Evolutionary Bioinformatics

ORIGINAL RESEARCH

FORS-D Analysis in *P. falciparum* can Differentiate Classes of Genes Under Selection

Atish Kathpal[2], Balakota Reddy Patakottu[1] and Swati Patankar[1]

[1]Department of Biosciences and Bioengineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400076, India.
[2]Birla Institute of Technology and Science, Pilani, Goa Campus, Zuarinagar 403726, Goa, India.
Corresponding author email: patankar@iitb.ac.in

Abstract: FORS-D is a measure of the contribution of base order to the stem loop potential of a nucleic acid sequence and can also give information on evolutionary pressures on sequences to move away from secondary structure. Negative FORS-D values in a gene are associated with exons and nucleotide substitutions such as SNPs. An analysis of *P. falciparum* genes under selection pressure shows a correlation between negative FORS-D values and SNP density for genes that drug targets but not for drug transporters or antigenic variation genes. Analysis of the *dhfr* gene shows that a majority of rare mutations that associate with drug resistance also fall into regions with negative FORS-D values. These data suggest that FORS-D values might be predictors for drug target genes and drug resistance mutations in these genes.

Keywords: *P. falciparum*, variation, FORS-D, drug targets

Evolutionary Bioinformatics 2011:7 21–29

doi: 10.4137/EBO.S6609

This article is available from http://www.la-press.com.

Introduction

The protozoan parasite *Plasmodium falciparum* causes severe malaria resulting in millions of deaths world-wide, a problem exacerbated by the emergence of resistance to almost all known anti-malarial drugs.[1] Drug resistance has arisen through acquisition of mutations in drug targets and in drug transporters.[2] For example, mutations in the target genes dihydrofolate reductase (*dhfr*) and dihydropteroate synthase (*dhps*) led to resistance against the anti-folate drugs pyrimethamine and sulphadoxine, while mutations in transporters *pfcrt* and *pfmdr1* resulted in resistance to chloroquine and quinine.[3] Apart from specific drug resistance mutations, high frequencies of single nucleotide polymorphisms (SNPs) are seen in genes encoding drug targets, transporters and cell surface antigenic genes, suggesting that genes under selection are highly polymorphic giving rise to a survival advantage.[4] Further, genes under selection pressure also exhibit high ratios of synonymous (S) to non-synonymous (N) mutations.[5] These studies on genome-wide diversity have also identified potential new drug targets.[6]

Once a drug target is identified, new drugs developed and deployed, the high genetic diversity of *P. falciparum* could result in acquisition of new mutations, potentially resulting in drug resistance. Indeed, anti-malarials like atovaquone succumbed to this problem in disturbingly short time frames.[7] Hence, it would be of interest to find genome features that associate with polymorphisms and resistance mutations in the hope that one could predict genetic variation especially for drug target genes.

One such genome feature that has been shown to associate with genetic variations, especially for genes under selection, is a parameter known as FORS-D.[8] In brief, the ability of a DNA sequence to form a stable secondary structure can be predicted by the free energy of folding (ΔG) of the sequence, with a negative ΔG value indicating more stable secondary structure. Forsdyke has termed this ΔG value of the natural sequence as Folding of the Natural Sequence (FONS). However, negative ΔG values might be observed simply due to base composition bias such as high GC content and have little to do with the sequence *perse*. To account for this, the same sequence is randomized and the average ΔG value of the randomized sequences is computed (Folding of Randomized Sequences-Mean or FORS-M). The difference between FORS-M and FONS is termed FORS-D (Folding of Randomized Sequences-Difference). Thus FORS-D is the base order dependant stem loop potential of a particular sequence.

