User Manual for FCF-MixP

Wenwu Xu, Zhiyan Zhang*

State Key Laboratory for Pig Genetic Improvement and Production Technology, Jiangxi Agricultural University, Nanchang, China

Xuwenwu248@outlook.com

bioducklily@hotmail.com

Catalogue

Introduction

FCF-MixP is written by Fortran90 and a stable and big data-oriented method to perform genomic selection. This genomic prediction model with four zero-mean normal distributions as the prior distribution. and the variance of the prior distribution in our model is precisely determined in advance. GEBV can be obtained accurately and quickly in combination with an iterative conditional expectation algorithm. For detailed theory, please refer to our paper FCF-MixP: A stable and big data-oriented genomic selection software based on iterative conditional expectation algorithm.

Note:

Data quality control can be carried out with PLINK software.

Genotype files should not contain missing genotypes, and all individuals at the same site cannot have the same genotype. For example, a column of genotype data cannot all be zero after the genotype file is converted to 0, 1, 2 formats

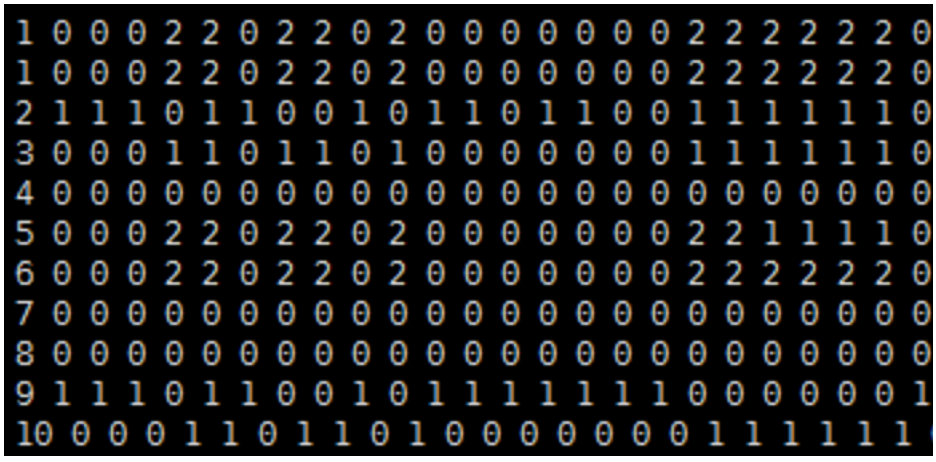
Dependencies

Fortran90

Input files

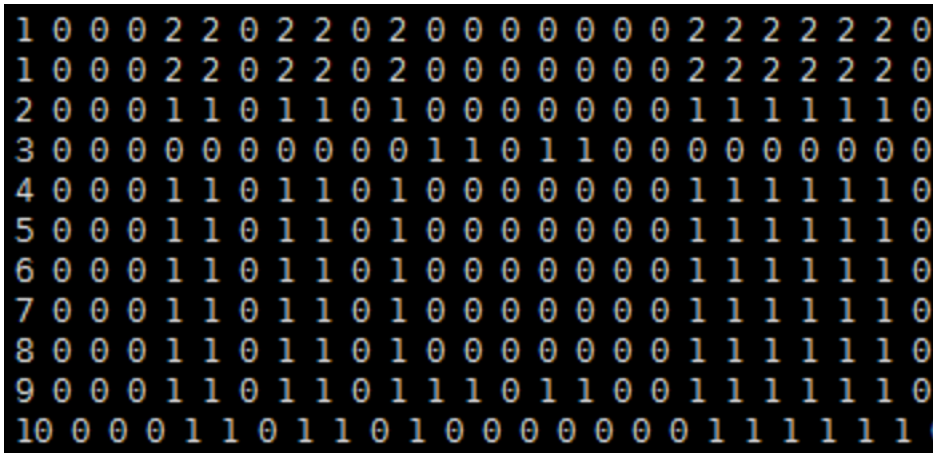
Training genotype file

The training Genotype file are in 0,1,2 formats, the first line of this file consists of the ID of the individual and the name for the markers. Starting with the second line, the first column is an individual number, and on the right is the marker genotype of the corresponding individual. Here is an example:



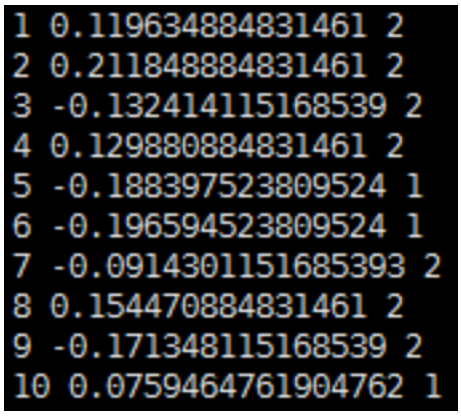
Validation genotype file

This file is in the same format as the training genotype file, and here is an example.



