Recombination and Diversification of the Variant Antigen Encoding Genes in the Malaria Parasite \textit{Plasmodium falciparum}

LAURA A. KIRKMAN$^{1,2}$ and KIRK W. DEITSCH$^2$

$^1$Department of Internal Medicine, Division of Infectious Diseases, Weill Medical College of Cornell University, New York, NY, 10065; $^2$Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY, 10065; Kirk W. Deitsch 1300 York Avenue, Box 62, New York, NY 10065

ABSTRACT The most severe form of human malaria is caused by the protozoan parasite \textit{Plasmodium falciparum}. These parasites invade and replicate within the circulating red blood cells of infected individuals leading to numerous disease manifestations, including severe anemia, altered circulation, and tissue inflammation. Malaria parasites are also known for their ability to maintain a chronic infection through antigenic variation, the ability to systematically alter the antigens displayed on the surface of infected cells and thereby avoid clearance by the host’s antibody response. The genome of \textit{P. falciparum} includes several large, multicopy gene families that encode highly variable forms of the surface proteins that are the targets of host immunity. Alterations in expression of genes within these families are responsible for antigenic variation. This process requires the continuous generation of new antigenic variants within these gene families, and studies have shown that new variants arise through extensive recombination and gene conversion events between family members. Malaria parasites possess an unusual complement of DNA repair pathways, thus the study of recombination between variant antigen encoding genes provides a unique view into the evolution of mobile DNA in an organism distantly related to the more closely studied model eukaryotes.

INTRODUCTION

Antigenic variation is of great importance for the success and survival of various pathogens ranging from trypanosomes to bacteria, fungi, and the focus of this paper, \textit{Plasmodium falciparum}, the most virulent of the human malaria parasites (1). For each pathogen, the pressure to diversify surface proteins exposed to the immune system is counterbalanced by the need to preserve function, which in the case of \textit{P. falciparum} is the maintenance of binding capacity to receptors on vasculature endothelial cells (2). Each pathogen has developed a systematic method to diversify surface proteins while balancing these strong but opposing selection pressures. This typically involves the generation of DNA sequence modifications to the genes that encode the surface proteins in ways that generate diversity without compromising function. These changes are created using the particular complement of DNA recombination and repair pathways present within the pathogen. Due to the critical nature of maintaining DNA integrity, DNA repair pathways are highly conserved across species from bacteria...
to mammals and components of most pathways can be readily identified in various organisms (3), making DNA recombination/repair a subject of interest both for evolutionary biologists as well as for those interested in host-pathogen interactions.

With the increasing availability of whole genome sequences from an ever increasing number of organisms, it is becoming clear that presence and preference for the use of specific DNA repair pathways does in fact vary considerably among different species. A well described example is the preference for which mechanism is utilized to repair DNA double-strand breaks (DSBs); higher eukaryotes show a pronounced propensity to use non-homologous end joining (NHEJ) while the yeast Saccharomyces cerevisiae exhibits a strong preference for homologous recombination (HR), though both repair pathways are complete and present in these organisms (4). Recently it was discovered that some lineages of very divergent lower eukaryotes have lost components of DNA repair pathways (5, 6, 7, 8), resulting in strong propensities toward certain types of DNA repair products and thus potentially influencing how these genomes change over time. Since DNA rearrangements play a crucial role in the diversification of important surface antigens, the availability and preferential use of different DNA repair pathways should be considered a key component of the process of antigenic variation. While DNA recombination and repair have long been considered to be intimately tied to antigenic variation in organisms like African trypanosomes (9, 10), they are just recently becoming appreciated for malaria parasites. The unique complement of repair pathways present in malaria parasites are thus likely to be tied to pathogenesis as well as the generation of antigenic diversity.

