This code is for transforming genotype of plink file to BLUPF90 form

笔记本:	R—code		
创建时间:	2019/8/7 15:11	**更新时间:**	2019/8/7 15:27
作者:	18170061468@163.com		

This code is for transforming genotype of plink file to BLUPF90 form

plink file form is as the following

- ped file：6+2*N columns and individual for each row

```
1       11103832        0       0       0       0       G       G       G       G       C
2       11103863        0       0       0       0       G       G       A       G       C
3       11103458        0       0       0       0       G       G       G       G       C
4       11109536        0       0       0       0       G       G       G       G       T
5       11102017        0       0       0       0       G       G       G       G       T
6       11199000        0       0       0       0       G       G       G       G       C
7       11105007        0       0       0       0       G       G       G       G       C
8       11105687        0       0       0       0       G       G       A       G       T
9       11104882        0       0       0       0       G       G       G       G       C
10      11102935        0       0       0       0       G       G       G       G       C
11      11104849        0       0       0       0       G       G       G       G       T
12      11103090        0       0       0       0       G       G       A       G       C
13      11101680        0       0       0       0       G       G       G       G       T
```

- map file：4 columns for chromosome，SNP name，Mendel distance and position

```
14      ARS-BFGL-BAC-10172      0       6371334
14      ARS-BFGL-BAC-1020       0       7928189
14      ARS-BFGL-BAC-10245      28.23   31819743
14      ARS-BFGL-BAC-10345      0       6133529
14      ARS-BFGL-BAC-10375      0       6616434
14      ARS-BFGL-BAC-10591      0       17544926
14      ARS-BFGL-BAC-10793      24.56396        29259114
14      ARS-BFGL-BAC-10867      36.53289        34639444
14      ARS-BFGL-BAC-10919      26.84388        31267746
10      ARS-BFGL-BAC-10952      43.86   18882288
10      ARS-BFGL-BAC-10960      45.20664        20609250
10      ARS-BFGL-BAC-10972      45.3343 20792754
```

using plink and the following step to transform A T G C into 0 1 2（example for cattle）

```
plink  --cow  --file test  --recodeA  --out test_qc
```

test.raw file will be exist

```
FID IID PAT MAT SEX PHENOTYPE ARS-BFGL-BAC-11718_C ARS-BFGL-BAC-13111_G AR
1 z11103832 0 0 0 -9 0 0 1 0 1 0 0 2 0 1 0 2 0 0 1 0 1 0 1 0 1 0 0 0 0 2 0
2 11103863 0 0 0 -9 0 0 2 0 0 0 0 2 1 0 0 1 1 0 0 0 2 1 1 0 2 1 0 1 0 2 1
3 11103458 0 0 0 -9 2 1 2 2 1 0 0 1 1 0 0 1 1 0 2 1 0 0 1 0 0 1 0 1 0 1 0
4 11109536 0 0 0 -9 0 0 1 0 1 1 0 0 0 0 2 0 0 1 1 0 2 1 0 0 0 0 1 2 0 1 1
5 11102017 0 0 0 -9 0 0 1 0 0 1 0 0 1 0 0 1 0 2 0 0 0 2 0 1 1 0 0 1 0 1 1
6 11199000 0 0 0 -9 0 0 1 0 0 1 0 0 0 0 0 1 0 2 0 1 1 1 0 0 2 0 0 1 0 1 0
```

Now we can use R to transform the file

```
#set path
setwd("D:\\2-test\\20190807SNP\\")
genotype_file_name<-"test_qc.raw"
map_file_name<-"test.map"
out_name<-"genotype"
```

```
plink_blupf90<-function(genotype_file_name,map_file_name,out_name){
  if(!require(data.table)) install.packages("data.table")
  ped<-fread(genotype_file_name,header = F)
  map<-fread("test_qc.map",header = F)
  #genotype for blupf90
  ped<-ped[-1,-c(1,3:6)]
  ped[is.na(ped)]=5
  ped_blupf90<-matrix(nrow = nrow(ped),ncol = 2)
  n<-max(nchar(ped[1,]))
  for(i in 1:nrow(ped)){
    id<-as.character(ped[i,1])
    ped_blupf90[i,1]<-sprintf(paste("%",n,"s",sep = ""),id)
    geno<-as.numeric(ped[i,-1])
    ped_blupf90[i,2]<-paste(geno,collapse = "")
  }
  fwrite(as.data.frame(ped_blupf90),paste(out_name,"_blupf90",sep
=  ""),row.names = F,col.names = F,sep = " ",quote = F)
  #map infomation for blupf90
  map$snp_order<-seq(1:nrow(map))
  fwrite(map[,c(5,1,4)],paste(out_name,"_map_blupf90",sep =  ""),row.names =
F,col.names = F,sep = " ",quote = F)
}
plink_blupf90(genotype_file_name = genotype_file_name,map_file_name
=  map_file_name,out_name = out_name)
```