

Abstract

Sequencing of a Sri Lankan personal genome and public availability of personal genomes and their variations enables genomic variation based intra-species comparison that would serve as first step in understanding how much Sri Lankan genome similar to and differ from the rest of the personal genomes. In this study, challenges and confounding factors of comparing genomic variations in personal genomes are considered and a single nucleotide polymorphism based comparison was designed.

Single nucleotide polymorphisms of a selected one hundred thousand base pair region of Sri Lankan genome was compared with the same region of the two of the published genomes; James Watson's personal genome and anonymous Chinese genome. The comparison was done using online public Bioinformatics tools and Databases. Coordinates of genomic variation data of all three genomes have been converted to the current build of the human genome reference sequence prior to comparison to acquiring comparable dataset from selected region of three personal genomes.

Selected region of the Sri Lankan genome was found to have 182 single nucleotide polymorphisms, James Watson 158 and Chinese 115. From 182 SNPs in the selected 100000 base pair region of Sri Lankan genome, 16 of these SNPs were novel variations not mapped to Database of Single Nucleotide Polymorphisms. The 166 Sri Lankan SNPs mapping to Database of Single Nucleotide Polymorphisms were compared with other two sets, 100 was common to Sri Lankan and James Watson. Another 67 were found to be common to Sri Lankan and Chinese. 50 of the SNPs were common to all three compared genomes. The methodology used here had succeeded in acquiring comparable dataset from selected region of three personal genomes assembled on different builds of reference sequence and Single nucleotide polymorphisms annotated with different builds of Database of Single Nucleotide Polymorphisms. Overall, these results agree with recent results of personal genome comparisons. In addition to requiring significantly smaller computational power and storage capacity, this method can be considered as a pilot for the future challenges of personalized genome comparisons.