FEBS Letters 581 (2007) 5751–5758

Comparative codon and amino acid composition analysis of Tritryps-conspicuous features of *Leishmania major*

Ipsita Chanda[a], Archana Pan[b], Sanjoy Kumar Saha[a], Chitra Dutta[a,*]

[a] Department of Structural Biology and Bioinformatics, Indian Institute of Chemical Biology, Kolkata 700 032, India
[b] Computational Biology Group, Theoretical Physics Department, Indian Association for the Cultivation of Science, Kolkata 700 032, India

Received 12 July 2007; revised 11 October 2007; accepted 9 November 2007

Available online 26 November 2007

Edited by Takashi Gojobori

Abstract **Comparative analyses of codon/amino acid usage in *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi* reveal that gene expressivity and GC-bias play key roles in shaping the gene composition of all three parasites, and protein composition of *L. major* only. In *T. brucei* and *T. cruzi*, the major contributors to the variation in protein composition are hydropathy and/or aromaticity. Principle of Cost Minimization is followed by *T. brucei*, disregarded by *T. cruzi* and opposed by *L. major*. Slowly evolving highly expressed gene-products of *L. major* bear signatures of relatively AT-rich ancestor, while faster evolution under GC-bias has characterized the lowly expressed genes of the species by higher GC_{12}-content.**

Keywords: Gene expressivity; GC-bias; Hydropathy; Aromaticity; Correspondence analysis; Principle of cost minimization

1. Introduction

The trypanosomatid pathogens *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*, often referred together as "Tritryps" [1], are three closely related kinetoplastid parasitic protozoa that cause some of the most debilitating diseases of humankind – cutaneous leishmaniasis, African sleeping sickness and Chagas disease, respectively [2]. All three parasites possess complex life-cycles alternating between the specific insect vectors and the mammalian hosts, undergoing distinct developmental changes in the insect vectors [3–5] that allow them to infect the human host. In spite of considerable research efforts, no vaccine could be approved yet for any of tance [6]. There is, therefore, an urgent need to understand the biology of these pathogens and people are trying to exploit their genome information [3–5] in this regard. *L. major*, *T. brucei* and *T. cruzi* contain about 32.8, 26 and 55-megabase size haploid genomes distributed in 36, 11 and 28 chromosomes with an average GC-content of 59.7%, 46.4% and 51%, respectively. Comparative analyses [1] revealed that the three genomes share 6158 ortholog clusters of protein-coding genes, which exist in large syntenic blocks containing 80% of the *T. brucei* and 93% of the *L. major* genes. They also share a number of molecular and biochemical characters [7]. Yet the Tritryps differ in features like mode of transmission by different insects, different target tissues, distinct disease pathogenesis and use of different strategies of immune evasion [1]. In *L. major* genes, a negative correlation exists between GC_{12} and GC_3, the origin of which has remained an open question [8]. For *T. brucei* and *T. cruzi*, however, this correlation is positive. In an effort to analyze the compositional similarities and divergence within and across these genomes in further details, we report a comparative multivariate analysis of their codon and amino acid usage patterns.

2. Materials and methods

2.1. Genome sequence data

The nuclear genome sequence of *L. major* with 8272 protein-coding genes was extracted from Sanger database (http://www.sanger.ac.uk/) and those of *T. cruzi* and *T. brucei* with 12570 and 9068 from TIGR Database (http://www.tigr.org). Annotations of the open reading frames (ORFs) were cross-checked with GeneDB. To reduce the sampling error, the genes with less than 100 codons, internal stop codons, untranslated codons and pseudogenes were excluded from the analysis, resulting in the datasets of 7806, 6084 and 11627 predicted ORFs for *L. major*, *T. brucei* and *T. cruzi*, respectively.