User Manual for Running KMC2D to Calculate Epistatic Effect Kinship Matrix

1. KMC2D Overview

The first step of GWAS analysis is computing a kinship matrix, which describes the relationships among individuals. The conventional LMM and the 1D GWAS analysis only can explain a very limited amount of the phenotypic variance, because its statistical model do not include the genes interaction (GXG) and gene by environment interaction(GXE). To explain more the phenotypic variance and address the G×G and G×E interaction effects, we proposed several novel LMMs and developed several related tools: PEPIS (https://bioinfo.noble.org/PolyGenic_QTL/) for polygenic epistasis related 2D GWAS analysis and PATOWAS(https://bioinfo.noble.org/PATOWAS/) for multiple omics marker' interaction analysis. In epistasis analyses, marker pairs are involved for kinship matrix calculation. For $\boldsymbol{m}$ markers, there are $\boldsymbol{c_m^2 = m(m-1)/2}$ marker pairs and calculation of an epistatic kinship matrix has a complexity of $\boldsymbol{O(m^2n^2)}$. Therefore, the combinational marker pairs for a moderate amount of markers will be very huge, which make it impossible to calculate such kind of interaction effect kinship matrices. Recently, the GPU-empowered parallel system has become a standard for HPC platform. Kinship matrix calculation involves pure matrix operations and thus is suitable for GPU-employed parallel structure.

We thoroughly investigated the principles of marker pair assisted kinship matrices and proposed a method to divide the high-dimensional marker pairs into successive blocks. We then calculate the kinship matrix for each block and merge together the block-wise kinship matrices to form the whole kinship matrix. The matrix operations thus can be parallelized by GPU kernels. Based on these, we developed a GPU-empowered pipeline: KMC2D for interaction effect kinship matrix calculation.

The user friendly web version can be accessed at https://bioinfo.noble.org/KMC2D/. However, the user also can download the CUDA source code, then compile and run it in Command Line version. KMC2D can be configured with the different option parameters to run at the mode of either GPU parallel, CPU Serial or Read Loop Through. The following sections will describe the three methods to run KMC2D to calculate your epistatic effect kinship matrix.

2. Running KMC2D to Calculate Epistatic Effect Kinship Matrix

Essentially, KMC2D only can run on a CUDA-capable GPU Linux Server. You can run it at our GPU sever, or download the source code, compile and run it at your CUDA-capable GPU server. Once you download the KMC2D repository, please unzip the two zipped files: Run_Test.zip and Simulated_Data.zip.

2.1 Application of KMC2D using our user-friendly web portfolio.

If you do not have CUDA-capable GPU Linux, no problem, you still can use KMC2D through our developed web portfolio at https://bioinfo.noble.org/KMC2D/ .

2.2Application of KMC2D using bash shell script

If you know well the Linux commands and bash script, and also have a CUDA-capable computation environment, but are not familiar with CUDA-NVIDIA Environment and want to test your hardware and software environment, via running KMC2D in your local Linux server. You can get the KMC2D repository from https://bioinfo.noble.org/KMC2D/ or GitHub at https://github.com/noble-research-institute/KMC2D. At the KMC2D repository, unzip the **Run_Test.zip** and **Simulated_Data.zip** at first, and then go to sub-directory **Run_Test**, and use the script run_test.sh and the following instruction to calculate the epistatic kinship matrix.

```
$ chmod a+x ./run_test.sh

$ chmod a+x ./KMC2D

./run_test.sh ../Simulated_Data/Z_16190_1390.CSV ../Simulated_Data/W_16190_1390.CSV
../Simulated_Data/Result/Kinship_ad.csv
```

2.3 Application of KMC2D using Command Line

If you know well the Linux command, and also are very familiar with CUDA GPU programming, you want to learn the KMC2D's source code, you can refer to INSTALL.md to set your CUDA-capable environment at first and then go to the sub-directory **Source_Code**, and use make command to compile the KMC2D.cu into an executable file KMC2D.

2.3.1 Display the help usage message

```
$ ./KMC2D –h
```

```
Welcome to use this program to do Kinship Matrix Calculation
The usage of input parameter arguments are listed as followings:
-h or -H: Output this Help usage message
-z or -Z: The full name of the 1st Genotype file
-w or -W: The full name of the 2nd Genotype file
-k or -K: The full name of Kinship Matrix file
-i or -I: The Individual number
-l or -L: The direct marker Length
-b or -B: The Block size for dividing the genotype marker pairs
-m or -M: Run_Mode(0, 1, 2 for the GPU/CPU/Read Through)
```

2.3.2 Run KMC2D at the mode of GPU_Parallel (-m 0)

```
$ ./KMC2D -z ../Simulated_Data/Z_16190_1390.CSV -w ../Simulated_Data/W_16190_1390.CSV -k
../Simulated_Data/Result/KM2D_GPU.txt -i 1390 -l 16190 -b 2000 -m 0

Hello, the elapse time is 1984.01 seconds
```

2.3.3 Run KMC2D at the mode of CPU_Serial (-m 1)

$ ./KMC2D -z ../Simulated_Data/Z_16190_1390.CSV -w ../Simulated_Data/W_16190_1390.CSV -k ../Simulated_Data/Result/KM2D_CPU.txt -i 1390 -l 16190 -b 2000 -m 1

Hello, the elapse time is 681755.36 seconds

2.3.4 Run KMC2D at the mode of Read_Loop_Through (-m 2)

$ ./KMC2D -z ../Simulated_Data/Z_16190_1390.CSV -w ../Simulated_Data/W_16190_1390.CSV -k ../Simulated_Data/Result/KM2D_None.txt -i 1390 -l 16190 -b 2000 -m 2

Hello, the elapse time is 204.07 seconds