A positive FORS-D value for a given sequence implies a pressure towards maintaining secondary structure in that part of the genome. Conversely, regions of the genome exhibiting negative FORS-D values may be sequences where there is an evolutionary pressure to move away from stable secondary structures. Forsdyke's work has shown that coding regions are associated with negative FORS-D values compared to introns since the presence of amino acid codons clearly has a high priority over secondary structure.[9,10] Interestingly, sequences that undergo substitutions in genes under selection also show negative FORS-D values, implying that selection pressure can override secondary structure constraints.[8,11]

In this report we ask whether FORS-D is correlated with SNPs in *P. falciparum* genes. For this analysis, a modified version of the Random Scan software[12] was developed. Using this software called SNP Scan, we analyzed genes that are associated with drug resistance ie, *dhfr*, *dhps*, *pfmdr1* and *pfcrt* and genes from the family of antigenic variable proteins (*var*) that are under selection pressure from the host immune response. The results of the analysis show that for genes encoding drug targets there is a negative correlation between FORS-D values and variation (SNPs, drug resistance mutations and rare drug resistance mutations) reinforcing similar data from previous studies. These data suggest that FORS-D is a feature that can be used to predict gene sequences that are hotspots for substitutions and mutations. Surprisingly, genes that encode transporters associated with drug resistance do not show negative correlations between FORS-D and SNPS that are seen for genes that encode drug targets; similarly most genes encoding variable surface antigens also do not show negative correlation between SNPS and FORS-D. Similar results were obtained with Random Scan software. Hence unlike SNP density and non-synonymous SNPs, FORS-D correlation with SNPs appears to be specific for drug target genes. To test this hypothesis we analyzed a gene that has been implicated in resistance to artemisinin, the *P. falciparum ATPase* 6 gene. This gene also shows a high negative correlation between SNPs and FORS-D values. We propose that FORS-D

analysis of the genome may reveal targets of drugs such as chloroquine; furthermore, FORS-D analysis of novel drug target genes may reveal early insights into drug resistance.

Materials and Methods

DNA sequences and SNPs

All DNA and SNP data was downloaded from PlasmoDB.[13] The *dhfr-ts*, *pppk-dhps* (PF08_0095), *pfmdr1* (PFE1150w), *pfcrt* (MAL7P1.27) genes, *pfatpase6* (PFA0310c) and 8 *var* genes (Mal13p1_405, PF11_0008, PFD0020c, PFB0020c, PFF1580c, PFF0845c, PFL0935c) were chosen for individual FORS-D analyses. DNA sequences of these genes were downloaded from PlasmoDB version 6 (www.plasmodb.org). All sequences used for FORS-D analysis were from the reference 3D7 genome.

SNPs and their location on each DNA sequence were also downloaded from PlasmoDB version 6. For assignment of SNPS to FORS-D windows, the location of each SNP was noted as were the number of strains each SNP was found in. Of note, all strains from PlasmoDB were selected during the download, including *P. reichenowi* since inclusion of this primate parasite gave higher numbers of SNPs and statistical power to the analysis. Further, the genome of *P. reichenowi* is is closest to *P. falciparum* in terms of phylogenetic distance and many genes are conserved. The total number of SNPs in a window was the sum of all SNPs found in different strains.

FORS-D and statistical analysis

For the purpose of carrying out FORS-D analysis, SNP Scan 2.0 (beta) software was developed. DNA sequences without introns ie, cDNA sequences were downloaded from PlasmoDB and were fed as input to SNP Scan 2.0. The software reads the sequence in steps of windows (each 200 nts wide) sliding forward by 50 nts. Hence, consecutive windows overlap by 150 nts. The software calculates FONS values for each of these windows by folding 100 nts against each other from within the 200 nts long window. FORS-M values are calculated by randomizing the natural sequence and taking the average of fold energies, all of which is coded into the software. The code generates 1000 randomized sequences, for the sake of accuracy, although 10 randomizations are known to serve the purpose.[12] The FORS-D values are then evaluated by the software by means of the following equation:

$$\text{FONS} = \text{FORS-M} - \text{FORS-D}$$

The stem loop potential associated with a given sequence is approximated by means of Tinoco's stability numbers. A-T base paring is assigned a value of 1 units and G-C base pairing in the given sequence is assigned value of 2 units. A sum of these pairings for a given stretch of nucleotides gives the stem loop potential (FONS) for the given sequence.