ANTIGENIC VARIATION AND IMMUNE EVASION BY MALARIA PARASITES

P. falciparum is a eukaryote parasite with a complex life cycle involving both a human host and a mosquito vector (Fig. 1). The pathogenic stage of the parasite’s life cycle involves invasion and asexual replication within red blood cells, a process that takes approximately 48 hours and results in the release of 20 to 30 daughter merozoites from each infected cell. These merozoites go on to invade new red blood cells, thus establishing a cyclic infection. Occasionally parasites differentiate into sexual stages (gametocytes) that when taken up by a mosquito further differentiate into male and female gametes, mate, and after several complex stage transformations, form new infectious cells (sporozoites) that migrate to the mosquito salivary gland, ready to establish a new infection with the next blood meal. It is only during the brief period in the mosquito midgut when the parasites are diploid and undergo meiotic recombination. With the mosquito’s next blood meal the now haploid sporozoites leave the salivary gland, enter the human blood stream and make their way to the liver where they transform to the liver stage of the life cycle. For the remainder of the life cycle the parasites remain in the haploid state. Within the hepatocytes the parasites undergo many rounds of DNA replication and division in a process termed merogony. Once the liver stage is completed, the parasites exit the liver and reestablish the erythrocytic cycle.

As with all pathogens, the ability to maintain a chronic infection requires the organisms to avoid clearance by the immune response of the infected individual. Modulation of the host immune system, stage transitions between dormant and active states, and antigenic variation are all important mechanisms that pathogens use for persistence. The antibody response is a major component of metazoan immunity and enables host organisms to recognize and destroy foreign antigens. Thus many infectious organisms have evolved specialized pathways devoted specifically to this problem. The significance of avoiding immune recognition is also reflected in how the genomes of pathogens have evolved. For example, African trypanosomes have dedicated 10 to 30% of their genomes to the multicopy gene families implicated in antigenic variation (11).

In the case of Plasmodium, being surrounded by the membrane of a host cell that is largely metabolically inactive renders the parasites almost entirely hidden from circulating antibodies, thus enabling them to mostly avoid this arm of the immune response. However, in the case of P. falciparum, growth within the red cell leads to a distorted shape and increased rigidity of the infected cell, properties that will result in its destruction within the spleen (12). To avoid splenic clearance, P. falciparum places the adhesive protein Plasmodium falciparum erythrocyte protein 1 (PfEMP1) on the red cell surface where it binds to receptors on the vascular endothelium, thereby sequestering the parasites from the peripheral circulation and avoiding movement through the spleen (13). The adhesive properties of the infected cells are also a primary virulence factor and are thought to be the cause of much of the pathogenesis associated with P. falciparum infection, including severe disease syndromes such as pregnancy-associated and cerebral malaria (14). The placement of parasite encoded proteins on the infected cell surface stimulates antibody
production against these surface antigens, leading to recognition and clearance of the infected cells. Thus variant PfEMP1 surface proteins are directly at the interface of host–parasite interactions, virulence, and the persistent nature of malaria infections.

Because PfEMP1 and other antigens expressed on the surface of *P. falciparum* infected red cells stimulate a strong antibody response, the parasites must continuously vary expression to antigenically distinct forms in order to maintain a chronic infection. Similar to many other infectious organisms, the genomes of malaria parasites contain large, multicopy gene families in which each individual gene encodes an antigenically distinct form of the surface protein (15). In *P. falciparum*, the best studied is the *var* gene family, which includes ~60 members, each encoding a different form of PfEMP1 (16, 17, 18). However *P. falciparum* also has three other multicopy gene families, called *rifin* (135 members), *stevor* (~35 members) and *Pfmc-2TM* (~15 members) (19, 20, 21). Genes from all three families are arranged in clusters or tandem arrays, primarily within the subtelomeric regions of most chromosomes. *var* genes are highly variant and different approaches to classify members of this gene family have identified three main
subfamilies based on 5’ upstream noncoding regions and the presence of particular functional domains within the encoded proteins \((22, 23)\). These subfamilies have been termed A, B, and C and expression of members of each different subfamily has been tied to different degrees of disease severity \((24, 25, 26, 27, 28, 29)\). Each \(\text{var}\) gene is complete, including a promoter and flanking regulatory regions, and genes are activated or silenced \emph{in situ}, without recombination or movement to alternative positions in the genome \((30)\). This is in contrast to antigenic variation in African trypanosomes or \emph{Babesia bovis}, in which silent genes are activated by transposition into a specific chromosomal expression site \((31, 32, 33)\). Only one \(\text{var}\) gene is actively transcribed at a time while the remainder are maintained in a silent state \((34, 35)\). Thus antigenic variation results from switches in which gene is expressed, a process that has been shown to be controlled epigenetically \((36)\).

