, 2 University of Basel, Basel, Switzerland, 3 Centre for Genetics and Genomics, The University of Nottingham, Queen’s Medical Centre, Nottingham, United Kingdom

Abstract

One of the major virulence factors of the malaria causing parasite is the *Plasmodium falciparum* encoded erythrocyte membrane protein 1 (PfEMP1). It is translocated to the membrane of infected erythrocytes and expressed from approximately 60 var genes in a mutually exclusive manner. Switching of var genes allows the parasite to alter functional and antigenic properties of infected erythrocytes, to escape the immune defense and to establish chronic infections. We have developed an efficient method for isolating VAR genes from telomeric and other genome locations by adapting transformation-associated recombination (TAR) cloning, which can then be analyzed and sequenced. For this purpose, three plasmids each containing a homologous sequence representing the upstream regions of the group A, B, and C var genes and a sequence homologous to the conserved acidic terminal segment (ATS) of var genes were generated. Co-transfection with *P. falciparum* strain ITG2F6 genomic DNA in yeast cells yielded 200 TAR clones. The relative frequencies of clones from each group were not biased. Clones were screened by PCR, as well as Southern blotting, which revealed clones missed by RFLP and end sequencing. Physical analysis of 36 clones revealed 27 distinct types potentially representing 50% of the var gene repertoire. Three clones were selected for sequencing and assembled into single var gene containing contigs. This study demonstrates that it is possible to rapidly obtain the repertoire of var genes from *P. falciparum* within a single set of cloning experiments. This technique can be applied to individual isolates which will provide a detailed picture of the diversity of var genes in the field. This is a powerful tool to overcome the obstacles with cloning and assembly of multi-gene families by simultaneously cloning each member.

Introduction

Malaria tropica is caused by the apicomplexan parasite *Plasmodium falciparum* and accounts for approximately 500 million cases with a death toll of 800 000 annually [1]. Nearly all pathology is caused by the intimate interaction between the infected erythrocyte and the host’s capillary endothelium, which is conferred by a parasite-derived molecule in the membrane of the erythrocyte, termed erythrocyte membrane protein 1 (*PfEMP1*). It is considered to be the major virulence factor of *P. falciparum* [2,3]. *PfEMP1* shows antigenic variation and is encoded by about 60 different var genes per haploid genome. The expression of a *PfEMP1* variant is mutually exclusive [4,5] and the protein is a major target of the adaptive immune response. It is believed that protection against clinical disease depends on the number of different *PfEMP1* variants previously seen by the immune system [6,7]. A significant number of these proteins are able to bind to different endothelial receptors such as CD36 or ICAM-1, leading to sequestration of infected red blood cells from the blood stream, thus avoiding clearance by the immune system and the spleen [8]. Switching expression of var genes is often accompanied by differences in the binding phenotype, allowing the parasite to escape the host’s immune response.

var genes vary in their size from 6 to 13 kb [9]. They are flanked by a large upstream promoter region that is considered to play an important role in the regulation of var genes [10,11]. The *P. falciparum* sequencing project revealed the complete set of var genes of the 3D7 strain. They can be grouped into group A, B and C var genes according to their 5’ upstream regions [12,13]. Group A and B var genes are located almost exclusively at telomere ends, whereas upsA var genes are transcribed towards the telomere and upsB var genes towards the centromere. In contrast, upsC var genes are located in central chromosomal regions [8]. Such distribution of group A, B, and C var genes has also been observed in parasite samples from naturally infected patients, however, sequence diversity appears to be vast, perhaps unlimited, with thousands
of different Duffy-binding-like (DBL) domains already identified in parasite field isolates [14–16].

var genes possess a two exon structure with the first exon between 3.5 to 9 kb in size [12]. Exon 1 encodes the highly variable region that is expressed on the surface of infected red blood cells as well as a transmembrane domain. Each var gene comprises a distinct and unique set of domains which most probably have evolved through recombination from a common ancestor. The most conserved domain is the DBL1alpha domain with identities of up to 52% between different var genes [17]. Following an AT-rich intron, exon 2 encodes the more conserved amino terminal segment (ATS) that is located in the intracellular part of the infected red blood cell.

Sequencing downstream of the DBL1alpha domain poses difficulties because of the high AT richness, long homopolymeric sequence stretches, repeat units of various sizes and the fact that var genes evolve by frequent recombination. Both Sanger sequencing and next generation sequencing of whole genomes fail to unequivocally determine sequences of full-length var genes due to difficulties in the assembly of var genes de novo or onto existing scaffolds. To overcome these limitations we applied a technique that could specifically generate clones containing single var gene units. This information could provide important insights into the pathology of malaria. So far, only a very small number of var genes have been associated with distinct clinical presentations, such as the var2csa gene (PPL 003c) encoding a PfEMP1 variant that binds to chondroitin sulfate A and is responsible for pregnancy-associated malaria [17]. Few studies were able to associate upstream regions of group A, B, and C var genes as well as the DBL1alpha domain in malaria infected children from endemic areas with clinical disease presentation [3,15,16,18,19,20,21]. However, analysis of complete var genes from patient samples might result in a different or more complete picture of parasite-host interactions involving PfEMP1.

