

## Mysterious, taxon-specific enrichment bias of Simple Sequence Repeats

Divya Tej Sowpati, Surabhi Srivastava, Akshay Kumar Avvaru, Rakesh Kumar Mishra  
CSIR - Centre for Cellular and Molecular Biology, Hyderabad, Telangana - 500007

Simple Sequence Repeats (SSRs), also known as microsatellites, are short tandem repeats of 1-6nt DNA motifs, present in all genomes. They have long been used in the fields of linkage analysis, genotyping, and forensics, and recent studies have highlighted the role of microsatellites in gene regulation and genome organization. Abnormal expansion of SSRs are associated with several neurodegenerative diseases in humans. SSRs are distributed non randomly in genomes, and show high mutation rates due to DNA polymerase slippage that can result in the increase or decrease of their length. We have previously developed an algorithm for accurate and exhaustive identification of microsatellites from genomic sequences, using which we collected the microsatellite data from several eukaryotic genomes. While analyzing this data, we identified a surprising enrichment pattern of some SSRs: a bias towards lengths of  $(kn - 1)$ , where  $k$  is the length of the repeating motif, and  $n$  is a positive integer  $\geq 12/k$ . The enrichment is both repeat-specific - only few repeat motifs out of 501 possible motifs show this trend, and clade-specific - the pattern for a given motif is only seen in closely related species. We believe further study of these enrichment biases can shed light on the mechanisms of repeat expansion and selection, with possible implications in speciation and disease.