**HYPERVERSIBILITY WITHIN THE \text{var} GENE FAMILY**

Surveys of \(\text{var}\) gene sequences from different parasite isolates have found that these genes display an extraordinary degree of diversity. While such studies detected extreme \(\text{var}\) gene diversity on a global scale, within a given location, especially one with lower transmission rates, the variability can be more limited \((37, 38)\). By comparing the degree of \(\text{var}\) gene diversity with that found within the rest of the genome, it is possible to gain insights into how the large multicopy gene families are maintained differently. Studies on parasites from areas of South America have been particularly informative because transmission rates are generally low or associated with clonal outbreaks, therefore the overall degree of genome heterogeneity is low. In addition, in certain areas malaria was close to being eradicated, thus upon reemergence the parasite genomes appear to have undergone selective sweeps \((39, 40)\). For example, a study of genome sequences from 14 \emph{P. falciparum} isolates collected in the Peruvian Amazon, an area notable for low transmission and a historic bottleneck from previous eradication efforts, detected a relatively genetically homogeneous population. Comparative microarrays found that genes annotated as “metabolic processes” displayed an average polymorphic probe frequency of only 2.6%, confirming the general lack of genetic diversity of these closely related parasites within the majority of the genome. In contrast, genes associated with antigenic variation were found to be hypervariable with a 60.2% average polymorphic probe frequency \((41)\).

Given the relative degree of sequence conservation throughout the rest of the genome, the hypervariability displayed by \(\text{var}\) genes suggests that they must be subject to much higher frequencies of mutation or recombination. Bioinformatic comparisons of large sequence sets of laboratory lines noted that \(\text{var}\) gene sequences appear to “shuffle,” predominantly between members of the same subfamily, creating mosaic sequences \((42)\). Further, \(\text{var}\) genes display a segmental organization that is reflected in “homology blocks” within the encoded PfEMP1 proteins. Relatively conserved domains are separated by hypervariable regions, potentially facilitating recombination between semihomologous genes \((43)\). While \(\text{var}\) genes in general are immensely diverse, the degree of variation is not uniform, with the most sequence diversity being observed at the 3′ end of exon 1. In contrast, exon 2, introns, and the 5′ and 3′ noncoding regions are relatively conserved \(\text{Fig. 2(A)}\) \((16, 22, 23)\).

It has also been proposed that the preferential recombination between \(\text{var}\) genes of the same subfamily is in part facilitated by the location of the genes within the subtelomeric regions of the chromosomes, with each subfamily displaying a particular orientation (transcribed either toward or away from the chromosome end; \text{Fig. 2(B)} \((42)\)). Like many other eukaryotes, the telomeres of \emph{P. falciparum} are tethered to the nuclear periphery and arranged in clusters or “bouquets,” an arrangement that brings members of each subfamily into alignment and potentially facilitates recombination \text{Fig. 2(C)} \((44, 45)\). Evidence of \(\text{var}\) gene reassortment has highlighted the likely predominance of gene conversion as a mechanism driving diversification. The resulting composites of partially conserved homology blocks are thought to balance the need for continuous generation of diversity with the constraints of maintaining functional cytoadhesive structures within the encoded forms of PfEMP1.