SNP Scan 2.0 was developed as the long term goal of this study was to use the software on genomic sequences, although this preliminary analysis used individual genes. SNP Scan 2.0 does not get into time consuming calculations of actual base stacking and loop destabilization energies of secondary structures. Instead, it draws a robust relative picture of the variations in FORS-D values of various windows, based on the sequence specific A-T and G-C Tinoco's stability numbers. The Tinoco stability number approximations,[14] which our software uses, are well documented.[15] As a result, SNP Scan 2.0 has a faster turnaround time than other software. The SNP scan code is available at http://www.bio.iitb.ac.in/~patankar/

Since the software generated FORS-D values for overlapping windows, for each window we assigned a number that was the average of the window *viz.* window 1–200 was assigned a value of 100, while window 50–250 was assigned a value of 150, etc. For correlation analysis, for each DNA sequence the number of SNPs and the corresponding FORS-D value was computed. Correlation between these two variables was performed using the statistical features in Microsoft Excel spreadsheets, SPSS and MiniTab software packages.

Results

SNPs and drug resistance mutations associate with negative FORS-D values for *dhps* and *dhfr*

FORS-D has been shown to correlate with SNPs in genes under selection for snake venom and MHC genes and for retroviral genomes.[8,11,16] To test whether this was true for *P. falciparum* drug resistance genes, two drug target genes *viz. dhps* and *dhfr* were analyzed. The proteins encoded by these genes are the only two validated drug targets of anti-malarial drugs known to date; although it would have been optimal to analyze

several drug target genes, the lack of such numbers precluded a wider analysis. The *pppk-dhps* locus was subjected to FORS-D analysis with SNPs and drug resistance mutations mapped onto the FORS-D graph (Fig. 1A). Both positive and negative FORS-D values as well as numerous SNPs can be seen through the length of this gene.

A large number of SNPs are associated with negative FORS-D regions in this gene. The majority of drug resistance mutations (S436, A437, A581) were also associated with negative FORS-D regions of *pppk* and *dhps*. These data indicate that both SNPs and drug resistance mutations associate with regions of negative FORS-D in the *dhps* gene. A similar analysis was performed for the *dhfr-ts* gene (Fig. 1B). In contrast to *pppk-dhps*, SNPs are found predominantly in *dhfr* and not in *ts*. Nevertheless, SNPs were associated with negative FORS-D regions. The five most commonly studied drug resistance mutations in the *dhfr* gene are A16, V51, C59, S108 and I164 as combinations of these mutations result in varying levels of pyrimethamine resistance in biochemical enzyme assays, in vivo drug resistance assays and clinical phenotypes.[17,18] Four out of five drug resistance mutations (A16, V51, C59, S108) were also associated with regions of negative FORS-D.

Recent reports have identified rare drug resistance alleles of *dhfr* by isolation of the gene from clinical samples and functional complementation of a yeast strain with a *dhfr* deletion.[19,20] We asked whether these the mutations identified in these rare alleles also map to negative FORS-D regions of the *dhfr* gene. Results shown in Figure 1B (mutations shown in black) indicate that although the rare drug resistance alleles are found at distinct positions on the gene, four out of six rare mutations map to regions with negative FORS-D values. Similarly, a novel SNP (S587F) in the *dhps* gene[21] also maps to a negative FORS-D window (window 1950–2000) as indicated in Figure 1A.

Negative correlation between SNP density and FORS-D for drug resistance genes

Having shown that SNPs and drug resistance mutations are associated with regions of negative FORS-D for two drug target genes, we next asked whether there is a statistical correlation between SNP density and FORS-D values for drug resistance genes. Correlation would suggest a linear relationship between the two variables and although no indication of causality, might allow the use of one variable (FORS-D) to predict the second variable (SNPs). We analyzed *dhfr-ts* and *pppk-dhps* and included *pfcrt* and *pfmdr1*, the transporters shown to confer resistance to quinoline drugs. Table 1 shows a negative correlation between SNP density and FORS-D value for two genes studied

