



**WORKSHOP
DE BIOINFORMÁTICA
APLICADA À GENÔMICA E
MELHORAMENTO ANIMAL**



AULA PRÁTICA 3

Métodos Bayesianos em GWS



Fabyano Fonseca e Silva

**Prof. Adjunto IV - Dep Zootecnia – UFV
Estatística Genômica e Bioinformática**

Campo Grande, 14/07 a 15/07 de 2014

Método Bayes A (pacote BGLR do R)

```
gen=read.table("gs1_gen.txt",h=T) #lendo arquivo genotipos
dim(gen)
head(gen)

freq=read.table("gs1_alefreq.txt",h=T) #lendo arquivo freq alel
dim(freq)
head(freq)

fen=read.table("gs1_fen.txt",h=T) #lendo arquivo fenotipos
dim(fen)
head(fen)

#corrigindo fenótipo para efeitos fixos - de los Campos et al. (2013)

y1=mean(fen$y) + lm(y~ factor(ym_slg) + factor(farm) + hcw, data=fen)$residuals

library(BGLR)
M=as.matrix(gen) # matriz de marcadores

BA=BGLR(y=y1,ETA=list(list(X=M,model='BayesA')),nIter=10000,burnIn=4000,thin=3)

a_hat_BA=BA$ETA[[1]]$b #vetor de efeitos estimados SNPs
snp_BA=data.frame(colnames(M),as.numeric(a_hat_BA))

colnames(snp_BA)=c("SNP","a_BA")
write.table(snp_BA,"snp_BA.txt", row.names=FALSE,quote=FALSE)
plot(a_hat_BA) #gráfico dos efeitos estimados SNPs

u_hat_BA=BA$yHat #valor genético genômico
```

Método Bayes A (Manhattan Plot)

```
library(gap) #pacote que contem o grafico Manhattan
```

```
map=read.table("mapa_gs1.txt", h=T)
```

```
#interseção de arquivos (SNPs com efeitos estimados e mapeados)
```

```
snp_BA1=merge(map,snp_BA,by=intersect("SNP","SNP"))
```

```
snp_BA2=snp_BA1[with(snp_BA1, order(chr,pos)),]
```

```
#ordenar chr e posicao
```

```
par(las=2, xpd=TRUE, cex.axis=0.9, cex=0.8)
```

```
color = cbind(rep(c("red","blue"),9),c("blue"))
```

```
#funcao mhplot
```

```
mhtplot(abs(snp_BA2[,-1]),mht.control(logscale=FALSE,  
colors=color,labels=cbind(seq(1,18),c("x")),srt=0), ylab="SNP  
effects - BayesA",pch=19)
```

```
axis(2, at = seq(0, 0.09, by = 0.01))
```

Método Bayes A (populações treinamento e validação)

```
library(BGLR)
M=as.matrix(gen)

M1=M[1:500,]
M2=M[501:634,]

y11=y1[1:500]
y12=y1[501:634]

BA=BGLR(y=y11,ETA=list(list(X=M1,model='BayesA')),nIter=10000,burnIn=4000,thin=3)

gebv2_BA=M2%*%BA$ETA[[1]]$b #GEBV para indivíduos da pop 2
cor(gebv2_BA,y12) #acurácia dada pela cor obs e pred
```

Método Bayes B (pacote BGLR do R)

```
gen=read.table("gs1_gen.txt",h=T) #lendo arquivo genotipos
dim(gen)
head(gen)

freq=read.table("gs1_alefreq.txt",h=T) #lendo arquivo freq alel
dim(freq)
head(freq)

fen=read.table("gs1_fen.txt",h=T) #lendo arquivo fenotipos
dim(fen)
head(fen)

#corrigindo fenótipo para efeitos fixos - de los Campos et al. (2013)

y1=mean(fen$y) + lm(y~ factor(ym_slg) + factor(farm) + hcw, data=fen)$residuals

library(BGLR)
M=as.matrix(gen) # matriz de marcadores

BB=BGLR(y=y1,ETA=list(list(X=M,model='BayesB',probIn=0.7)),nIter=10000,burnIn=4000,
thin=3)

a_hat_BB=BB$ETA[[1]]$b #vetor de efeitos estimados SNPs
snp_BB=data.frame(colnames(M),as.numeric(a_hat_BB))

colnames(snp_BB)=c("SNP","a_BB")
write.table(snp_BB,"snp_BB.txt", row.names=FALSE,quote=FALSE)
plot(a_hat_BB) #gráfico dos efeitos estimados SNPs

u_hat_BB=BB$yHat #valor genético genômico
```