**\text{var} GENE RECOMBINATION – MEIOSIS VERSUS MITOSIS**

While there is clear evidence that \(\text{var}\) genes are subject to increased rates of recombination and diversification, the mechanisms underlying this characteristic are poorly understood. During transmission by the mosquito vector, parasites undergo an obligate meiotic stage and corresponding sexual recombination, providing a potential opportunity for \(\text{var}\) gene shuffling and diversification. Previous studies have highlighted that the meiotic recombination rate of \(\text{var}\) and other clustered multicopy gene families is higher than that observed for the rest of the genome. Specifically, close examination of the
FIGURE 2 Structure and genomic arrangement of the var gene family in *P. falciparum*. (A) Schematic showing the two exon structure of all var genes. Note that the 5' UTR, intron, exon 2 and 3' UTR are all highly conserved (gray). The sequence of the 5' UTR and upstream regulatory domains can be classified into three basic types called A, B and C. Exon 1 encodes the polymorphic portion of PfEMP1 and is arranged as an alternating series of highly diverse sequences separating regions of higher similarity. Overall the greatest degree of sequence diversity is found at the 3' end of exon 1 as represented by the color gradient. Exon 2 encodes a highly conserved region of PfEMP1 that is not exposed to the immune system. (B) General chromosomal arrangement of var genes, with type C genes typically found in tandem arrangement in the internal regions of the chromosomes while types A and B genes are located next to the telomere repeats. The telomeres are known to cluster into “bouquets” which align the genes in a way that is proposed to facilitate recombination and gene conversion events. (C) Illustration (left) showing the subnuclear localization of the telomere bouquets (yellow spots) that are found near the nuclear envelope within regions of dense heterochromatin (dark blue). Regions of less dense euchromatin are typically found near the center of the nucleus (light blue). Fluorescent in situ hybridization showing the location of the var gene clusters within the parasite’s nucleus. A probe that hybridizes to the conserved exon 2 of var genes is shown in yellow while the nuclear DNA is stained with DAPI (4',6-diamidino-2-phenylindole) and is shown in blue. Image in Fig. 2(B) modified with permission from reference 84. Image in Fig. 2(C) modified with permission from reference 85. doi:10.1128/microbiolspec.MDNA3-0022-2014.f2
progeny from experimental genetic crosses found that recombination between var genes is more frequent than expected when compared to the overall rate of meiotic recombination (40). In addition, examples of meiotic gene conversion events between var genes have been documented (43, 46). A recent paper proposed that DNA secondary structures within var coding regions, specifically potential hairpins formed during DNA replication, might play a role as recombination inducers (46). A probable homologue of Spo11, the endonuclease responsible for generating the DNA DSBs that initiate sexual recombination between homologous chromosomes, has been identified and is expressed in the gametocyte to ookinete stage as expected (Plasmodb.org; PF3D7_1217100). Accessory proteins known to play a role in meiotic recombination have not been identified but might simply be too divergent to be recognized by ordinary comparative analysis. It has also been observed that var gene recombination during meiosis can be ectopic (between gene copies in different chromosomal positions) (46, 45), promoting an even greater degree of potential recombination products.

In addition to meiotic recombination during transition through the mosquito, it is also possible that var genes could undergo recombination and diversification during asexual replication within the human host. While each individual erythrocytic stage parasite is initially haploid, after invasion of a red blood cell they undergo repeated genome replications and nuclear divisions in a process called schizogony, resulting in the release of 20 to 30 haploid daughter cells. These repeated rounds of replication provide ample opportunities for the occasional formation of DSBs as the DNA strands are copied. In addition, the mammalian immune response is known to include the production of numerous substances that can cause DNA damage, for example reactive oxygen and nitrogen species (47, 48, 49). Repair of DNA DSBs often includes mechanisms of HR, which when applied to multicopy gene families with large tracts of homologous sequences, could easily result in the generation of hybrid or mosaic genes. If such recombination happens at even low frequencies, given the large numbers of parasites present within an infected individual (up to 10^5 parasites per μl of blood), the probability is high that new var genes are being generated over the course of a single infection.

To investigate the frequency of mitotic recombination, studies mapped chromosomal regions from cloned laboratory lines grown in vitro as asexual parasites for prolonged periods of time. The most common finding has been large scale deletions of subtelomeric and telomeric sequences at the chromosome ends (50, 20). Because of their location within subtelomeric domains, these deletions often included the loss of members of multicycopy gene families. Similar deletions have also been noted in field isolates, thus the most frequently observed structural variation is related to genetic loss (51, 52, 53, 54). The most thorough evaluation of genetic changes resulting from long-term asexual replication was performed with clones of laboratory lines grown in vitro for up to a year, then recloned and analyzed by both microarray hybridization and whole genome sequencing (55). This study indeed detected several re combination events leading to the formation of chimeric var genes. Three events were characterized in detail and likely represented examples of break induced recombination (BIR), with the hybrid sequence extending from a break site within the var gene through the end of the chromosome. No small gene conversion events were observed in this study, although evidence for small (∼100 bp) gene conversions was found in a separate study of closely related parasite lines (56). Taken together, there is cumulative evidence for both mitotic and meiotic recombination between var gene family members, thus driving the continuous diversification and hypervariability that is a hallmark of these genes.

