# User Manual for

# TRCMGene:

## A two-step referential compression method for the efficient storage of genetic data

**Last updated on June 01, 2018**

# Preparation

1 Run the self-installing executable file to unpack and install JDK. As part of JDK, this installation may include the Java Runtime Environment. (http://www.oracle.com/technetwork/java/javase/downloads/index.html).

2 "**trcm.jar**","**Demo data**" and "**User Manual**" can be downloaded from our website (https://github.com/tangyou79/TRCM). The .jar and Demo data must be saved in the same directory (Fig.1).

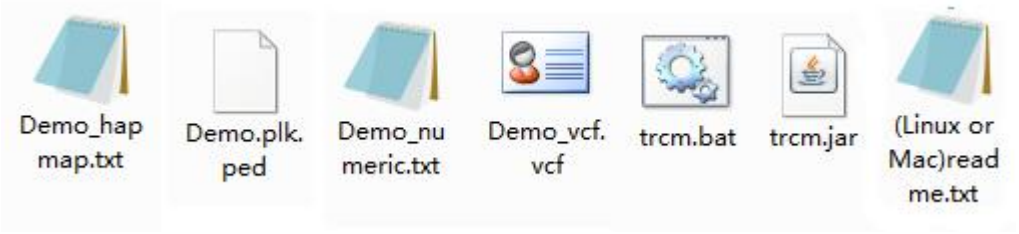


Fig.1 Working folder

3 Operate **trcm** according to the **user's manual**.

# I Compression of genotype files by ORCM

In this section, the genotype files should be compressed using **trcm.jar** in Windows or Linux.

**1 Windows**

**Step 1: Double-click the .bat file; you will then see three options (Fig.2).**

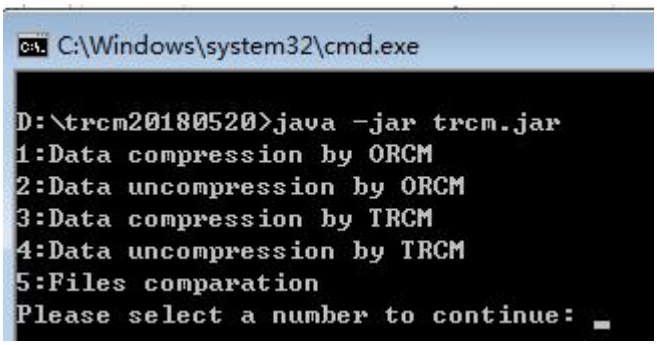


Fig.2

**Step 2: Choose "option 1" to compress the data (Fig.3a). Type the name of the file to be compressed (Fig.3b) followed by the name of the generated file (Fig.3c).**

