

# Spatial Architecture of Nucleotides in Eukaryotic Genomes

**Charu Kapil<sup>1,2</sup>,  
Samrat Chatterjee<sup>1</sup>  
Tahir Husain<sup>1</sup>,  
Mohamad  
Aman Jairajpuri<sup>2</sup>,  
Manickam Yogavel<sup>1</sup>  
and Amit Sharma<sup>1</sup>**

<sup>1</sup> Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology, New Delhi 110 067, India

<sup>2</sup> Department of Biosciences, Jamia Millia Islamia University, Jamia Nagar, New-Delhi 110 025, India

**Address for correspondence:**

E-mail: [charukapil.bioinfo@gmail.com](mailto:charukapil.bioinfo@gmail.com)

Malaria is one of the most widespread infectious diseases and a global health problem. Of the four Plasmodium species causing human malaria, *Plasmodium falciparum* is

the deadliest. The completion of Plasmodium genome sequencing and availability of PlasmoDB database has provided a platform for systematic study of parasite genome.

Till date no effective vaccines for malaria is available, which necessitates a comparative in-depth study of this AT-rich genome with other eukaryotic genomes. Already genomic sequencing of various organisms from pathogenic and non-pathogenic class has provided a wealth of information about coding sequences. The content of the nucleotides (A, T, G and C) varies in different organisms making them AT/GC-rich or neutral. Several hypothesis about codon usage leading to residue biasness such as selection pressure, mutations and genetic drifts have been proposed, however, the underlying reasons for varied compositions of AT and GC nucleotides causing different prevalence of specific residues in certain regions is still unclear. We focused on the pyrimidines and purines which form the basic genomic architecture of an organism from 85 eukaryotic organisms belonging to different genomic compositions including pathogens (such as *P. falciparum*, *C. parvum*, *T.gondii*, *L. major*) and

non-pathogens, at the codon level analysis. On analyzing the spatial distribution of four nucleotides at each position of a codon, we found guanine (G) and adenine (A) to be less varied as compared to thymine (T) for all the positions. Irrespective of the nature of nucleotide at first position, the probability of occurrence of 'G' or 'A' are very similar at the second and the third positions of a codon. G is widely spread at the first position as compared to other nucleotides. A+T and G+C at second and third position have upside down distribution. We presume that particular codons preference in the genome could actually enlarge the evolutionary space to gain new functions or also to adapt to stressful conditions mainly in case of pathogenic organisms. We found no correlation of pathogenicity to the biased base composition. The above analysis, gives a brief overview of the preference of specific nucleotide by organisms belonging to different genomic compositions.