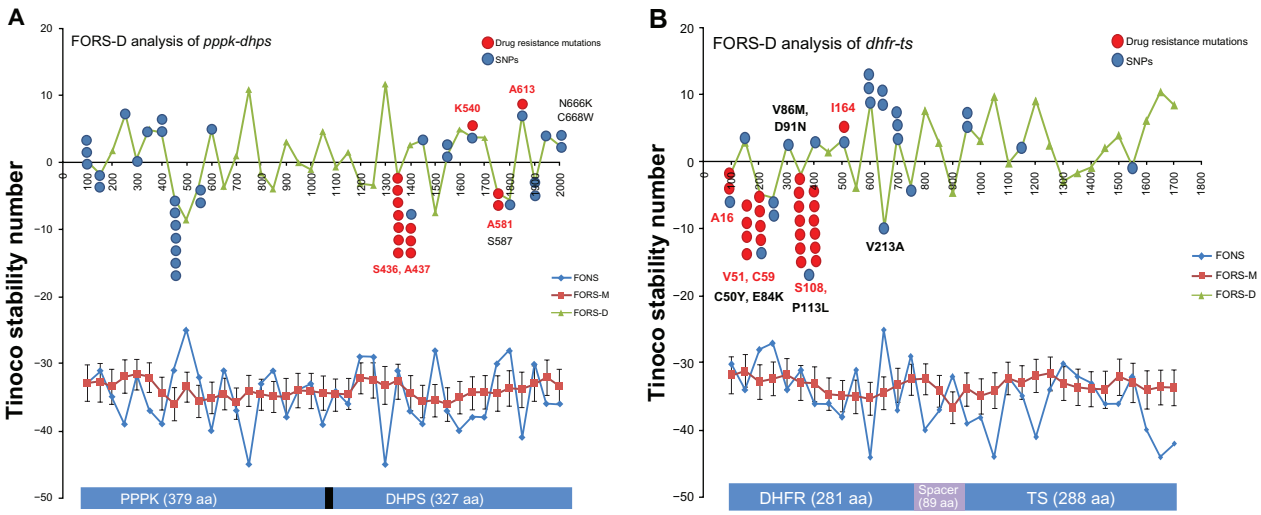

Figure 1. FORS-D analysis of the *pppk-dhps* locus (**A**) and the *dhfr-ts* locus (**B**). The y-axis shows the folding energy in Tinoco stability numbers of each 100 base window. The x-axis represents the sequence of the gene. SNPs are shown as blue circles and drug resistance mutations as red circles. Amino acid changes that have been shown to cause drug resistance are indicated in red. Rare mutations found in field isolates are shown in black.

[*dhfr* (–0.24), *dhps* (–0.17)] and a positive correlation for the *pfmdr1* gene (0.23). In contrast, no correlation between FORS-D values and SNPs was observed for *pfcrt* (0.09). Although the negative correlation values are relatively small, they are comparable with results seen in other reports that analyze FORS-D and variation.[10,16]

This result suggests that for genes which encode drug targets and hence are under selection pressure, SNP density is correlated to a numerical parameter (FORS-D) that indicates a propensity to move away from stable secondary structure. However, for other genes that are also under selection pressure such as drug resistance transporters, there is poor or positive correlation between FORS-D and SNP density. We next tested whether the *pfatpase6* gene that has been implicated in in vitro resistance to artemisinin also shows negative correlation between SNPs and FORS-D values. Interestingly, this gene also shows a negative correlation (–0.20) between these two parameters, suggesting that perhaps this gene is indeed under selection pressure from the drug.

To confirm the results, we used the Random Scan software developed by Zhang et al to compute FORS-D values of the four genes.[12] The Random Scan software subtracts the FORS-M value from the FONS value (opposite to our software) hence with this software a positive value reflects the propensity to move away from secondary structure. A similar result was obtained with the Random Scan software *viz*. drug target genes show a positive correlation between FORS-D and SNP density indicating that SNPs are found in regions that are under evolutionary pressure to move away from stable secondary structures (Table 1). In contrast the genes that encode drug transporters do not show this property. Since the results obtained by the SNP Scan 2.0 software were similar to those from Random Scan, SNP Scan 2.0 was used for the remainder of the study.