Método Bayes B (Manhattan Plot)

```
library(gap) #pacote que contem o grafico Manhattan
```

```
map=read.table("mapa_gs1.txt", h=T)
```

```
#interseção de arquivos (SNPs com efeitos estimados e mapeados)
```

```
snp_BB1=merge(map,snp_BB,by=intersect("SNP","SNP"))
```

```
snp_BB2=snp_BB1[with(snp_BB1, order(chr,pos)),]
```

```
#ordenar chr e posicao
```

```
par(las=2, xpd=TRUE, cex.axis=0.9, cex=0.8)
```

```
color = cbind(rep(c("red","blue"),9),c("blue"))
```

```
#funcao mhplot
```

```
mhtplot(abs(snp_BB2[,-1]),mht.control(logscale=FALSE,  
colors=color,labels=cbind(seq(1,18),c("x")),srt=0), ylab="SNP  
effects - BayesB",pch=19)
```

```
axis(2, at = seq(0, 0.09, by = 0.01))
```

Método Bayes B(populações treinamento e validação)

```
library(BGLR)
M=as.matrix(gen)

M1=M[1:500,]
M2=M[501:634,]

y11=y1[1:500]
y12=y1[501:634]

BB=BGLR(y=y11,ETA=list(list(X=M1,model='BayesB')),nIter=10000,burnIn=4000,thin=3)

gebv2_BB=M2%*%BB$ETA[[1]]$b #GEBV para indivíduos da pop 2
cor(gebv2_BB,y12) #acurácia dada pela cor obs e pred
```

Método Bayes LASSO (pacote BGLR do R)

```
gen=read.table("gs1_gen.txt",h=T) #lendo arquivo genotipos
dim(gen)
head(gen)

freq=read.table("gs1_alefreq.txt",h=T) #lendo arquivo freq alel
dim(freq)
head(freq)

fen=read.table("gs1_fen.txt",h=T) #lendo arquivo fenotipos
dim(fen)
head(fen)

#corrigindo fenótipo para efeitos fixos - de los Campos et al. (2013)

y1=mean(fen$y) + lm(y~ factor(ym_slg) + factor(farm) + hcw, data=fen)$residuals

library(BGLR)
M=as.matrix(gen) # matriz de marcadores

BL=BGLR(y=y1,ETA=list(list(X=M,model='BL')),nIter=10000,burnIn=4000,thin=3)

a_hat_BL=BL$ETA[[1]]$b #vetor de efeitos estimados SNPs
snp_BL=data.frame(colnames(M),as.numeric(a_hat_BL))

colnames(snp_BL)=c("SNP","a_BL")
write.table(snp_BL,"snp_BL.txt", row.names=FALSE,quote=FALSE)
plot(a_hat_BL) #gráfico dos efeitos estimados SNPs

u_hat_BL=BL$yHat #valor genético genômico
```


Método Bayes LASSO (Manhattan Plot)

```
library(gap) #pacote que contem o grafico Manhattan
```

```
map=read.table("mapa_gs1.txt", h=T)
```

```
#interseção de arquivos (SNPs com efeitos estimados e mapeados)
```

```
snp_BL1=merge(map,snp_BL,by=intersect("SNP","SNP"))
```

```
snp_BL2=snp_BL1[with(snp_BL1, order(chr,pos)),]
```

```
#ordenar chr e posicao
```

```
par(las=2, xpd=TRUE, cex.axis=0.9, cex=0.8)
```

```
color = cbind(rep(c("red","blue"),9),c("blue"))
```

```
#funcao mhplot
```

```
mhtplot(abs(snp_BL2[,-1]),mht.control(logscale=FALSE,  
colors=color,labels=cbind(seq(1,18),c("x")),srt=0), ylab="SNP  
effects - BayesB",pch=19)
```

```
axis(2, at = seq(0, 0.09, by = 0.01))
```

