Genomic analysis of gene structure and expression of Plasmodium falciparum rifins

Pizzi E, Bultrini E, Silvestrini F, Alano P

Dipartimento di Malattie Infettive, Parassitarie e Immunomediate,
Istituto Superiore di Sanità, Roma

Motivation

Rifins are members of the most abundant multigene family of Plasmodium falciparum genome. They are two-exon genes coding for transmembrane proteins that are probably located at the surface of infected erythrocytes where they have been supposed to contribute to antigenic variability of the parasite. Because of their subtelomeric location they are subjected to frequent recombination events leading to new repertoires of genes coding for proteins with novel antigenic properties. Although the completion of genome sequencing an extensive analysis on rifin organization and distribution has not been carried out to date, despite their probable important role in host-parasite interactions. In this work we present a sequence analysis of the entire repertoire of rifin genes in P. falciparum genome. We investigated 5' upstream and 3' downstream sequences for the presence of potential regulatory elements and for conformational and compositional properties. Then we analysed available gene expression data.

Methods

We extracted 5' upstream (500 bp), coding and 3' downstream (500 bp) sequences for each of the 150 rifin genes in the P. falciparum genome. Within each group of sequences a comparative analysis was carried out: each sequence was compared with all the others and a distance (D=100-percentage of identity) was calculated for each comparison. These data were used as input for a multidimensional scaling. This procedure corresponds to map sequences on a two-dimensional space in which relative distances are maintained and hence allowed us to easily recognize subfamilies of them. According to these results we proposed a classification for rifin genes and on the basis of a simple probabilistic model we studied their distribution in the genome. Mean profiles for bendability propensity and nucleotide composition were constructed for 5' upstream and 3' downstream sequences. Further, we looked for common oligomers 8 bp long that are overrepresented with respect a given background (lexicon-partitioning). We considered as significant only oligomers that occur at least in the 80% of sequences with an obs/exp ratio higher than 1.5.We analysed available gene expression data to identified rifin transcripts regulated during the asexual life cycle of the parasite.

Results

We applied several methods to characterize sequences corresponding to 5' upstream, coding and 3' downstream regions, to identify putative regulatory elements and to analyse available gene expression data. We found that 5' upstream sequences, as well as coding ones, can be grouped into two sub-families, whereas 3' downstream sequences are organized into three main clusters. On the basis of these results, we classified rifin genes as combinations of different 5', coding and 3' sequences. We found that only some arrangements of the three modules occur in the genome, some are completely absent and some others occur at the same frequency as the random expectation, suggesting positive selection acting on these genes. We investigated 5' upstream and 3' downstream sequences to look for potential regulatory elements and we found that in the case of 5' upstream sequences potential transcription start sites are probably located about at -200 bp from the first codon. Finally we studied available gene expression data. We found that only 7 out of 150 possible rifin transcripts are regulated during the asexual life cycle of the parasite, supporting the hypothesis that these genes are subjected to antigenic variation mechanisms.

Contact email: epizzi@iss.it