**THE ROLE OF DNA REPAIR PATHWAYS IN GENERATING ANTIGENIC DIVERSITY**

Recombination between DNA strands invariably initiates at the site of a DNA break. How the break is repaired therefore dictates the nature of the resulting repair product. In the case of large, variant gene families like var, the DNA repair process appears to be skewed toward recombination between family members, thereby generating the high degree of diversity observed when var gene complements are compared between different parasite lines. Given the important role that DNA repair likely plays in generating var gene diversity, a closer examination of the DNA repair pathways utilized by malaria parasites to maintain genome integrity is warranted. In eukaryotic organisms, DNA DSBs are repaired by two basic pathways: NHEJ, in which the broken DNA ends are ligated without the involvement of DNA sequences from elsewhere in the genome (57), and HR, which utilizes sequences homologous to those surrounding the break to serve as either a template to guide repair or as a site of exchange between the two DNA strands (58). NHEJ can be further classified into two distinct pathways, canonical (C-NHEJ) and alternative (A-NHEJ) (59, 60, 61). These three pathways rely on different
molecular machinery and generate different types of repair products, thus the propensity of an organism to use one pathway versus another can have a profound influence on genome evolution, and in particular on how large, repetitive, and hyper-recombinogenic regions of the genome are maintained and diversify.

Pathogens have developed varied ways to maintain genome integrity in the face of DNA damage due to both metabolic stress and immune pressure, and mechanisms of re-arrangement of the parasite surface proteins are in part dictated by DNA repair pathways active and present in the pathogen. The repair of DSB via NHEJ occurs in organisms ranging from some bacteria (although not in E. coli, for example) to mammals, indicating that this type of repair has been conserved during evolution (57). Most of the major factors involved were identified in the mammalian system, reflecting the major contribution of NHEJ to cell survival in the face of DNA damage. However, in diverse lower eukaryotes it appears there are many exceptions to the standard DNA repair pathways described in model organisms. By searching an organism’s genome to identify genes encoding required components for the different repair pathways and by examining the products of repair, it is possible to infer how different evolutionary lineages have adapted their DNA repair pathways to suit their specific needs. For example, the DNA repair proteins Ku70/80 are thought to be required for C-NHEJ, but the genes encoding these proteins cannot be identified in the genomes of the protozoan parasites Trichomonas vaginalis and Encephalitozoon cuniculi, and only Ku70 is found in the gut parasite Entamoeba histolytica (62, 6). Similarly, the kinetoplast parasites Trypanosoma brucei, Trypanosoma cruzi and Leishmania sp. all possess Ku70/80, but appear to be missing other key components required for C-NHEJ, including DNA ligase IV and XRCC4/Lif1 (63). In these organisms, the Ku proteins appear to be involved in telomere maintenance rather than DSB repair (64). The loss of key components of C-NHEJ and the resulting shift toward the use of HR to repair DNA breaks are predicted to lead to a higher frequency of gene conversion events, the resulting product when an organism utilizes a homologous sequence within the genome as a template for repair (65). In T. brucei, gene conversion from alternative or even pseudogenes is important in the creation of mosaic vsg antigens (66), and similar mechanisms have been proposed for the tick borne parasite Babesia bovis (32), a parasite that also appears to lack components of the C-NHEJ pathway.

There has been discussion in the literature regarding the presence or absence of C-NHEJ in Plasmodium (15, 7, 67). Early analysis of the complete genome sequence noted that the several key components of the pathway could not be identified. Subsequent analyses have also failed to identify the Ku proteins, Ligase 4, Artemis, DNA-PKcs, XRCC4 or Cernunnos/XLF, all key components of C-NHEJ. Remarkably, in the closely related Apicomplexan parasite Toxoplasma gondii, these components were readily identified and genetic manipulations found that NHEJ was by far the dominant pathway utilized to repair DNA DSBs (8). In fact, NHEJ was so dominant that the Ku proteins had to be knocked out, thus disabling the pathway, to enable genetic manipulations that rely on HR (68, 69, 70). This stark difference in DNA repair pathways in closely related organisms suggests a selective advantage for utilizing one type of repair and the resulting effect this has on the evolution of the genome. For example, the relatively recent loss of C-NHEJ in the Plasmodium lineage suggests a selective advantage of skewing repair toward HR.

