

**Abstract ID:-** 194

**Abstract Topic:-** Evolutionary and population genetics

**Abstract Title:-** Profiling STR de novo mutations using trios from 1KGP.

**Presenting author name :-** Malini Nemaikanti

**Presenting author institute:-** CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India

**Co-authors name:-** Akshay Kumar Avvaru, Divya Tej Sowpati

**Co-authors institute:-** CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India ;  
Academy of Scientific and Innovative Research, Ghaziabad 201002, India

**Aims:-** Short Tandem Repeats (STRs) are the DNA sequences with motif sizes of 1-6 repeated in tandem, with an average mutation rate of  $10^{-5}$  to  $10^{-3}$  per generation in humans. Profiling the genome-wide distribution of STRs has a functional significance ranging from understanding the regulation of gene expression to roles in genetic diseases. The inherent multiallelic nature of STRs has made it important to profile its genetic variance in different human populations. We study STR variations in 603 trio datasets from the 1000 Genomes Project. The important aspect this study holds is cataloging the STR genetic variation and population-wise STR de novo mutation rates, which has been very less studied in this cohort till now.

**Methods:-** We genotyped about 4.4M STR regions in the genome, the largest dataset so far, using two different tools GangSTR & HipSTR. The trios are genotyped jointly, and the outputs of both tools include allele lengths for each locus as biallelic genotype. Loci are filtered based on parameters including, sample call rate, Hardy-Weinberg equilibrium and also presence in segmental duplications. Post filtering, identification of de novo mutations and their Parent-of-origin (PoO) is done using MonSTR, which considers the genotype likelihoods of the parents and offspring to confidently call de novo alleles. To understand correlation between the rate of the de novo mutations and the effect sizes of the SNPs, a Genome-wide Association Study (GWAS) has been performed using the PLINK tool set.

**Results:-** The initial results suggest about 0.5M loci to be variable across different population groups ascertaining the polymorphism. We identify an average of  $\sim 1050$  de novo mutations per sample. In-house analyses observes an average of 20 % allele dropout cases. Analysis of STR loci with de novo mutations reveals that dimers, tetramers have higher mutation rates while trimers see less mutations. Overall, expansion vs contraction ratios are equal but we observe an interesting equilibrium with shorter STRs exhibiting higher rate of expansions and longer lengths having higher contractions. Association of de novo mutation rates with LD pruned SNPs and indels yielded a set of significant genomic variants which include both known and unknown variants.

**Conclusions:-** Unlike previous studies limiting analyses to a single population group, we catalog de novo STR variants in 603 trios of 1KGP from 18 different population groups. Our pipeline standardization reveals HipSTR and GangSTR, relatively advanced STR genotyping tools, show variability in their results. While GangSTR gives more consistent de novo mutation numbers, it picks up many allele dropout cases. This raises the need for more robust STR genotyping tools.

**Keywords:-** short tandem repeats, de novo mutations, trio datasets, STR genotyping, GWAS