WeIMPUTE FORMAT Conversion Tool User Manual

1. Overview of the weIMPUTE_FORMAT Conversion Tool

The weIMPUTE_FORMAT Conversion Tool is designed to facilitate the conversion of genomic data from various formats (e.g., PLINK, Hapmap) to the Variant Call Format (VCF), which is the standard input format for weIMPUTE. This tool also supports flexible quality control (QC) measures, such as minor allele frequency (MAF) filtering and Hardy-Weinberg equilibrium (HWE) checks. With this tool, researchers working with different genomic data formats can seamlessly integrate their datasets into the weIMPUTE pipeline for genotype imputation, regardless of their initial format.

The weIMPUTE_FORMAT Conversion Tool is built to support diverse species, from human genetics to non-human species like plants and animals, making it highly adaptable for various research purposes. This includes but is not limited to, species like yaks, fruit flies, soybeans, and maize. Researchers can now process and analyze their datasets with minimal bioinformatics expertise, thanks to this user-friendly tool.

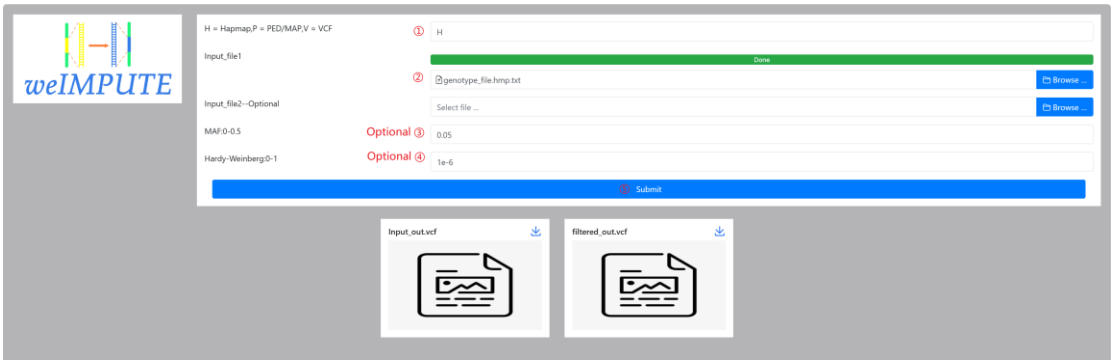
2. System Requirements

Operating System: Linux / macOS / Windows
Java Version: openJDK 17
Software Requirements: R
Dependencies: The relevant directory structure and configuration files should already be set up for proper execution of the tool.

3. How to Use

Example URL: You can access the service via the following address:
http://144.34.160.128:8081/.

3.1 Detailed Steps for using the weIMPUTE_FORMAT Conversion Tool



① Select Input Format:

On the starting interface, first choose your input data format:

H = Hapmap
P = PLINK (PED/MAP)
V = VCF

② **Upload Files:**

Based on your selected input format, upload the corresponding files:

If you choose P (PLINK format):

Upload the PED file in Input_file1.

Upload the MAP file in Input_file2.

If you choose H (Hapmap format):

Upload the necessary Hapmap file (e.g., .ped, .map, or .hapmap.txt) in Input_file1.

If you choose V (VCF format):

Simply upload the VCF file in Input_file1.

③ **Optional Step: Perform MAF Filtering:**

If you need to perform MAF filtering, enter the minimum allele frequency threshold (e.g., 0.01, 0.05) in the MAF input box.

This will exclude variants with a minor allele frequency below the specified threshold.

④ **Optional Step: Perform Hardy-Weinberg Equilibrium (HWE) Test:**

If you want to perform a Hardy-Weinberg equilibrium (HWE) test, enter the p-value threshold (e.g., 1e-6) in the Hardy-Weinberg input box.

This will exclude variants that do not conform to Hardy-Weinberg equilibrium.

⑤ **Submit:**

If MAF filtering and HWE test are not selected, simply click Submit, and the system will only perform the format conversion.

If MAF filtering or HWE test is selected, clicking Submit will trigger both the format conversion and quality control (MAF and HWE checks) steps.

These steps allow you to perform data format conversion with optional quality control measures, based on your specific requirements.

3.2 Starting the weIMPUTE_FORMAT Conversion Service

Prepare Input Files:

Upload the source genomic files (PLINK, Hapmap, or VCF) to the input directory.

If you are working with PLINK (PED/MAP) or Hapmap formats, make sure the respective files are correctly paired and named.

Build and Run the weIMPUTE_FORMAT Tool:

Use Maven to build the tool:

mvn clean install

Once the build is complete, run the tool with the following command:

java -jar weIMPUTE-format-V0.1.jar

The system will initiate the conversion process and be ready to accept user inputs through the command-line interface or a graphical web interface, depending on your preference. You can access the web interface at http://IP:8081/

3.3 Converting Genomic Data Formats

Input File Selection:

PLINK (PED/MAP): If you are using PLINK format files, upload the PED, MAP, and FAM files to the input directory.
Hapmap: For Hapmap format, upload the necessary Hapmap files (e.g., .ped, .map, .hapmap.txt).
VCF: If you already have VCF files, you can directly upload them.
Select Output Format:

Choose VCF as the output format. The conversion tool will generate the corresponding VCF files, which are compatible with weIMPUTE.
Initiate Conversion:

Click on the Convert button to start the conversion process. The system will process the input files and produce a corresponding VCF file.

3.4 Quality Control and Filtering

The tool also supports additional QC measures before or after imputation:

MAF Filtering:

Before or after conversion, you can filter out variants with a minor allele frequency (MAF) below a specified threshold.
Specify the MAF threshold (e.g., 0.01, 0.05) to exclude variants that do not meet the threshold.

Hardy-Weinberg Equilibrium (HWE) Filtering:

You can also perform HWE tests to exclude variants that do not conform to Hardy-Weinberg equilibrium.
Enter the HWE p-value threshold (e.g., 1e-6) to filter out variants that fail the test.
Start Filtering:

Once the necessary QC parameters are set, click on the Apply Filters button to run the MAF and HWE checks.

3.5 View and Download Results

Result Generation:

After the conversion and QC processes are complete, the tool will generate a VCF file containing the filtered genomic data.
Download the Output:

Users can download the converted VCF file, which can be directly used for downstream processing in weIMPUTE.
If additional QC steps were applied, the final output will be available for download as a .vcf file with the applied filters.

4. Summary

The weIMPUTE_FORMAT Conversion Tool provides a user-friendly solution for converting genomic data formats (PLINK, Hapmap, VCF) to the standard VCF format for use in the weIMPUTE pipeline. Additionally, the tool includes flexible quality control features, such as MAF and Hardy-Weinberg equilibrium filtering, making it a valuable resource for researchers working with various genomic data formats. Whether you are working with human genetics or non-human species like animals or plants, the weIMPUTE_FORMAT tool helps streamline your data conversion process, ensuring compatibility with weIMPUTE and consistency in your research workflow.

For more information or to access the service, visit http://144.34.160.128:8081/.