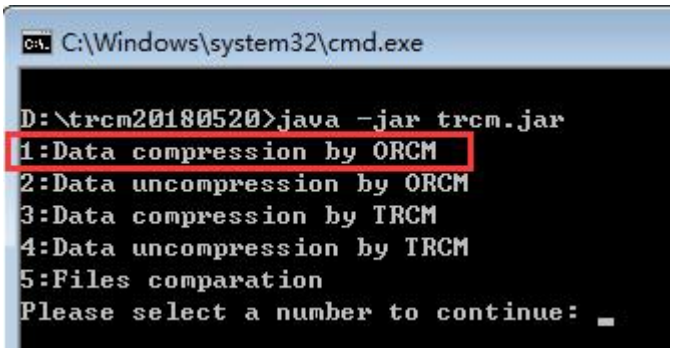


Fig.3a

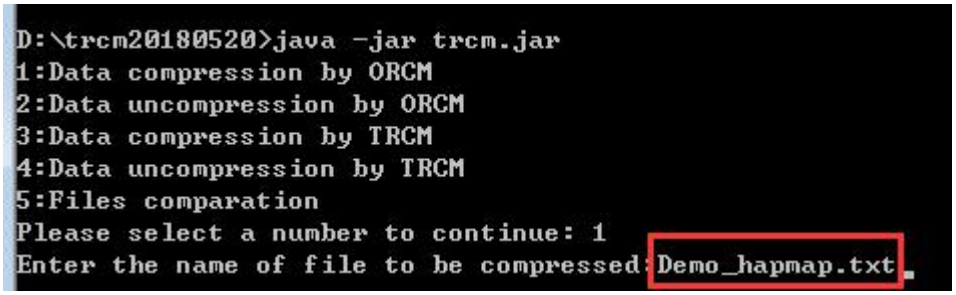


Fig.3b

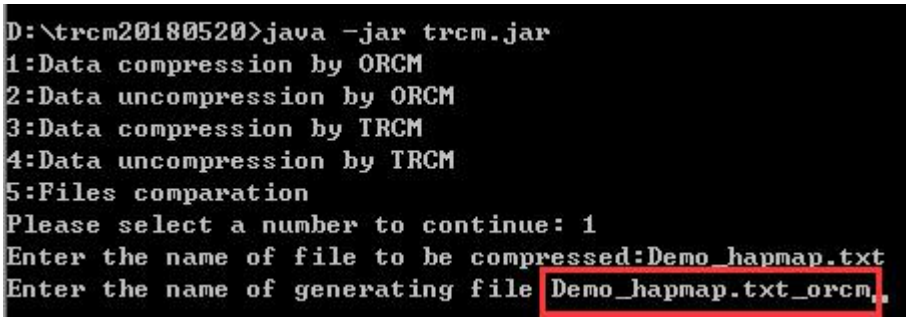


Fig.3c

**Step 3: Select the compression mode (Fig.4). "Option 1" represents data compression through special methods according to the genotype file format (Fig.5a), whereas "Option 2" corresponds to data compression by grouping according to the number you input. The number of groups should vary between 1 and 20 (Fig.5b).**

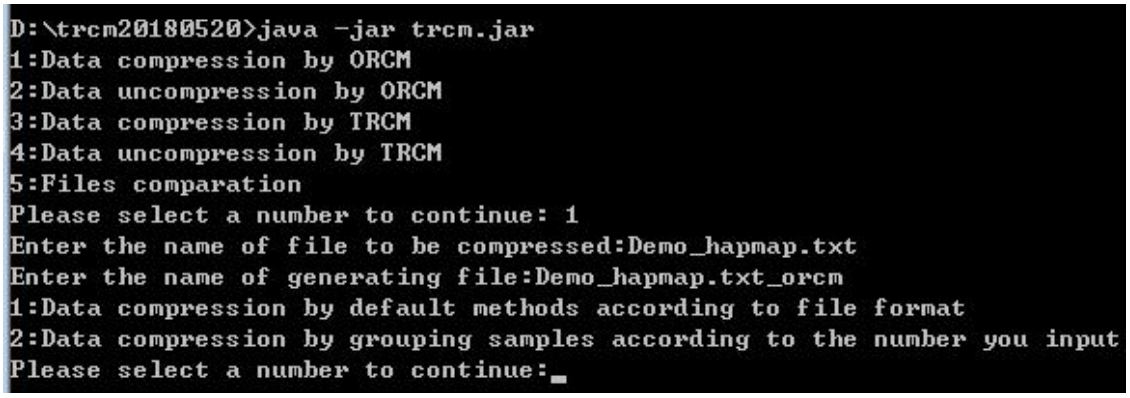


Fig.4

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 1
Enter the name of file to be compressed:Demo_hapmap.txt
Enter the name of generating file:Demo_hapmap.txt_orcm
1:Data compression by default methods according to file format
2:Data compression by grouping samples according to the number you input
Please select a number to continue:1
Please choose the file format:
1.HapMap
2.VCF
3.PLink
4.Numeric
Please select a number to continue:1
Data is being processed,please wait...
Demo_hapmap.txt_orcm is created
total time:0s
```

Fig.5a

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 1
Enter the name of file to be compressed:Demo_hapmap.txt
Enter the name of generating file:Demo_hapmap.txt_orcm2
1:Data compression by default methods according to file format
2:Data compression by grouping samples according to the number you input
Please select a number to continue:2
Enter the sample Numbers:5
Data is being processed,please wait...
Demo_hapmap.txt_orcm2 is created
total time:0s
```

Fig.5b

Note 1: The compression file will be saved in the same directory as the original file (Fig.6).

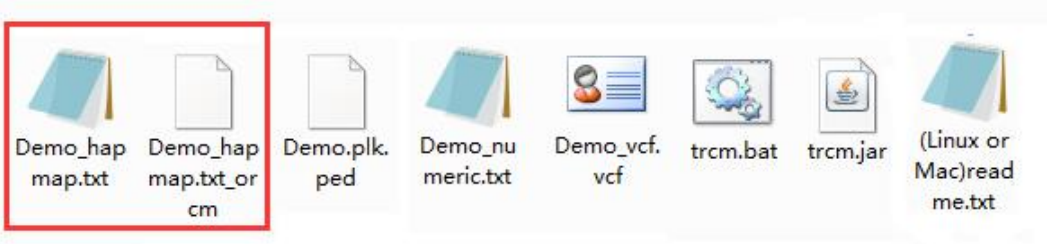


Fig.6

Note 2: The compression methods in options 1 and 2 are both reference-based.