Low negative correlation between SNP density and FORS-D for *var* genes

Genes under selection show a negative correlation between SNP density and FORS-D, suggesting that sequences with low base order dependent stem loop potential are associated with higher variation. To assess whether this is true of other genes under positive selection, we performed FORS-D analysis of 8 *var* genes that encode proteins involved in antigenic variation. These 8 genes were selected based on literature that showed that these genes are indeed expressed in a mutually exclusive manner during lab culture of *P. falciparum*.[22–26]

PfEMPs are parasite-encoded proteins found on the surface of infected blood cells. By switching the PfEMP that is expressed the parasite can evade immune responses leading to some of the pathogenic phenotypes of *P. falciparum*.[27] Hence, these proteins are under selection pressure, however the response of the parasite to host immune pressure is at the level of transcriptional and epigenetic control of gene expression although *var* genes do show high frequencies of SNPs.[28] Analysis of individual *var* genes located at sub-telomeric and internal positions on different chromosomes (Table 2) indicated that most *var* genes studied did not show a strong negative correlation between FORS-D values and SNP density. However two genes (PFF1580c and PFF0845c) showed a correlation coefficient of –0.12 and –0.11. Based on results of drug targets, we speculate that these *var* genes may

Table 1. Correlation between number of SNPs and FORS-D values associated with each SNP for drug resistance genes. The four genes were analyzed using SNP Scan 2.0 software where negative FORS-D values indicate that the sequence is under evolutionary pressure to accommodate substitutions. The Random Scan software differs from SNP Scan 2.0 in that a sequence under evolutionary pressure to accommodate SNPs would show a positive FORS-D value. Pearson correlation coefficients for SNPs and FORS-D are shown as are the chromosome numbers for each gene.

Gene	Chromosome	Pearson correlation coefficient	
		SNP scan program	Random scan FORS-D program
dhfr-ts	4	–0.24	0.122
pppk-dhps	6	–0.17	0.22
pfmdr1	5	0.23	–0.061
Pfcrt	7	0.09	–0.017

be under selection to alter the DNA sequence such that the encoded PfEMP1 proteins exhibit altered binding to the host receptors. In conjunction with the data on drug target genes and drug transporters, this data suggests that correlation between FORS-D and SNP density appears to be able to differentiate the type of selection pressure on a given gene.

Discussion

The SNP Scan software shows negative correlations between FORS-D values and SNPs for *P. falciparum* drug target genes

The SNP Scan software performed similar to Random Scan at assigning FORS-D values that showed inverse correlations with the presence of SNPs in *P. falciparum* drug target genes. The SNP Scan tool is based on a time-efficient algorithm to evaluate randomizations of windows of nucleotides and perform the appropriate calculations of FONS, FORS-M and FORS-D values, in that particular order. It approximates and predicts the relative differences in the folding energies based on Tinoco stability numbers.[15] SNP Scan simplifies the procedure of evaluating folding energies, concentrating on relative variation in FORS-D values over the windows, rather than evaluating the absolute folding energies for every window and has the advantage of rapid data output.

Additionally, for the AT-rich genome of *P. falciparum* we observed that SNP Scan analysis gives relatively flat curves for FORS-M shown for *dhfr-ts* and *pppk-dhps* (Figs. 1A and 1B) and observed for all the other genes tested. This results in FORS-D values essentially giving the same trend as the FONS curve for our analysis and hence, SNPs would also show a strong correlation with FONS. As FONS is simply the folding of the natural sequence, Tinoco stability numbers from SNP Scan data suggest that *P. falciparum* SNPs are found in regions of drug target genes that show lower FONS values are hence have less stable secondary structure. These observations suggest that FONS alone may be sufficient to predict SNPs in the *P. falciparum* genome.

Negative FORS-D values may predict SNPs and drug resistance mutations

There is a negative correlation between SNP density and FORS-D values for the drug target genes *dhfr-ts* and *pppk-dhps* (Table 1) and the putative drug target *pfatpase6*. Correlation of SNPs with negative FORS-D values has been shown previously for viral genomes, snake venom genes and MHC genes.[8,11,16] Hence, it has been proposed that genes under selection have regions that show a propensity to move away from stable secondary structure. Such correlation analyses provide valuable data, however cause and effect is impossible to discern ie, are SNPs caused due to sequence constraints in regions with negative FORS-D values or conversely are the negative FORS-D values caused due to SNPs?