Transformation associated recombination (TAR) cloning has been successfully used for cloning of telomeric ends in various organisms [22], for cloning of centromere regions [23,24], for gap closure of difficult sequence stretches [25], and for selectively cloning vsg-expression sites from Trypanosoma brucei [26,27]. We applied a modified protocol of transformation-associated recombination [28,29] to isolate individual var genes from P. falciparum strain ITG2F6. The Brazilian strain IT4/25/4, the parental strain of ITG2F6, has been partially sequenced and its var gene sequences of the ITG2F6 strain have been compared to those of 3D7 and HB3 [30]. This allows comparison of isolated var gene sequences of the ITG2F6 strain with IT4/25/4 sequences and facilitates sequence analysis. TAR cloning offers several advantages over traditional cloning methods when genomic regions are difficult to deal with [29] as demonstrated with the cloning and characterization of bloodstream form expression sites from T. brucei [26,27]. In the case of var genes, the alternative to TAR would be whole genome sequencing projects or the screening of random insert libraries rather than efficient targeted library construction using TAR. Here, we present the construction of a var clone library containing about 200 clones that were obtained in experiments using three different TAR cloning vectors specific for upsA, upsB and upsC var genes. We demonstrate the feasibility of analyzing these clones by PCR colony screening and Southern blotting. A selected number of var clones were transformed into E. coli, allowing plasmid preparations of high concentrations, for restriction fragment length polymorphism (RFLP) analyses or sequencing. Moreover, we used deep sequencing on three isolated var clones and were able to assemble them de novo.

TAR cloning can be used as a rapid and efficient method for the isolation of var genes of different P. falciparum strains. Moreover, utilization of this technique on field samples of infected patients will greatly support studies aiming to associate certain var gene types (present or expressed) with particular clinical presentations. Combined with long range PCR, this technique will provide a means to study the regulation and dynamics of var genes in naturally infected individuals and the associated functional significance of var gene expression that currently is an impossible task. Furthermore, it will provide data on the diversity of the entire full-length var gene repertoire.

Results

Construction of TAR library containing var genes from P. falciparum strain ITG2F6

In order to isolate the repertoire of single var genes from the P. falciparum strain ITG2F6, a modified TAR cloning protocol was applied. Three TAR cloning vectors were designed to contain upstream sequences of group A, B and C var genes as 5’ targeting sites as well as the ATS domain as 3’ targeting site. In addition, they contained a yeast centromere, an autonomously replicating sequence (ARS), positive and negative selectable markers as well as an ori and ampR gene for replication in E. coli cells. Each of the three vectors was co-transformed with genomic DNA of the ITG2F6 strain into yeast spheroplasts. Homologous recombination within the spheroplasts resulted in circular yeast plasmids presumably containing a single var gene (Figure 1). We performed three independent experiments for each of the three TAR cloning vectors and obtained a total of 205 TAR clones of which 46 resulted from recombination with pEB_upsA_ATS, 91 from recombination with pEB_upsB_ATS, and 68 from recombination with pEB_upsC_ATS. This proportion is in accordance with the known distribution of var groups within the genomes of 3D7, IT4/25/4 and HB3 (Table S1).

PCR screening and Southern blot analysis of TAR clones

In an initial PCR screen, using primers for the most conserved DBL1alpha domain, 48% of TAR clones were identified as positive, i.e. containing a var gene sequence. Subsequent Southern blot analysis conducted on a subset of 104 TAR clones identified 95 putative positive TAR clones (an example is shown in Figure 2), of which only 74 had been positive for the DBL1alpha domain by PCR. This demonstrates that DNA hybridization can reveal more positive clones than PCR due to variation in the primer region. Yeast colonies resulting from transformation with ‘empty’ vector without insert produced a band of approximately 6.0 kb (vector pEB_upsA-ATS was used as negative control, Figure 2). We considered all clones larger than this as positive, taking into account the possibility that partial var genes could also be isolated. The additional 21 identified by Southern blotting may also contain var genes that are more divergent. A total of 119 putative TAR clones were identified with either or both methods within the 3 sets of experiments. TAR clone sizes ranged from less than 6 kb to approximately 30 kb (examples shown in Table 2). Sequencing of a 30 kb clone generated two contigs which could be linked by gap closure PCR yielding a total length of 11.5 kb which was further confirmed by RFLP analysis (see below).

Restriction fragment Length polymorphism (RFLP) of selected TAR clones

In order to perform RFLP analysis, we transformed E. coli with DNA from yeast containing several candidate TAR clones. Despite the large size of the plasmids and the toxicity of yeast DNA to bacteria, we obtained 36 E. coli clones containing the TAR plasmids. To analyze the variability of the isolated TAR
Figure 1. Construction of the var TAR vectors. (A) TAR cloning vectors pEB_upsA/B or C_ATS were constructed from the general TAR vector pEB2 (M. Becker and E.J. Louis, unpubl.). The 5' targeting sites upsA (185 bp), upsB (263 bp) or upsC (203 bp) are indicated by black arrows. The 3' targeting site is identical in all three TAR cloning vectors and contains part of the ATS region (287 bp) indicated with black squares. The yeast centromere and ARS are labelled as CEN and ARS, the positive selectable marker gene URA3 is indicated as light grey box (URA3), the negative selectable marker CYH2 with a white box (CYH2), and the Ampicillin resistance gene with a dark grey box (Amp-R). SalI and ClaI restriction sites are indicated. Homologous recombination between the targeting sites and P. falciparum genomic DNA results in the cloning of a var gene flanked by those sites as a circular molecule. (B) 5' and 3' targeting sequences used in the TAR cloning vectors pEB_upsA/B or C_ATS.