You can choose “option 2” after you execute **.jar** command line to extract your compressed file (Fig.7). The extracted file is also located in the working directory (Fig.8); Choose “option 5” to

compare the original and decompression files (Fig.9).

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 2
Enter the name of file to be uncompressed:Demo_hapmap.txt_orcm
Please enter the name of generating uncompression file:Demo_hapmap.txt_r
Data is being processed,please wait...
Demo_hapmap.txt_r is created
total time:0s
```

Fig.7


Fig.8

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 5
please input filename1:Demo_hapmap.txt
please input filename2:Demo_hapmap.txt_r
Two files are the same
```

Fig.9

**2 Linux**

The compression operation in Linux is similar to that in Windows, but the pattern of their save directory varies from each other (Fig.10).

```
[tangyou@bogon 桌面]$ java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 1
```

```
Demo_hapmap.txt          2796282
Demo_hapmap.txt_orcm      836256
Demo_hapmap.txt_trcm      316297
Demo_hapmap.txt_trcm_r   2796282
```

Fig.10

# II Compression of genotype files by TRCM

In this section, the genotype files should be compressed using **trcm.jar** in Windows or Linux.

**1 Windows**

**Step 1: Double-click the .bat file; you will then see three options (Fig.11).**

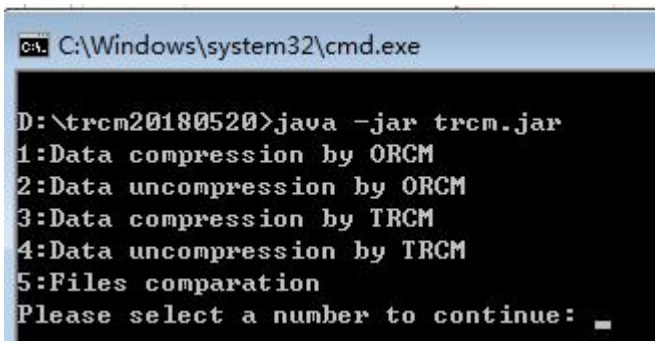


Fig.11

**Step 2: Choose "option 3" to compress the data (Fig.12a). Type the name of the file to be compressed (Fig.12b) followed by the name of the generated file (Fig.12c).**