Results of this study indicate that for *P. falciparum*, regions of genes that have negative FORS-D values may have a higher predisposition to acquire SNPs. We propose this cause and effect relationship for the following reasons. The sequence that was analyzed for generating FORS-D values is the 3D7 strain genome sequence; 3D7 parasites are sensitive to anti-malarial drugs and their genome should not show signatures

Table 2. Correlation between number of SNPs and FORS-D values associated with each SNP for *var* genes. Chromosome number and chromosomal location (sub-telomeric and internal) are indicated. Pearson correlation coefficients were calculated for SNP density versus FORS-D values using the SNP Scan 2.0 software.

PfEMP gene	Chromosome	Pearson correlation coefficient	Chromosomal location
PF11_0008	11	0.08	Sub-telomere
PFD0020c	4	0.02	Sub-telomere
PFF1580c	6	–0.12	Sub-telomere
PFF0845c	6	–0.11	Internal
PFL0935c	12	–0.01	Internal
PF07_0049	7	–0.03	Internal
PFD1015c	4	0.00	Internal
PFB1055c	2	–0.01	Sub-telomere

of selection pressure. Secondly, SNPs were obtained from all strains of *P. falciparum* and also from *P. reichenowi* where available. If SNPs from a closely related primate parasite are also found in regions of negative FORS-D values, this implies that those regions of the drug target genes may indeed have sequence characteristics that promote variability. Thus the data suggest that *dhfr-ts* and *pppk-dhps* genes have hotspots for mutation that can be predicted by negative FORS-D values. To address this possibility, we studied rare SNPs[19,20] that have been detected in drug resistant parasites by a novel yeast complementation assay for *dhfr*. If regions of negative FORS-D in a gene do indeed predispose those sequences to higher variation, the rare SNPs should also fall into FORS-D negative windows in the *dhfr* gene. Indeed, four out of six rare SNPs (Fig. 1B), were associated with negative FORS-D windows. One rare mutation in the *dhps* gene was also associated with a negative FORS-D window (Fig. 1A). These data suggest that FORS-D may be able to predict regions in genes under selection that may be mutational hotspots; such hotspots have been suggested for *dhfr*[29] and may be true for *dhps* as well.

If the prediction that negative FORS-D values indicate mutational hotspots in genes under selection is correct, a FORS-D analysis of genes that are targets of future drugs may be able to predict hotspots for SNPs and drug resistance mutations. A case in point is the deoxy-xylulose phosphate reductoisomerase (*dxpr*) gene, the target of an anti-malarial drug, fosmidomycin that is in clinical trials. The *dxpr* gene has only 2 SNPs in PlasmoDB however a recent paper[30] has found 5 amino acid substitutions in *dxpr* (not associated with drug resistance) in clinical samples. Correlation of these with FORS-D gives a Pearson coefficient of −0.85. Although *dxpr* resistance in vitro so far appears to be through duplication of the gene,[31] we speculate that SNPs may appear in negative FORS-D regions of this gene and if these regions have amino acids that affect drug binding they may give rise to resistance. This would be similar to *pfmdr1* where increased copy number is associated with mefloquine resistance[32,33] however specific mutations have been shown to confer resistance to mefloquine, quinine and halofantrine.[34] We propose that FORS-D analysis may become part of the analysis of future drug targets.

What might be the biological significance of the negative correlation observed between FORS-D and SNPs? DNA sequences analyzed here would be double stranded throughout the life cycle of the parasite except during DNA replication where the single-stranded DNA could potentially form transient stem loops. We speculate that low FORS-D values suggest a low secondary structure context and perhaps the DNA replication machinery incorporates variations in such regions. Alternative explanations for high variation in regions of low secondary structure include potential for recombination during the meiotic stages of the parasite life cycle (sexual stages formed in the mosquito) where weak secondary structure may allow recombination.

FORS-D analysis can distinguish between different responses to selection pressure

Whole genome SNP mapping has shown that genes having a higher frequency of SNPs and a higher ratio of synonymous to non-synonymous mutations are under selection and hence may be potential drug and vaccine targets.[4,31,35] In this report we show that there is a negative correlation between FORS-D values, SNP density and known drug resistance mutations for the *dhfr-ts* and *pppk-dhps* genes that are under selection pressure. However two other genes also under selection pressure (*pfmdr* and *pfcrt*) do not show a negative correlation. Some reasons for these differences could be relative GC and purine content of the four genes, features that might result in differences in FORS-D values; however, analysis of GC and purine content showed that there is no statistical difference between these parameters in the four genes (data not shown).

