Sweep Discovery Tool: A machine learning based software for predicting Selective Sweep
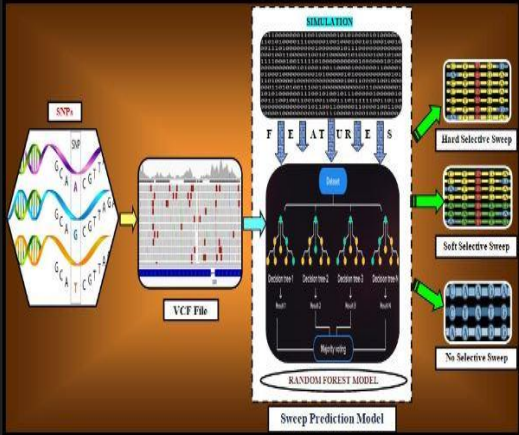
Sweep Discovery Tool is a user friendly tool for detecting selective sweep in genomic regions. Selective sweep is a biological phenomenon in which genetic variation on a selected locus gets swept away due to increase in frequency of beneficial allele along with its surrounding alleles after fixation of beneficial mutation. Sweep Discovery Tool has been made using machine learning based Random Forest model in its server logic. The interface of tool has been developed through R Shiny program. User interface of the Sweep Discovery Tool consists of five tabs namely- 1) Home, 2) Predict, 3) Data Availability, 4) Developers and 5) Contact Details. Sweep Discovery Tool takes VCF file along with tabixed VCF file as input. User has to specify the region of interest for his/her study such as chromosome no, start position and end position. The tool predicts the selective sweep status of that specified region in terms of three classes such as i) No Selective Sweep, ii) Soft Selective Sweep and iii) Hard Selective Sweep. Sweep Discovery Tool is accessible through http://cabgrid.res.in:5599/ and maintained by ICAR-Indian Agricultural Statistics Research Institute, New Delhi, 110012.

- ▪ **Instruction for users**
- ▪ **Input: -**
    - ➢ Input File Format: - User has to provide input in **gzipped VCF format (.vcf.gz)** along with the **tabix-indexed VCF file (.vcf.gz.tbi)**. VCF file must contain sample information properly (Check Example Data under "Data Availability" tab for further information).
    - ➢ Input Parameters: - User has to mention required input parameters such as **"Chromosome No/Identifier", "Start Position" and "End Position"** along with the input files.

- ▪ **Prediction procedure: -**
    - ➢ Open the Sweep Discovery Tool through the link http://cabgrid.res.in:5599/.
    - ➢ Choose the **"Predict Tab"** from the home page.
    - ➢ Provide **Input files** together at first in the required format and mention a **Directory** name. Then press the **"Save"** button to save the files.
    - ➢ Now provide the necessary **input parameters** (Chromosome No/Identifier, Start Position and End Position) and finally click on **"Predict"** button for prediction.
    - ➢ After prediction, the result will be shown in the interface as well as user can **download** the result in **".csv"** format from the **"Download Results"** button.
    - ➢ Steps are being shown in the following figure for the ease of user accessibility.