5' targeting site for upsA:
TTTATTATTTTTGTGAGGTGAAAAATTTGAAATAATAGGTTTTATATTACCAAT
ATTATTATAATATGTATTATCCATGAGAAATAATATTGTTTTCCACCAGTTAATAATATT
TTTAAATATATATTGAGGATATATTGTTTTATAATAATATGTTTTATAG (185 bp)

5' targeting site for upsB:
ATGTAATTGTGTTTTTTTTGTTAGAATATTTAATAAATATATAAATATATATTTTT
TTTTAATAATATAATATAAATATATATATATAATATATAAATATATATATAAATATATATA
ATACCTCTTATATATATATATATATATATATATATATATATATATATATATATATAT
TTTTCCATATATAGTGAGGAGAGATATATTGTTGGCAAAATATATATATATGTTGGAGCAC
ACCCAG (263 bp)

5' targeting site for upsC:
CACATATTAAAGCCTAGAAAACAAATACATCAGACACAAAACATAGTGACTACATTACATGTR
TTTCCCCATATTTTCATGGTACATTACCTCATGTTAGTAATACCCATATATATATATATAT
ACCGTATATAGTGAGGAGATATATTGTTGGCAAAATATATATATATGTTGGAGCAC
ACCCAG (203 bp)

3' targeting site for ATS:
CTGAGTTTTCATATACAAATGAGATGCTAATCTATAAAATATGTTTTATTTAATATATGTT
CACCTTTAATATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
ATAGTGTGCGAGATGATTATATATATGTTGGAAATGTTTATATATATGTTGGAAATGTTTATAT
ATAGTGTGCGAGATGATTATATATGTTGGAAATGTTTATATATATGTTGGAAATGTTTATAT
AAATTCCTATAGCCAGTATGTTATCG (287 bp)
clones, we performed RFLP using restriction enzymes NotI and NdeI. The GC-rich NotI restriction site is unlikely to cut the AT-rich \( P. falciparum \) gene sequence. There is a single NotI and one NdeI restriction site in the vector backbone of pEB2 which results in a distinct characteristic band of 1600 bp in all TAR clones. Screening the published 3D7 and IT4/4/25 \( var \) gene sequences, NdeI restriction sites were found to be present at one to eight locations in individual \( var \) genes, resulting in expected distinct banding patterns for different \( var \) gene inserts. All 36 \( E. coli \) clones were analyzed in this way. All upsA clones showed different NotI/NdeI digestion patterns indicating that different upsA \( var \) genes have been isolated. For upsB 23 TAR clones were analyzed, resulting in 20 unique patterns. One pattern was represented three times and one pattern was represented twice. The restriction digest of ten upsC clones revealed six patterns of which four were unique, one was represented four times and one twice.

These RFLP data demonstrate a high variability of isolated \( var \) genes. Although \( var \) gene IT4 \( var23p \) (Table 2) was represented four times, no bias towards cloning certain \( var \) genes was observed which could lead to the isolation of a restricted set of \( var \) genes.

Figure 2 shows the RFLP pattern of TAR clones upsAIII\_10, upsBl\_29 and upsCl\_3, representative of all other examined clones. The total size of the bands summed to 15.5 kbp for upsAIII\_10, 15 kbp for upsBl\_29 and 11.5 kbp for upsCl\_3. For upsAIII\_10 and upsBl\_29 this is consistent with the results obtained from Southern blot analysis and shotgun sequencing (Table 2 and section below), for clone upsCl\_3 it is consistent with shotgun sequencing (section below).

**Sequence analysis of TAR clones**

We sequenced a number of PCR-products obtained directly from the colony screening 7 TAR clones using both forward and reverse DBL1alpha -primers. As the degree of DBL1alpha conservation is only about 50% [17], annealing of the primers to different \( var \) genes is probably not optimal. The AT-richness of the \( var \) genes may explain why low sequence quality was obtained in some instances. However, BLAST analysis of all sequences confirmed the presence of \( P. falciparum \) \( var \) gene sequences, demonstrating that the TAR clones contained \( var \) gene inserts.