Another difference between *dhfr-ts*, *pppk-dhps* and *pfmdr*, *pfcrt* is that the former are drug targets while the latter are drug transporters. This led to the hypothesis that negative FORS-D values may be associated with certain classes of genes, especially those that directly bind to ligands or proteins. To test this hypothesis, another class of genes that is also under selection pressure was studied: the antigenic variation or *var* genes. Most *var* genes showed no correlation between FORS-D and SNP density except for two genes.

Although drug targets, drug transporters and *var* genes all are under selection, there is a fundamental difference between the response of *var*

genes and drug resistance genes to selection pressure in *P. falciparum*. Drug resistance genes in the parasite respond to selection pressure by acquiring mutations[17,18] while some drug transporters such as *pfmdr* show increased copy number in response to drug pressure.[32,33] In a completely different mechanism, *var* genes respond to selection pressure through switching at the level of transcription or through epigenetic mechanisms.[27,28] It is interesting that correlations between SNP density and FORS-D values appear to differentiate between different types of responses to selection pressure.

Several studies have shown that SNP density and dN/dS ratios can reveal genes that are under selection and *var* genes fall into this category.[4] Simply looking at SNP density of a gene or dN/dS ratios does not distinguish the types of responses to selection pressure. However, data shown in this report suggests that a negative correlation between SNP density and FORS-D values indicates genes under selection pressure to acquire mutations such as the drug target genes *dhfr* and *dhps*. The observation that some *var* genes also show negative correlation between FORS-D and SNP density suggests that these *var* genes acquire mutations in response to the selective pressure imposed by the host immune system. As a future application of FORS-D analysis, genome-wide scans for genes that show high negative correlation between FORS-D and SNP density may be a strategy to identify drug targets that have yet to be discovered *viz*. the target of chloroquine. In an attempt to do precisely this, we subjected the *pfatpase6* gene to FORS-D analysis and showed that this gene also shows a negative correlation between FORS-D values and SNPs. These data suggest that the *pfatpase6* gene that has been implicated in resistance to artemisinin in vitro may indeed be the target of the drug.

In conclusion, given the rapid acquisition of drug resistance, it would be immensely useful to be able to predict sequences that are susceptible to selection pressure and undergo variation in the malaria parasite genome sequence. FORS-D analysis will be an excellent starting point to uncover genes that are under selection pressure and analyze drug target genes for signatures of variation in *P. falciparum*.

Acknowledgements

A.K. is grateful to Dr. C.K. Prakash, Assistant Professor, BITS-Pilani, Goa Campus for the introduction to the field of secondary structure analysis and guidance during the project where a simplified version of SNP Scan was developed. We acknowledge IIT Bombay for providing a summer research fellowship to A.K. during which time this project was initiated.