A closer examination of the phylogenetic tree shows that amongst the Alveolates, the presence of C-NHEJ (as defined by identifiable Ku proteins) is variable (Fig. 3). Ku proteins are easily identified by using Psi blast (http://blast.ncbi.nlm.nih.gov) or Hmmer (http://hmmer.janelia.org) with the yeast or human Ku 70 core sequence as query, or alternatively using the prokaryotic core domains (71). With this strategy Ku proteins (and Ligase IV) were identified for the Alveolates Perkinsus marinus and Paramecium tetraura. However within the Apicomplexan parasites, the same search strategy only identified orthologs of these proteins in the Coccidian branch, which includes Toxoplasma, Eimeria and Neospora. A similar approach for core components of HR (Rad 51, Ligase I, Mre11) easily identified this pathway as intact across all the Alveolates. Interestingly, both Cryptosporidium and the hematozoa (Plasmodium, Babesia, Theileria) do not possess any of the components required for C-NHEJ, implying that the pathway was lost twice while being maintained by the Coccidia.

The remarkable differences in the DNA repair pathways found in these relatively closely related organisms raises interesting evolutionary questions, in particular why both the Cryptosporidium and hematozoa lineages appear to have lost the C-NHEJ pathway. Cryptosporidia are distantly related to the Coccidia and have undergone substantial reductive evolution, including the loss of both the plastid (a chloroplast like organelle distinct to Apicomplexan parasites) and mitochondrial genomes, and thus the loss of C-NHEJ might be related to this general loss of genomic material (72). Toxoplasma, the best studied representative of the Coccidia,
is an intracellular parasite that evades the immune system primarily by transitioning to a latent, intracellular cyst stage. No hyper-recombinogenic, large multicopy gene families involved in antigenic variation have been identified. In contrast, multicopy gene families with potential roles in antigenic variation have been identified in most hematozoa and it is tempting to posit that the hematozoa evolved to favor HR (and ultimately lost C-NHEJ) as part of their strategy for maintaining and diversifying these gene families. Gene conversion through HR preserves open reading frames and maintains functional domains intact, characteristics that are key to preserving large gene families in which virtually all copies are functional.

**THE POTENTIAL ROLE OF CHROMATIN STRUCTURE IN var GENE DIVERSIFICATION**

Unlike some other pathogens where antigenic variation is mediated by the transposition of silent genes or gene fragments into an active expression site (for example African trypanosomes or Babesia bovis) (31, 32, 33), antigenic variation in P. falciparum results from the *in situ* activation of previously silent var genes without recombination (30). Activation of a silent gene is always coupled to silencing of the previously active gene, resulting in strict mutually exclusive expression. The molecular mechanisms underlying this tightly coordinated activation and silencing process are only partly understood, however it has become clear in recent years that
epigenetic modifications to the chromatin surrounding the genes play a definitive role. Modifications to histone H3 at the lysine 9 position (H3K9) appear to play a prominent role, with acetylation of this residue (H3K9ac) associated with the single active gene and trimethylation (H3K9me3) found at the silent remaining members of the family (73, 74). In addition, trimethylation of the lysine at the 36th position of the same histone (H3K36me3) is found at all var genes and has been shown to be important for maintaining mutually exclusive expression (75, 76).

An unusual aspect of the histone modifications associated with var genes (H3K9me3 and H3K36me3) is that they are almost exclusively localized to genes that undergo variable expression, in particular the multicopy gene families involved in antigenic variation (77, 78). This is in contrast to most eukaryotic organisms in which these marks are found associated with genes throughout the genome. Malaria parasites appear to have dedicated these specific epigenetic marks to very narrow portions of their genomes, and the chromatin that is assembled in these regions is different than that found elsewhere. The multicopy gene families are also only found in limited regions of the chromosomes, specifically within the subtelomeric domains or within defined tandem arrays within internal chromosomal regions. Further, these chromosomal regions are tethered to the nuclear envelope and thereby anchored at the nuclear periphery, localized within distinct spots or clusters (79, 80) [Fig. 2(C)]. Collectively, these observations suggest that malaria parasites have segregated the hyper-recombinogenic regions of their genome away from the typical euchromatin that surrounds genes required for “housekeeping” functions, both along the linear length of the chromosomes and also spatially within the nucleoplasm. This model raises the interesting possibility that the unique chromatin structure found at the multicopy gene families that is known to play a role in transcriptional regulation might also be important for controlling the accessibility of templates for DNA DSB repair. If so, this would have major implications for recombination between genes and therefore the ongoing diversification of var genes.