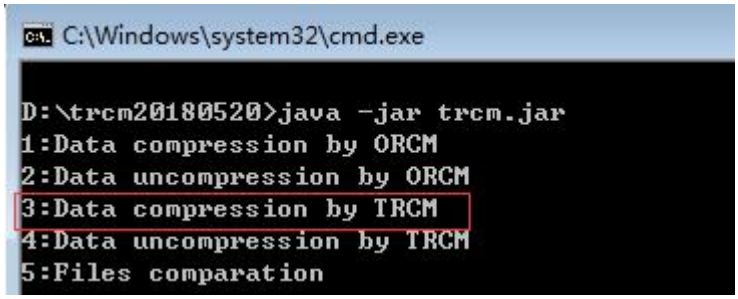


Fig.12a

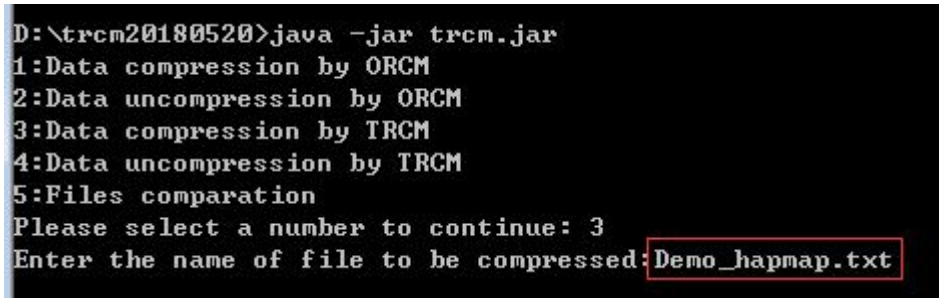


Fig.12b

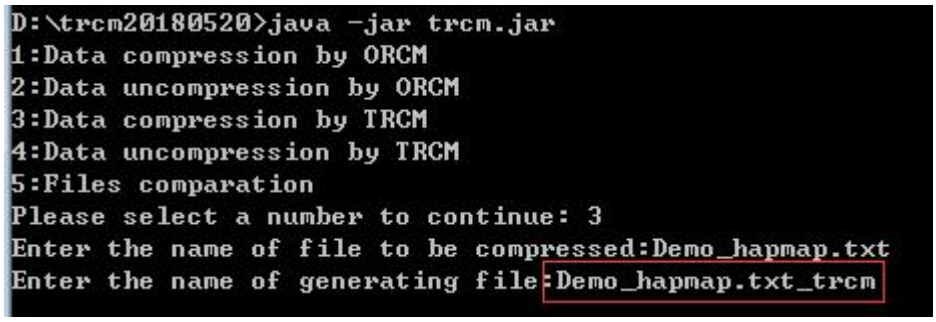


Fig.12c

**Step 3: Select the compression mode (Fig.13). "Option 1" represents data compression through special methods according to the genotype file format (Fig.14a), whereas "Option 2" corresponds to data compression by grouping according to the number you input. The number of groups should vary between 1 and 20 (Fig.14b).**