To confirm that nearly complete \( var \) genes were cloned, a representative clone for each of the different RFLP patterns was selected and 5\' and 3\' sequences were generated using primers located on either recombination site. Correct 5\' and 3\' \( var \) gene sequences were obtained for 30 clones. For two clones only a correct 3\' sequence was obtained. Three sequences represented group A, 20 sequences group B and nine group C \( var \) genes (Table 2). All sequences extended beyond the targeting sequence confirming successful recombination and cloning. As expected from the RFLP results, BLAST analysis gave identical hits for some TAR clones (Table 2). One clone, isolated with the upsC 5\' targeting sequence showed higher similarity to a group B \( var \) gene (PlasmoDB accession number PFD1015c) confirming previously observed limited fidelity within these two groups of \( var \) genes [16].

For each \( var \) group, we sequenced one yeast clone by deep sequencing (Illumina) and the sequences aligned well with the 5\' and 3\' sequences obtained by Sanger sequencing of the plasmids. For clone upsAIII\_10 and clone upsBl\_29 we were able to...
assemble all reads \textit{de novo} into a single contig. We were able to identify the backbone of the TAR cloning vector pEB2, the 5’ targeting sites upsA and upsB, and the 3’ targeting site ATS in both contigs. BLAST searches were performed with the complete isolated var sequences against the NCBI non-redundant nucleotide database. Clone upsAII_10 consisted of 11396 bp. BLAST search identified a partially sequenced \textit{P. falciparum} var gene of the FCR3 strain (Accession number AJ139301) with 99% identity (13 mismatches over 9024 bp).

BLAST search of TAR clone upsBL29 identified the IT4\_var13 var gene (Accession number EF158072) with an identity of 99%. This similarity was expected since the strain ITG2F6 used here is a subclone of strain IT4/4/25. IT4\_var13 is a group B var gene confirming the specificity of the upsB targeting sequence.

Sequencing of the upsC TAR clone upsCL3 revealed two contigs each overlapping at one end with the pEB2 backbone. The gap between the contigs was closed by PCR using primers CI\_3\_pF and CI\_3\_pR (Table 1) corresponding to a sequence 90 bp before the unmatched end of either contig. The 500 bp fragment (Figure S1) resulted in an extremely AT rich sequence. The 5’ end generated with the forward primer showed a 100% match with the last 49 bp of contig 1 and similarly, the sequence generated with the reverse primer matched to the first 45 bp of the of contig 2 (Supplementary Figure 1). Further analysis revealed that the missing sequence stretch lies within the intron sequence. BLAST searches of the gap sequence resulted in a hit on the intron of \textit{var} gene PF10005w (Accession number AL844508). The complete sequence of clone upsCL3 showed highest similarities with the partial IT4\_var23 var gene (Accession number EF158077) belonging to group C \textit{var} genes. This \textit{var} gene has only been sequenced partially over 2049 bp displaying a 99% identity to the upsCL3 TAR clone. The 3’ end of clone upsCL3 showed an 85% identity with another partially sequence \textit{var} gene (VARC1, Accession number AF100791).

All three isolated clones contained an open reading frame beginning with a methionine and were interrupted by an intron. The 3’ targeting sites of all three TAR cloning vectors do not cover the complete ATS domain, therefore the isolated \textit{var} genes lack a stop codon in the second exon.

\textbf{Table 1.} Primers used for targeting, screening, sequencing, and gap closure.

| primer name | sequence 5’to 3’ |
|-------------|-----------------|
| upsA\_fw    | tataaatgcttttctttatatattttagggagaa |
| upsA\_rev   | taatatagctttatataattttagggagaa |
| upsB\_fw    | tataaatgcttttctttatatattttagggagaa |
| upsB\_rev   | taatatagcttttctttatatattttagggagaa |
| upsC\_fw    | tataaatgcttttctttatatattttagggagaa |
| upsC\_rev   | taatatagcttttctttatatattttagggagaa |
| ATS\_fw     | tataaatgcttttctttatatattttagggagaa |
| ATS\_rev    | taatatagcttttctttatatattttagggagaa |
| DBL1alpha\_fw | gcaagcagctttatattttagggagaa |
| DBL1alpha\_rev | gcaagcagctttatattttagggagaa |
| CI\_3\_pf   | ggttatatatatttttagggagaa |
| CI\_3\_pR   | ggttatatatatttttagggagaa |

Restriction sites are underlined. W denotes nucleotides A or T, R denotes A or G, and K denotes G or T.

doi:10.1371/journal.pone.0017782.t001

\textbf{Discussion}

We present the successful adaptation of the transformation-associated recombination (TAR) method [22] to isolate sets of individual \textit{var} genes from the \textit{P. falciparum} ITG2F6 strain. We obtained a library of 205 independent TAR clones, 98 (48\%) of which tested positive for the presence of \textit{var} gene sequence using the most conserved DBL1alpha domain in colony PCR screening. The low percentage of positives is a reflection of the limited conservation (around 50\%) of the diagnostic DBL1alpha domain. PCR screening is adequate for first and fast results but is not reliable for the detection of the entire repertoire in the library. Using Southern blot analysis on a subset of 104 yeast clones 95 putative positives (91\%) were detected, indicating that the library is likely to contain almost exclusively total or partial \textit{var} genes. To further analyze these clones, 36 TAR clones were selected, transformed into \textit{E. coli} cells and plasmids analyzed by restriction fragment length polymorphism (RFLP) and sequencing. As proof of principle, we further showed that pooling and bar-coding of three clones containing \textit{P. falciparum} \textit{var} genes allowed the subsequent \textit{de novo} reassembly after deep sequencing (Illumina).