Método Bayes LASSO(populações treinamento e validação)

```
library(BGLR)
M=as.matrix(gen)

M1=M[1:500,]
M2=M[501:634,]

y11=y1[1:500]
y12=y1[501:634]

BL=BGLR(y=y11,ETA=list(list(X=M1,model='BL')),nIter=10000,burnIn=4000,thin=3)

gebv2_BL=M2%*%BL$ETA[[1]]$b #GEBV para indivíduos da pop 2
cor(gebv2_BL,y12) #acurácia dada pela cor obs e pred
```

FastBayesA: executável Blup_FastBayesB-Meuwissen (software GWP)

```
?=====
?   Genome wide prediction of EBU using BLUP
?   Theo Meuwissen
?   Norwegian University of Life Sciences, As, Norway
?
?   June 2009
?=====
How many SNPs?
2500
How many phenotypes?
634
What is the genetic variance (will be divided by #SNPs to get the per SNP varia
nce)?
2
What is the error variance (actually only ratio of error to genetic variance ne
eds to be OK)?
4
Filename of SNP genotypes; 0,1,2, or 9 (=missing))?
M.txt
Filename of phenotypes (optional: additional column of weights))?
(weights may reflect 'no. of daughters'; weight=0: phenotype is not used, but E
BU is calculated)
y1.txt
Which solver option is to be used?
  1 : GWBLUP
  2 : fastBayesB (Genet Sel Evol 2009; 41:2) (note: this option does not a
llow for Weights)
2
No of records      =      634
using fastBayesB
What is the prior probability of a SNP having an effect (1: implies BayesA)
1
```

FastBayesA: lendo resultados Blup_FastBayesB-Mewissen (software GWP) no R

```
id = read.table("id_names.txt")      #lendo arquivo com nomes de ID's
snp = read.table("snp_names.txt")    #lendo arquivo com nomes de SNPs

u_bayA=data.frame(read.table("fast_bb.EBV")[,2]) #lendo arquivo de GEBVs
rownames(u_bayA)=id[,1]
colnames(u_bayA)=c("GEBV")
u_bayA = as.matrix(u_bayA)

plot(hist(u_bayA))    #distribuição dos GEBVs
top10=quantile(u_bayA[,1], probs = c(0.9)) #top10%
top10

# individuos selecionados e respectivos GEBVs
sel_id=data.frame(u_bayA[,1][u_bayA[,1]>=top10])
colnames(sel_id)=c("GEBV")

#h2 aproximada
y1= read.table("y1.txt")      #fenótipo lido no GWP
sig2e=var(y1-u_bayA)
sig2u= var(u_bayA)
h2_bayA=sig2u/(sig2u + sig2e)

a_bayA=read.table("fast_bb.SOL") #lendo arquivo de efeitos marcadores
a_bayA1=cbind(snp,a_bayA)
colnames(a_bayA1)=c("SNP","a_bayA")
write.table(a_bayA,"snp_bayA.txt", row.names=FALSE,quote=FALSE)
plot(a_bayA1[,2])
```

FastBayesA: cross-valid no executável Blup_FastBayesB-Meuwissen (software GWP)

OBS.: os arquivos y11,y12, M1 e M2 são os mesmos usados para RR-BLUP e BL

```
=====
?      Genome wide prediction of EBU using BLUP
?      Theo Meuwissen
?      Norwegian University of Life Sciences, As, Norway
?
?      June 2009
?=====
How many SNPs?
2500
How many phenotypes?
500
What is the genetic variance (will be divided by #SNPs to get the per SNP vari
nce)?
2
What is the error variance (actually only ratio of error to genetic variance ne
eds to be OK)?
4
Filename of SNP genotypes; 0,1,2, or 9 (=missing))?
M1.txt
Filename of phenotypes (optional: additional column of weights))?
(y11.txt)
<weights may reflect 'no. of daughters'; weight=0: phenotype is not used, but
BU is calculated)
y11.txt
Which solver option is to be used?
1 : GWBLUP
2 : fastBayesB (Genet Sel Evol 2009; 41:2) (note: this option does not
allow for Weights)
2
No of records      =      500
using fastBayesB
What is the prior probability of a SNP having an effect (1: implies BayesA)
1_
```