Training Phenotype file

The first column is the individual ID, which must be the same as the ID of the training genotype file. The second and third columns are the phenotypes and the fixed effect coefficients, respectively. Here is an example.



Validation Phenotype file

This file is in the same format as the training phenotype file. Missing phenotypes are replaced by zero, and the prediction accuracy cannot be calculated without phenotypes. Here is an example.

```
1 -0.243725523809499 1
2 -0.010119523809525 1
3 -0.0552015238095239 1
4 -0.0442991151685393 2
5 -0.0832331151685393 2
6 -0.0156101151685393 2
7 0.0847988848314606 2
8 -0.0756925238095239 1
9 0.168159476190476 1
10 -0.173397115168539 2
```

Parameter file

There are 7 parameters in this file, which are the number of individual in the training group, the number of marker in the training group, the number of fixed effects, the proportion of the number of markers in the four categories, the heritability of the trait, and the number of individual in the validation group, the last parameter is 0 or 1, indicating whether the prediction accuracy is calculated.

Run

with the relevant input files for FCF-MixP ready, it is easy for us to run the program with the following code in servers (Linux operating system).

./FCF-MixP

note: we compile the script file target.f90 with the gfortran compiler, which generates 64-bit version of FCF-MixP. If your server has a 32-bit operating system, you can compile a 32-bit FCF with the following code:

gfortran -o FCF-MixP32 target.f90 -m32

(note: Running the executable file FCF-MixP to get the final result)

the following results indicate that our program is running successfully.

```
 The cycle is running:           5   Convergent coefficient=  2.42560860E-02
 The cycle is running:          10   Convergent coefficient=  1.59827527E-03
 The cycle is running:          15   Convergent coefficient=  2.14253625E-04
 The cycle is running:          20   Convergent coefficient=  3.21973857E-05
 The cycle is running:          25   Convergent coefficient=  4.93577909E-06
 The cycle is running:          30   Convergent coefficient=  7.71572843E-07
 The cycle is running:          35   Convergent coefficient=  1.23921538E-07
 The cycle is running:          40   Convergent coefficient=  2.00270289E-08
 The cycle is running:          43   Convergent coefficient=  6.94577373E-09
 The accuracy of MixP in this population is:  0.68035120       1.3446254
the start time is: date = 2020/ 3/31, time = 22:19:58
the end time is: date = 2020/ 3/31, time = 22:20:23
```

output files

ebv.txt

The first and second columns are the individual ID and EBV, respectively. Here is an example.

Individual	EBV
1	-0.1367
2	0.0388
3	-0.0262
4	-0.0284
5	-0.0545
6	0.0008
7	0.0115
8	-0.0279
9	-0.0219
10	-0.0952

Effect of marker.txt

The first and second columns are the marker ID and the effect of each marker, respectively. Here is an example.

No.marker	Effect
1	-0.000000
2	-0.000000
3	-0.000015
4	0.000018
5	0.000017
6	-0.000000
7	0.000018
8	0.000018
9	-0.000015
10	0.000018
11	-0.000025
12	-0.000025
13	-0.000019
14	-0.000018
15	-0.000018
16	-0.000019
17	-0.000025
18	0.000021
19	0.000016
20	0.000024
21	0.000024
22	0.000022
23	0.000024
24	-0.000019

Predictive accuracy.txt

(note: 0.68 and 1.34 represent the prediction accuracy and regression conefficients, respectively)

```
Training size (number of individuals in training set) is:          739
Number of marker loci is:       33901
The accuracy of MixP in this population is:  0.68035120       1.3446254
```

Process of iteration.txt

This file record the total genetic variance and environment variance during each iteration.

Iteration	convergent_coefficient
1	1.00000000
2	0.64897227
3	0.14526892
4	0.05408973
5	0.02425609
6	0.01154948
7	0.00646307
8	0.00406012
9	0.00259149
10	0.00159828