The use of yeast artificial chromosomess to clone large fragments of DNA dates back to the late 80s when Silverman and colleagues cloned segments of human chromosomes and specifically members of gene families [22,31,32]. Later, this technique was successfully applied to specifically clone human centromeric regions [25,24]. TAR cloning was also used for gap closure in difficult areas in the human genome project [25]. With the development of higher sequencing capacity, it became feasible to target specific gene families by TAR cloning and an excellent example was the specific cloning of \textit{T. brucei} brucei, \textit{T. gambiense} and \textit{T. equiperdum} \textit{vsg} expression sites [26,27].

The majority of \textit{var} genes of \textit{Plasmodium falciparum} are located in subtelomeric locations which represent approximately 10\% of the whole genome. These areas are highly diverse and thought to undergo frequent recombination [33]. In addition, \textit{var} genes are composed of repetitive domains with a high AT content. Sequencing projects struggle with the assembly of telomeric ends due to the limited information content in small sequence fragments from these areas. This poses large problems in the correct assembly of large multi-gene family genes such as \textit{var} genes. Here we have shown that TAR cloning, which would generate clones containing only a single full-length \textit{var} gene would greatly facilitate sequencing of the \textit{var} gene repertoire of individual isolates overcoming the many ambiguities associated with such repetitive stretches.

To prove the feasibility and the principle of this approach we aimed at the specific isolation of \textit{var} genes from the ITG2F6 parasite strain. We searched for conserved regions in the 5’ upstream regions which we used to define three groups and which show a relatively high similarity within but not between the groups [12,13]. Based on the identified sequences we constructed for each of the groups a TAR cloning vector. We obtained 46 clones of group A \textit{var} genes, 91 of group B \textit{var} genes, and 68 clones of group C \textit{var} genes and this distribution is broadly similar to the observed distribution of \textit{var} genes in \textit{P. falciparum} strains 3d7, 14, and HB3 [13,30] (Supplemental Table 1). Sequence analysis of a number of obtained TAR clones demonstrated that most of the isolated \textit{var} genes derived from the group which the cloning vectors had targeted, indicating that the selected targeting sequences occurred sufficiently specific to isolate \textit{var} genes of a defined group. The one exception observed (one B clone obtained with the upsB vector) might be due to some alignment between sequences used as target sites for upsB and upsC \textit{var} genes, which exhibit 44\% identity.
Reliable screening of positive clones by PCR from colonies directly proved to be difficult. PCR amplification of a 400 bp fragment within the DBL1alpha domain, which is present in all known var genes (except in var2), gave a significant number of false negative clones (around 20%) based on subsequent Southern blot analysis. Although PCR tolerates some diversity, sequence conservation between DBL domains of as little as 52% [17] might be responsible for the missed sequences, particularly from colonies directly. In contrast, Southern blotting of total genomic or plasmid DNA identified many more potential positive clones. We considered all TAR clones with a size $> 6.5$ kb as positive and demonstrated that all clones with total sizes above $9$ kb contained var genes as confirmed by sequencing of the 5' and 3' ends of the insert.

RFLP analysis and sequencing of the 5' and 3' termini indicated that we isolated a broad range of different var genes. However, BLAST analysis using the sequence fragments and RFLP pattern also suggested some clones to be identical. The 5' sequences of four TAR clones, i.e. upsBI_23 and upsCII_12 could not be determined consequently var group affiliation could not be determined. n.d. = not done.