FastBayesA: lendo resultados Croos-validação (Blup_FastBayesB-Mewissen) no R

```
#lendo arquivo de efeitos SNPs
a_bayA=read.table("fast_bb.SOL")

#lendo arquivo de genótipos da pop de validação
M2=read.table("M2.txt")
GEBV2=as.matrix(M2)*%as.matrix(a_bayA)

#lendo arquivo de fenótipos da pop de validação
y12=read.table("y12.txt")

#calculando a acurácia do método BayesA

cor(y12, GEBV2)
```

FastBayesB com $\pi=0.7$: executável Blup_FastBayesB-Meuwissen (software GWP)

```
?=====
?   Genome wide prediction of EBU using BLUP
?   Theo Meuwissen
?   Norwegian University of Life Sciences, Ås, Norway
?
?   June 2009
?=====
How many SNPs?
2500
How many phenotypes?
634
What is the genetic variance (will be divided by #SNPs to get the per SNP variance)?
2
What is the error variance (actually only ratio of error to genetic variance needs to be OK)?
4
Filename of SNP genotypes; 0,1,2, or 9 (=missing))?
M.txt
Filename of phenotypes (optional: additional column of weights)?
(weights may reflect 'no. of daughters'; weight=0: phenotype is not used, but EBU is calculated)
y1.txt
Which solver option is to be used?
  1 : GWBLUP
  2 : fastBayesB (Genet Sel Evol 2009; 41:2) (note: this option does not allow for Weights)
2
No of records      =      634
using fastBayesB
What is the prior probability of a SNP having an effect (1: implies BayesA)
0.7
```

FastBayesB com $\pi=0.7$: lendo resultados Blup_FastBayesB-Mewissen (software GWP) no R

```
id = read.table("id_names.txt")    #lendo arquivo com nomes de ID's
snp = read.table("snp_names.txt")  #lendo arquivo com nomes de SNPs

u_bayB=data.frame(read.table("fast_bb.EBV")[,2]) #lendo arquivo de GEBVs
rownames(u_bayB)=id[,1]
colnames(u_bayB)=c("GEBV")
u_bayB = as.matrix(u_bayB)

plot(hist(u_bayB))    #distribuição dos GEBVs
top10=quantile(u_bayB[,1], probs = c(0.9)) #top10%
top10

# individuos selecionados e respectivos GEBVs
sel_id=data.frame(u_bayB[,1][u_bayB[,1]>=top10])
colnames(sel_id)=c("GEBV")

#h2 aproximada
y1= read.table("y1.txt")    #fenótipo lido no GWP
sig2e=var(y1-u_bayB)
sig2u= var(u_bayB)
h2_bayB=sig2u/(sig2u + sig2e)

a_bayB=read.table("fast_bb.SOL") #lendo arquivo de efeitos marcadores
a_bayB1=cbind(snp,a_bayB)
colnames(a_bayB1)=c("SNP","a_bayB")
write.table(a_bayB,"snp_bayB.txt", row.names=FALSE,quote=FALSE)
plot(a_bayB1[,2])
```


FastBayesB com $\pi=0.7$: executável Blup_FastBayesB-Meuwissen (software GWP)

```
!=====
!   Genome wide prediction of EBU using BLUP
!   Theo Meuwissen
!   Norwegian University of Life Sciences, Ås, Norway
!
!   June 2009
!=====
How many SNPs?
2500
How many phenotypes?
500
What is the genetic variance (will be divided by #SNPs to get the per SNP variance)?
2
What is the error variance (actually only ratio of error to genetic variance needs to be OK)?
4
Filename of SNP genotypes; 0,1,2, or 9 (=missing))?
M1.txt
Filename of phenotypes (optional: additional column of weights)?
(weights may reflect 'no. of daughters'; weight=0: phenotype is not used, but EBU is calculated)
y11.txt
Which solver option is to be used?
  1 : GWBLUP
  2 : fastBayesB (Genet Sel Evol 2009; 41:2) (note: this option does not allow for Weights)
2
No of records          =          500
using fastBayesB
What is the prior probability of a SNP having an effect (1: implies BayesA)
0.7
```

