

## Abstract

The Sri Lankan Human Genome Project was initiated by the University of Colombo with the collaboration of Institute of Genomics and Integrative Biology, New Delhi, India. The first ever complete Human Genome was successfully sequenced and launched on 10th December 2010. The sequence had 2,811,918 Single Nucleotide Polymorphisms and 250,463 novel variations by mapping to the database of Single Nucleotide Polymorphism build 131 of National Center for Biotechnology Information. Subsequently the database of Single Nucleotide Polymorphism was upgraded with the identification of new variations as well as the deletions and merging of previous variations. Therefore the current figures of the Sri Lankan Human Genome variations are not up to date.

Any personal genome upgrade process consists of identification of novel variations, deletion of inactive variation and alteration of existing variations with reference to a upgraded database of Single Nucleotide Polymorphism. Along with the each Single Nucleotide Polymorphism database upgrade process, National Center for Biotechnology Information' produces archives for the deleted and merged variations up to that particular database and the complete list of currently active Single Nucleotide Polymorphism with chromosomal locations and these files are make publicly available at their ftp site.

The objective of this project is to identify the required resources to upgrade a personal genome and efficiently incorporate those resources to design and implement a Single Nucleotide Polymorphism re-annotation pipeline. The re-annotation pipeline was implemented and automated by using free and open source technology including Perl programming language and MySQL relational database management system. All the required data resources were downloaded from the National Center for Biotechnology Information's ftp site.

The Single Nucleotide Polymorphism re-annotation pipeline was deployed to upgrade the Sri Lankan Human Genome to the database of Single Nucleotide Polymorphism build 135 of National Center for Biotechnology Information and found out there were 7865 newly identified Single Nucleotide Polymorphism, 5,732 deleted Single Nucleotide Polymorphism and 237 merged Single Nucleotide Polymorphism from the previous Sri Lankan Human Genome variations.