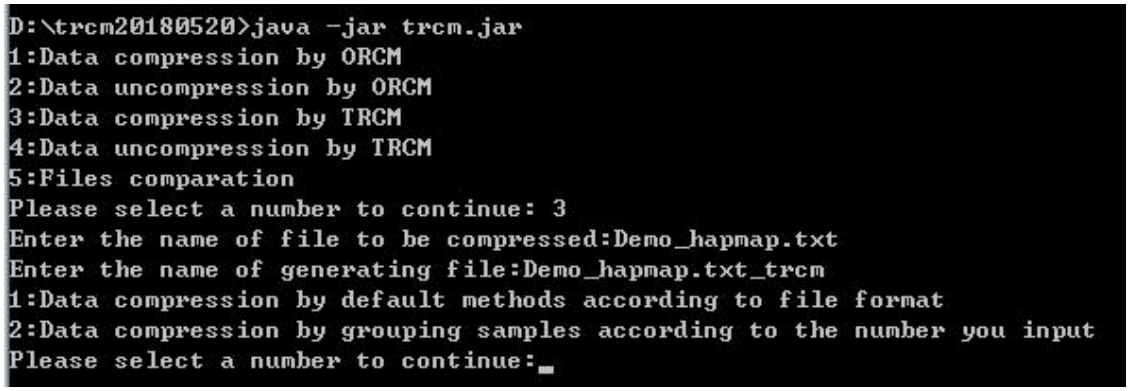


Fig.13

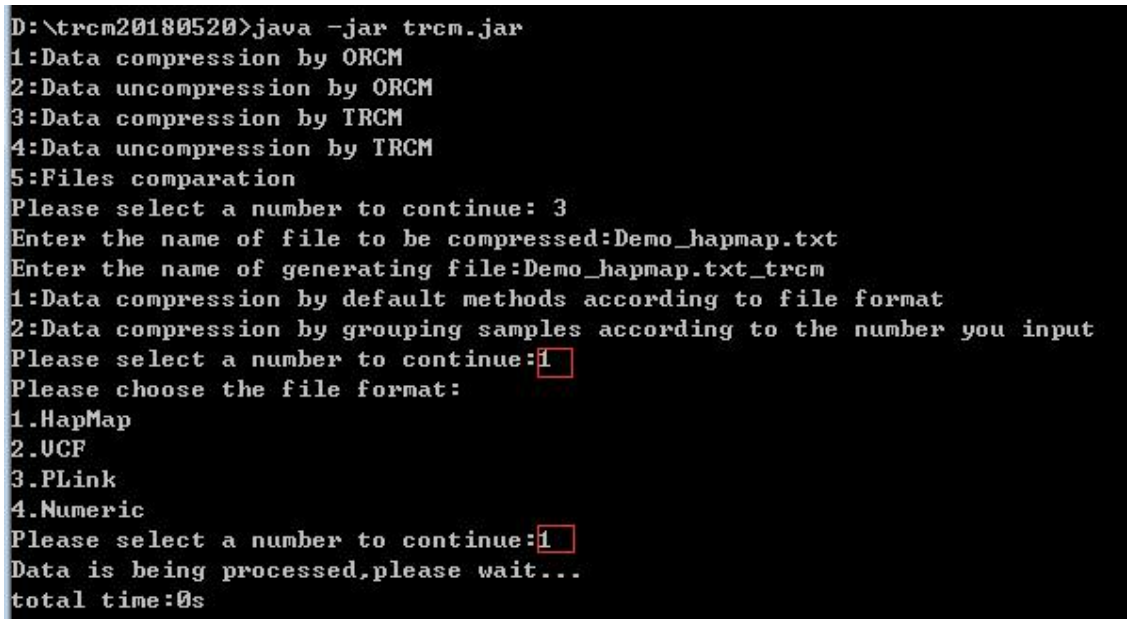


Fig.14a

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 3
Enter the name of file to be compressed:Demo_hapmap.txt
Enter the name of generating file:Demo_hapmap.txt_trcm2
1:Data compression by default methods according to file format
2:Data compression by grouping samples according to the number you input
Please select a number to continue:2
Enter the sample Numbers:5
Data is being processed,please wait...
total time:0s
```

Fig.14b

Note 1: The compression file will be saved in the same directory as the original file (Fig.15).

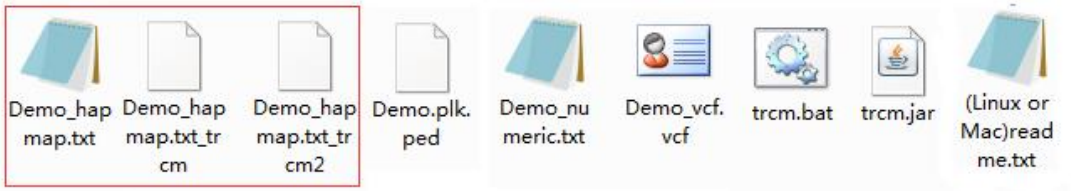


Fig.15

Note 2: The compression methods in options 1 and 2 are both reference-based.

You can choose “option 4” after you execute **.jar** command line to extract your compressed file (Fig.16). The extracted file is also located in the working directory (Fig.17); Choose “option 5” to compare the original and decompression files (Fig.18).

```
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 4
Enter the name of file to be uncompressed:Demo_hapmap.txt_trcm
Please enter the name of generating uncompression file:Demo_hapmap.txt_trcm_r
Data is being processed,please wait...
Demo_hapmap.txt_trcm_r is created
total time:0s
```

Fig.16

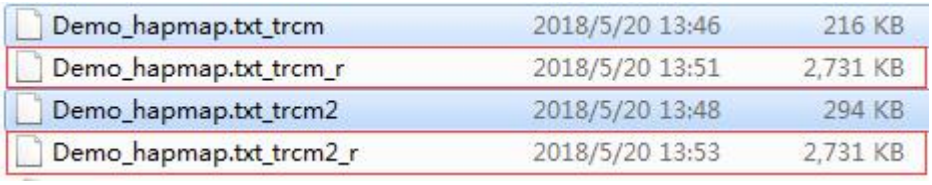


Fig.17

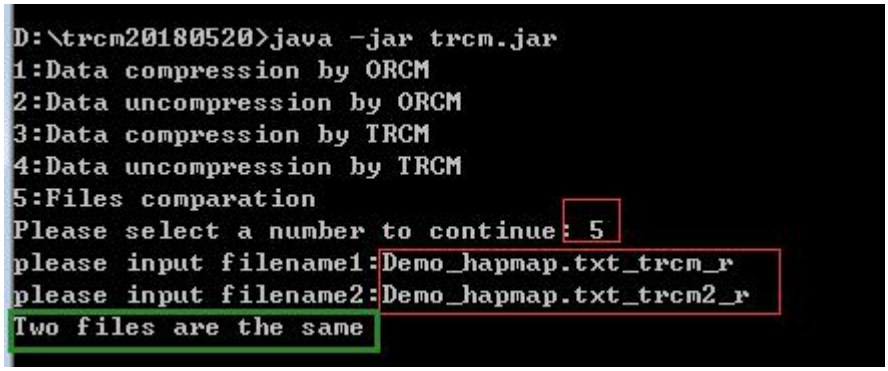
D:\trcm20180520>java -jar trcm.jar
1:Data compression by ORCM
2:Data uncompression by ORCM
3:Data compression by TRCM
4:Data uncompression by TRCM
5:Files comparation
Please select a number to continue: 5
please input filename1:Demo_hapmap.txt_trcm_r
please input filename2:Demo_hapmap.txt_trcm2_r
Two files are the same

Fig.18