| TAR clone | Blast result 5' sequence (gene-ID) | Blast result 3' sequence (gene-ID) | var group | TAR clone size$^b$ | TAR clone size$^b$ |
|-----------|------------------------------------|------------------------------------|-----------|-------------------|-------------------|
| upsAI_2   | IT4_var3                           | IT4_var3                           | upsA      | 9.5 kb            | 8 kb              |
| upsAI_9   | PFD00020c                          | VARC2                              | upsA      | 9.5 kb            | 8.6 kb            |
| upsAlll_10| TM180var2                          | IT4_var54                          | upsA      | 15 kb             | 15.5 kb           |
| upsBI_1   | IT4_var54                          | IT4_var54                          | upsB      | 13.5 kb           | 12 kb             |
| upsBI_2   | MAL7P1.55                          | PFD0005w                           | upsBh     | 13.5 kb           | 11 kb             |
| upsBI_6   | IT4_var41                          | IT4_var26                          | upsB      | 16 kb             | 16 kb             |
| upsBI_8   | Var13.1                            | IT4_var28                          | upsB      | 17 kb             | 17 kb             |
| upsBI_11  | Dd2_LT141                          | IT4_var13                          | upsB      | 16 kb             | 15 kb             |
| upsBI_12  | IT4_var26                          | IT4_var26                          | upsB      | 13 kb             | 12 kb             |
| upsBI_14  | IT4_var13                          | Dd2_LT141                          | upsB      | 16 kb             | 17 kb             |
| upsBI_17  | IT4_var11                          | IT4_var11                          | upsB      | 14 kb             | 13.5 kb           |
| upsBI_21  | var-3                             | var-3                              | upsB      | 15.5 kb           | 14 kb             |
| upsBI_23  | -                                 | IT4_var24                          | ?         | 10 kb             | 11.5 kb           |
| upsBI_29  | IT4_var13                          | Dd2_LT141                          | upsB      | 16.5 kb           | 15 kb             |
| upsBI_31  | var (EU787536)                     | IT4_var24                          | upsB      | 13.5 kb           | 10.5 kb           |
| upsBI_32  | PFEMP1(PFU31083)                   | FCR3-varT11-1                      | upsB      | 15 kb             | 15 kb             |
| upsBI_33  | var (EU787596)                     | IT4_var24                          | upsB      | 10 kb             | 11 kb             |
| upsBI_35  | var-3                             | PFD0005w                           | upsB      | 13 kb             | 10 kb             |
| upsBI_38  | IT4_var54                          | IT4_var54                          | upsB      | 13 kb             | 15 kb             |
| upsBI_41  | var-3                             | var-3                              | upsB      | 15.5 kb           | 13 kb             |
| upsBI_61  | var (AY462592)                     | IT4_var36                          | upsB      | 13 kb             | 13 kb             |
| upsBI_1   | IT4_var33                          | IT4_var33                          | n. d.     | 12 kb             |                   |
| upsBI_3   | var13.1                           | var13.1                            | upsB      | n. d.             | 12.7 kb           |
| upsCI_3   | IT4_var23p                         | PF07_0049                          | upsC      | 28 kb             | 12.5 kb           |
| upsCI_10  | Var (EU787639)                     | PF09995c                           | upsC      | 23 kb             | 10 kb             |
| upsCI_7   | IT4_var23p                         | PFL1960w                           | upsC      | 14 kb             | 12 kb             |
| upsCI_13  | IT4_var23p                         | var13.1                            | upsC      | >30 kb            | 12.5 kb           |
| upsCI_6   | PF07_0051                          | PFL1970w                           | upsC      | n. d.             | 12.5 kb           |
| upsCI_14  | PF07_0051                          | IT4_var24                          | upsC      | n. d.             | 13 kb             |
| upsCI_10  | IT4_var40                          | PFD1015c                           | upsB      | n. d.             | 13.5 kb           |
| upsCI_25  | IT4_var23p                         | PF07_0049                          | upsC      | n. d.             | 12.5 kb           |
| upsCI_12  | -                                 | VarC2                              | ?         | 13 kb             | 12.5 kb           |

var groups were determined by sequence alignment of targeting sites and from 5' BLAST analysis. 5' sequences of clones upsBI_23 and upsCII_12 could not be determined consequently var group affiliation could not be determined. n.d. = not done.

doi:10.1371/journal.pone.0017782.t002

Table 2. Comparison of selected TAR clones by Sanger sequencing of 5’ and 3’ ends of inserts and size analysis by Southern blot (a) and RFLP (b).
TAR Cloning of P. falciparum var Genes

Methods

P. falciparum strain
Parasites of P. falciparum strain ITG2F6 were cultured at 5% hematocrit as previously described [37], using RPMI medium supplemented with 0.5% AlbuMax [38]. Cells were lysed in 0.03% saponin in PBS for 10 minutes on ice, centrifuged at 4000 × g for 10 minutes and washed in PBS until the supernatant was clear. The parasite containing pellet was subjected to proteinase K digestion and genomic DNA was subsequently prepared by phenol/chloroform precipitation [39].

S. cerevisiae strain
The S. cerevisiae strain TYC1 (genotype: MATα, ura3-52, leu2Δ1, cyh2Δ), containing a knockout of DNA Ligase 4) was used as the yeast host in TAR experiments as previously published [26].

Identification of var gene positive TAR transformants
TAR transformants, grown selective medium, were initially screened by colony PCR using primers for the DBL1alpha domain (Table 1). Single yeast colonies were suspended in a mix of 10 μM pF_DBL1alpha, 10 μM pR_DBL1alpha (Table 1), 0.4 μl Taq and 0.9 μl 11.1xPCR-buffer (45 mM Tris-HCl (pH 8.8), 11 mM ammoniumsulphate, 4.5 mM MgCl2, 6.7 mM 2-mercaptoethanol, 4.4 mM EDTA (pH 8.0), 113 μg/ml BSA, 1 mM dATP, dGTP, dCTP and dGTP) and adjusted with H2O to a final volume of 10 μl. PCR conditions were as follows: 5 min at 95°C, 40 cycles of

TAR procedure
The TAR cloning procedure was essentially performed according to Kouprina and Larionov [29] with the following modifications: Yeast cells were incubated with zymolyase solution (10 μg/ml) for 30 min at 30°C, for each transformation reaction 450 μl of yeast spheroplast solution was used, 4.5ml PEG 8000 solution was used and spheroplasts were resuspended in 1ml SOS solution. Spheroplasts from each transformation reaction were mixed with 15 ml melted synthetic complete agarose medium lacking uracil but containing 1 M Sorbitol and 10 μg/ml cycloheximide and poured onto plates containing the same medium. This allows for positive selection of the TAR vector that provides URA3 function. The addition of cycloheximide prevents transformation of empty plasmids, which remain circular due to incomplete Cbl and SaI restriction digestion, as those vectors still maintain the CH2 gene that confers dominant sensitivity to cycloheximide. A full manual can be downloaded from www.nottingham.ac.uk/biology/people/louis/principleoftarcloning.
30 sec 94°C, 40 sec 57°C and 2 min 72°C and a final step of 5 min at 75°C.

