J Mol Evol. 2005 Oct;61(4):513-23. doi: 10.1007/s00239-005-0023-5. Epub 2005 Jul 21.

Proteome composition in Plasmodium falciparum: higher usage of GC-rich nonsynonymous codons in highly expressed genes

Ipsita Chanda [1], Archana Pan, Chitra Dutta

Affiliations + expand

PMID: 16044241 DOI: 10.1007/s00239-005-0023-5

Abstract

The parasite Plasmodium falciparum, responsible for the most deadly form of human malaria, is one of the extremely AT-rich genomes sequenced so far and known to possess many atypical characteristics. Using multivariate statistical approaches, the present study analyzes the amino acid usage pattern in 5038 annotated protein-coding sequences in P. falciparum clone 3D7. The amino acid composition of individual proteins, though dominated by the directional mutational pressure, exhibits wide variation across the proteome. The Asn content, expression level, mean molecular weight, hydropathy, and aromaticity are found to be the major sources of variation in amino acid usage. At all stages of development, frequencies of residues encoded by GC-rich codons such as Gly, Ala, Arg, and Pro increase significantly in the products of the highly expressed genes. Investigation of nucleotide substitution patterns in P. falciparum and other Plasmodium species reveals that the nonsynonymous sites of highly expressed genes are more conserved than those of the lowly expressed ones, though for synonymous sites, the reverse is true. The highly expressed genes are, therefore, expected to be closer to their putative ancestral state in amino acid composition, and a plausible reason for their sequences being GC-rich at nonsynonymous codon positions could be that their ancestral state was less AT-biased. Negative correlation of the expression level of proteins with respective molecular weights supports the notion that P. falciparum, in spite of its intracellular parasitic lifestyle, follows the principle of cost minimization.

Similar articles