Disclosure

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. WHO. World Malaria Report 2008. World Health Organization. 2008.
2. Gregson A, Plowe CV. Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev*. 2005;57:117–45.
3. May J, Meyer CG. Chemoresistance in *falciparum* malaria. *Trends Parasitol*. 2003;19:432–5, discussion, 435–6.
4. Volkman SK, Hartl DL, Wirth DF, et al. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science*. 2002;298:216–8.
5. Neafsey DE, Schaffner SF, Volkman SK, et al. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol*. 2008;9:R171.
6. Kidgell C, Volkman SK, Daily J, et al. A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog*. 2006;2:e57.
7. Vaidya AB, Mather MW. Atovaquone resistance in malaria parasites. *Drug Resist Updat 3*. 2000;283–7.
8. Forsdyke DR. Reciprocal relationship between stem-loop potential and substitution density in retroviral quasispecies under positive Darwinian selection. *J Mol Evol*. 1995a;41:1022–37.
9. Barrette IH, McKenna S, Taylor DR, Forsdyke DR. Introns resolve the conflict between base order-dependent stem-loop potential and the encoding of RNA or protein: further evidence from overlapping genes. *Gene*. 2001; 270:181–9.
10. Forsdyke DR. A stem-loop "kissing" model for the initiation of recombination and the origin of introns. *Mol Biol Evol*. 1995b;12:949–58.
11. Forsdyke DR. Stem-loop potential in MHC genes: a new way of evaluating positive Darwinian selection? *Immunogenetics*. 1996;43:182–9.
12. Zhang CY, Wei JF, Wu JS, Xu WR, Sun X, He SH. Evaluation of FORS-D analysis: a comparison with the statistically significant stem-loop potential. *Biochem Genet*. 2008;46:29–40.
13. Aurrecoechea C, Brestelli J, Brunk BP, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*. 2009;37:D539–43.
14. Tinoco I Jr, Uhlenbeck OC, Levine MD. Estimation of secondary structure in ribonucleic acids. *Nature*. 1971;230:362–7.
15. Forsdyke DR. An alternative way of thinking about stem-loops in DNA. A case study of the human G0S2 gene. *J Theor Biol*. 1998;192: 489–504.
16. Forsdyke DR. Conservation of stem-loop potential in introns of snake venom phospholipase A2 genes: an application of FORS-D analysis Molecular Biology and Evolution. 1995c;12:1157–65.
17. Yuthavong Y. Basis for antifolate action and resistance in malaria. *Microbes Infect*. 2002;4:175–82.
18. Yuthavong Y, Kamchonwongpaisan S, Leartsakulpanich U, Chitnumsub P. Folate metabolism as a source of molecular targets for antimalarials. *Future Microbiol*. 2006;1:113–25.
19. Bates SJ, Winstanley PA, Watkins WM, et al. Rare, highly pyrimethamine-resistant alleles of the *Plasmodium falciparum* dihydrofolate reductase gene from 5 African sites. *J Infect Dis*. 2004;190:1783–92.

20. Zindrou S, Nguyen PD, Nguyen DS, Skold O, Swedberg G. *Plasmodium falciparum*: mutation pattern in the dihydrofolate reductase-thymidylate synthase genes of Vietnamese isolates, a novel mutation, and coexistence of two clones in a Thai patient. *Exp Parasitol*. 1996;84:56–64.
21. Garg S, Saxena V, Kanchan S, et al. Novel point mutations in sulfadoxine resistance genes of *Plasmodium falciparum* from India. *Acta Trop*. 2009; 110:75–9.
22. Chookajorn T, Ponsuwanna P, Cui L. Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation. *Trends Parasitol*. 2008;24:455–61.
23. Deitsch K, Duraisingh M, Dzikowski R, et al. Mechanisms of gene regulation in *Plasmodium*. *Am J Trop Med Hyg*. 2007;77:201–8.
24. Frank M, Deitsch K. Activation, silencing and mutually exclusive expression within the var gene family of *Plasmodium falciparum*. *Int J Parasitol*. 2006;36:975–85.
25. Kyriacou HM, Stone GN, Challis RJ, et al. Differential var gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol*. 2006;150:211–8.
26. 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.
27. Pasternak ND, Dzikowski R. PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*. *Int J Biochem Cell Biol*. 2009;41:1463–6.
28. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol*. 2008;62:445–70.
29. Volpato JP, Pelletier JN. Mutational 'hot-spots' in mammalian, bacterial and protozoal dihydrofolate reductases associated with antifolate resistance: sequence and structural comparison. *Drug Resist Updat*. 2009;12:28–41.
30. Tahar R, Basco LK. Molecular epidemiology of malaria in Cameroon. XXV. In vitro activity of fosmidomycin and its derivatives against fresh clinical isolates of *Plasmodium falciparum* and sequence analysis of 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *Am J Trop Med Hyg*. 2007;77:214–20.
31. Dharia NV, Sidhu AB, Cassera MB, et al. Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol*. 2009;10:R21.
32. Cowman AF, Galatis D, Thompson JK. Selection for mefloquine resistance in *Plasmodium falciparum* is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. *Proc Natl Acad Sci U S A*. 1994;91:1143–7.
33. Price RN, Uhlemann AC, Brockman A, et al. Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet*. 2004;364:438–47.
34. Reed MB, Saliba KJ, Caruana SR, Kirk K, Cowman AF. *Pgh1* modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature*. 2000;403:906–9.
35. Mu J, Awadalla P, Duan J, et al. Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat Genet*. 2007;39:126–30.