There is growing evidence from model organisms consistent with the hypothesis that higher order chromosomal architecture exerts just as profound an influence on DNA repair as it does on nuclear processes like transcription and DNA replication (81, 82). Some differences in chromatin structure can be accounted for by chromosomal position, for example proximity to centromeres or telomeres, an observation directly relevant to recombination of multicopy genes in P. falciparum given their arrangement in mostly subtelomeric and several internal clusters. The A, B and C groups of var genes mentioned previously are distinguishable by chromosomal location as well as by 5’ UTR sequence and domain architecture, with A and B subtypes more frequently arranged in the subtelomeric regions and C types more often located in internal tandem arrays (22, 23). The aforementioned “bouquets” specifically align the different var subtypes (A genes with A genes, B genes with B genes, etc.) thus potentially facilitating the steady divergence of these families over time.

In model organisms, DSBs occurring in areas of heterochromatin are repaired with much slower kinetics than those in euchromatic regions (81). In addition, telomeric DSB repair is regulated differently than breaks found in more central chromosomal locations, in part due to the role of DNA repair enzymes in maintaining telomeric ends. For example, the addition of telomeric repeats to a plasmid integrated into an internal site dramatically decreased repair of DSBs regardless of which repair pathway, HR or NHEJ, was used (82). To repair a DSB regardless of the surrounding chromatin structure, the same general machinery is required, however in a state of heterochromatin an additional set of enzymes and histone modifiers are necessary (81). This might be particularly relevant for DSB repair at var genes due to the unique histone modifications and chromatin structure found only at these particular chromosomal regions. The possibility that the unique chromatin environment found at var genes influences the DNA repair pathways used to repair DSBs might help to explain recent data reported on DNA repair in P. falciparum. Examination of var gene sequences from large datasets provides ample evidence for frequent gene conversion events driving var gene diversification (42, 56). However, experimental induction of DSBs at a non-var site in the parasite’s genome found that sequence divergence greater than 2% resulted in exclusive utilization of a repair pathway resembling A-NHEJ, and no gene conversion events were detected (67). Since var gene sequences typically diverge by much greater than 2%, DSBs that occur within var-associated regions of the genome must be handled differently if HR induced gene conversions are indeed the driving force behind var gene recombination.

**CONCLUSIONS**

The forces driving recombination and diversification of var and other multicopy gene families in P. falciparum remain poorly understood. The domain architecture of
the gene coding regions, the chromatin structure found at these chromosomal regions, location of the genes along the length of the chromosome and within the nuclear space are all likely to be key players. To date descriptions of var gene recombination events, either those detected in the progeny of a genetic cross or arising in lab isolates grown over time, remain limited to the extent that it is difficult to account for the actual degree of diversity observed in the field. Recombination during meiosis or through BIR/gene conversion during mitotic replication are likely to contribute to overall var diversification, but our understanding of the molecular mechanism that influences these processes is far behind what is known in other pathogenic organisms including trypanosomes or various bacterial species. Acquisition of large amounts of data has also been limited by difficulties in sequencing through the most AT-rich genome yet described, and assembly of repetitive gene families poses particular difficulties for the current generation of high-throughput genome sequencing platforms that rely on massive numbers of short sequence reads. This problem has been significant enough that most genome sequences to date have excluded the multicopy gene families and focused on the core of the genome. Similar analyses using microarrays have been limited to studying closely related parasites to which the array was designed. Advancing sequencing technologies that yield much longer read lengths and an improved ability to assemble closely related, repetitive sequences will solve many of these problems. Combined with a better understanding of nuclear structure, mechanisms of DNA repair, and knowledge about how parasites handle DNA stress will combine to greatly enlighten our concept of “mobile DNA” in malaria parasites.

REFERENCES


Kirkman and Deitsch