All yeast TAR clones were propagated on selective medium and yeast genomic DNA was isolated by phenol/chloroform purification (as described in [20]) with subsequent precipitation. Alternatively, yeast plasmid DNA was directly purified using the Zymoprep II™ Yeast Plasmid Miniprep Kit (Zymo Research) or the QIAprep Spin Miniprep Kit (Qiagen).

Southern Blot analysis
Genomic DNA of yeast TAR clones was digested with NotI and separated on 1% agarose gels in 0.5 x TBE using a CHEF mapper system (BIORAD) with the following conditions: forward: 9 V/cm, 10 s initial switch, 40 s final switch, reverse: 6 V/cm, 10 s initial switch, 40 s final switch, for 24 hrs at 14°C.

Southern blot, hybridization and detection were performed as per manufacturer’s instruction for non-radioactive nucleic labeling and detection (Roche Diagnostics) using Blocking reagent (Roche), CDP-Star (PerkinElmer or Roche), Maleic Acid (Sigma). The probe consisted of fluorescein-labeled vector backbone (pGEM7, Promega) using Fluorescein-High Prime labeling kit (Roche) and was immunologically detected using Anti-fluorescein-AP, Fab fragments (Roche) on positively charged nylon membranes (Roche).

Transformation of positive yeast clones into E. coli and Restriction Fragment Length Polymorphism (RFLP) analysis
Genomic DNA of individual yeast clones was prepared as described above and transformed into E. coli PMC103 either by electroporation or by chemical transformation into ultra competent DH5α cells. Plasmid DNA from bacterial cells was purified using the QIAprep Spin Miniprep Kit (Qiagen).

For RFLP analyses, purified plasmid DNA, of individual clones, were digested with NotI and NdeI. The digests were separated on a 0.7% agarose gel and visualized by ethidiumbromide staining. The vector backbone contains one restriction site each for NotI and NdeI. The digests were separated on a 0.7% agarose gel and visualized by ethidiumbromide staining. The vector backbone contains one restriction site each for NotI and NdeI.

Sequence analysis
All alignments were performed with the ClustalW multiple alignment tool (www.ebi.ac.uk/Tools/clustalw2) at default parameters. BLAST analyses were performed against the NCBI nucleotide database or against the non-redundant protein sequences database at NCBI. A The open reading frame of three complete sequences were analyzed using the translation tool of the ExPASy proteomics server (www.expasy.ch) from the Swiss Institute of Bioinformatics (SIB) [41] and the InterProScan Sequence Search of EBI-MBL-EBI.

Supporting Information
Figure S1 Gap closure of clone upsCI_3. (A) Diagram of the structure of the clone with gap and PCR primers designed shown. (B) PCR product covering the gap. (C) Sequence analysis of the PCR product showing AT richness. (TIF)

Table S1 Distribution of var genes within the three var groups upsA, upB and upsC obtained by the TAR experiment compared with P. falciparum strains IT4/25/4, 3d7 and HB3 [19,30]. * For the 3d7 strain, the number of upsB var genes include B/A and B/C var groups [19]. (DOCX)

Acknowledgments
We thank members of the HPB and EJL laboratories for comments and discussion.

Author Contributions
Conceived and designed the experiments: EJL HPB. Performed the experiments: AG MMB TB. Analyzed the data: AG MMB CDS. Contributed reagents/materials/analysis tools: EJL HPB. Wrote the paper: AG MMB CDS TB EJL HPB.

References
1. World Health Organization (2008) The world health report 2008: primary health care now and more than ever. Available at: http://www.who.int/whr/2008.
2. Banerje DE, Gomory JA, Ma C, Howard RJ, Pasloke BL (1996) Plasmodium falciparum erythrocyte membrane protein 1 is a paraxialized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. Proc Natl Acad Sci U.S.A 93: 3497–3502.
3. Jensen ATR, Magistrado P, Sharp S, Joergensen L, Lavstsen T, et al. (2004) Plasmodium falciparum associated with severe childhood malaria preferentially expresses PFEMP1 encoded by group A var genes. J Exp Med 199: 1179–1180.
4. Chookajorn T, Ponsuwanna P, Cui L (2008) Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. Trends Parasitol 24: 455–461.
5. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, et al. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell 121: 25–36.
6. Bull PI, Lowe BS, Kortok M, Molyneux CS, Newbold CI, et al. (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med 4: 358–360.
7. Bull PI, Lowe BS, Kortok M, Marsh K (1999) Antibody recognition of Plasmodium falciparum erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. Infect Immun 67: 733–739.
8. Kyys SA, Kramer SM, Smith JD (2007) Antigenic variation in Plasmodium falciparum: gene organization and regulation of the var multigene family. Eukaryotic Cell 6: 1511–1520.
9. Kraemer SM, Smith JD (2006) A family affair: var genes, PEMP1 binding, and malaria disease. Curr Opin Microbiol 9: 374–380.

10. Voss TS, Kaestli M, Vogel D, Bopp S, Beck H (2003) Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. Mol Microbiol 46: 1593–1607.

11. Voss TS, Healer J, Marry AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. Nature 439: 1004–1008.

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

13. Lavstsen T, Salanti A, Jensen ATR, Hermsen CC, Salanti A, et al. (2006) Expression of Plasmodium falciparum var genes in Plasmodium falciparum var gene subgroups in a case-control study. J Infect Dis 193: 1567–1574.

14. Falk N, Kaestli M, Qi W, Ott M, Baea K, et al. (2009) Analysis of Plasmodium falciparum var genes expressed in children from Papua New Guinea. J Infect Dis 200: 347–356.

15. Roux JA, Kyes SA (2004) The role of Plasmodium falciparum var genes in malaria. Mol Microbiol 53: 1011–1019.

16. Row J, Kyes SA, Papenbrock A, Baez A, Raza A (2002) Identification of a conserved Plasmodium falciparum var gene implicated in malaria in pregnancy. J Infect Dis 185: 1207–1211.

17. Voss TS, Magistrado P, Salanti A, Jensen A, Jensen ATR, et al. (2005) Expression of Plasmodium falciparum erythrocyte membrane protein 1 in experimentally infected humans. Malar J 4: 21.

18. Kyriacou HM, Stone GN, Chalfs RJ, Raza A, Lyke KE, et al. (2006) Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitemia. Mol Biochem Parasitol 150: 211–218.

19. Rottmann M, Lavstsen T, Magova JP, Kaestli M, Jensen ATR, et al. (2006) Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum falciparum infection in Tanzanian children. Infect Immun 74: 3904–3911.

20. Lainson R, Arnot DE, Theander TG (2003) Subcloning of Plasmodium falciparum 3DF7 var genes based on sequence analysis of coding and non-coding regions. Malar J 2: 27.

21. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Mago-Nabias F, et al. (2007) Population genomics of the immune evasion (var) genes of Plasmodium falciparum. PLoS Pathog 3: e34.

22. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419: 498–511.

23. Young R, Taylor JE, Kurioka A, Becker M, Louis EJ, et al. (2008) Isolation and analysis of the genetic diversity of repertoires of VSG expression site containing telomeres from Trypanosoma brucei gambiense, T. b. brucei and T. equiperdum. BMC Genomics 9: 385.

24. Kouprina N, Larionov V (2006) TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. Nat Rev Genet 7: 805–812.

25. Kouprina N, Larionov V (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat Protoc 3: 371–377.

26. Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, et al. (2007) Patterns of gene recombination shape var gene repertoires in Plasmodium falciparum: comparisons of geographically diverse isolates. BMC Genomics 8: 45.

27. Silverman GA, Ye RD, Pollock KM, Sadler JE, Korsmeyer SJ (1989) Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 16q21.3. Proc Natl Acad Sci U.S.A 86: 7483–7489.

28. Kouprina N, Graves J, Cancilla MR, Rosnick MA, Larionov V (1997) Specific isolation of human rDNA genes by TAR cloning. Gene 197: 269–276.

29. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, et al. (2000) Frequent ectopic recombination of var gene promoters in telomeric chromosome clusters of P. falciparum. Nature 407: 1018–1022.

30. Petter M, Haagström M, Khatib A, Fernandez V, Klinik M, et al. (2007) Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. Mol Biochem Parasitol 156: 51–61.

31. Trager W, Jenson JB (1976) Cultivation of malarial parasites. Nature 263: 420–421.

32. Kyes SA, Kant J, Acheson N, Cotten M, Cuff J, et al. (2002) Comparative analysis of the genetic diversity of var gene subgroups in a case-control study. J Infect Dis 193: 1567–1574.

33. Trager W, Jenson JB (1976) Cultivation of malarial parasites. Nature 263: 420–421.

34. Dorn A, Smith DJ, Koeppe J, Stoffel F, Kier C, et al. (1998) Stevored and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 97: 161–176.

35. Voss TS, Healer J, Marry AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. Nature 439: 1004–1008.

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

37. Trager W, Jenson JB (1976) Cultivation of malarial parasites. Nature 263: 420–421.

38. Dorn A, Stoffel R, Matile H, Bubendorf A, Ridley RG (1995) Malarial haemoglobin supports haem polymerization in the absence of protein. Nature 374: 269–271.

39. Beck H (2002) Extraction and purification of Plasmodium parasite DNA. Methods Mol Med 72: 159–163.

40. Raymond CK, Sims EH, Olson MV (2002) Linker-mediated recombinational subcloning of large DNA fragments using yeast. Genome Res 12: 190–197.

41. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31: 3784–3788.