FastBayesB: lendo resultados Croos-validação (Blup_FastBayesB-Mewissen) no R

```
#lendo arquivo de efeitos SNPs
a_bayB=read.table("fast_bb.SOL")

#lendo arquivo de genótipos da pop de validação
M2=read.table("M2.txt")
GEBV2=as.matrix(M2)*%as.matrix(a_bayB)

#lendo arquivo de fenótipos da pop de validação
y12=read.table("y12.txt")

#calculando a acurácia do método BayesB

cor(y12, GEBV2)
```

#Geração dos arquivos de genótipos e fenótipos para serem usados no GS3

```
M=read.table("M.txt") #lendo matriz genótipos (mesma BayesA e B)
id_novo=format(seq(1,nrow(M)),digits=3,width=3) # ID's
renumerados (1 até numero IDs) e FORMATADOS

y1=read.table("y1.txt") #lendo arquivo com fenótipos

write.table(M,"Mgs3_0.txt",col.names=FALSE,row.names=FALSE,sep=" ",
, quote=FALSE)

Mgs3_0=read.table("Mgs3_0.txt",colClasses="factor")

#inserindo col ID's na matriz Mgs3_0
Mgs3=cbind(id_novo,Mgs3_0)

write.table(Mgs3,"Mgs3.txt",row.names=FALSE,col.names=FALSE,quote
=FALSE) #exportando a matrix Mgs3

fen=cbind(id_novo,y1,rep(1,nrow(M))) # 3 cols: id, fen e col 1's

write.table(fen,"y1gs3.txt",row.names=FALSE,col.names=FALSE,quote
=FALSE)
```

BayesCpi: código GS3

DATAFILE

y1gs3.txt

PEDIGREE FILE

GENOTYPE FILE

Mgs3.txt

NUMBER OF LOCI (might be 0)

2500

METHOD (BLUP/MCMCBLUP/VCE/PREDICT)

VCE

SIMULATION

F

GIBBS SAMPLING PARAMETERS

NITER

3000

BURNIN

1000

THIN

2

CONV_CRIT (MEANINGFULL IF BLUP)

1d-4

CORRECTION (to avoid numerical problems)

1000

VARIANCE COMPONENTS SAMPLES

bayesCpi_var

SOLUTION FILE

bayesCpi_sol

TRAIT AND WEIGHT COLUMNS

2 0 #weight

NUMBER OF EFFECTS

2

POSITION IN DATA FILE TYPE OF EFFECT NUMBER OF LEVELS

3 cross 1

4 add_SNP 0

VARIANCE COMPONENTS (fixed for any BLUP, starting values for VCE)

vara

0.001 2

vard

0 2

varg

1.08 2

varp

0 2

vare

3.44 2

RECORD ID

1

CONTINUATION (T/F)

F

MODEL (T/F for each effect)

T T T T T

A PRIORI a

1 1

a PRIORI D

1 1

USE MIXTURE (BAYES Cpi)

T

BayesCPi : lendo resultados do software GS3 no R

#lendo arquivo de efeitos SNPs

```
sol=read.table("bayesCpi_sol",h=T)
a_bayC=sol$solution[-1]
```

#lendo arquivo de genótipos

```
M=as.matrix(read.table("M.txt"))
```

u_bayC=M%*%a_bayC #EBV para bayesCPi

```
id=read.table("id_names.txt") #Id's originais
```

```
rownames(u_bayC)=t(id)
```

```
plot(hist(u_bayC)) #distribuição dos GEBVs
```

```
top10=quantile(u_bayC[,1], probs = c(0.9)) #top10%
top10
```

individuos selecionados e respectivos GEBVs

```
sel_id=data.frame(u_bayC[,1][u_bayC[,1]>=top10])
colnames(sel_id)=c("GEBV")
```

calculando h2

```
var=read.table("bayesCpi_var",h=T) #lendo cadeias MCMC
```

```
varg=var$X2varapppi
```

```
vare=var$vare
```

```
h2=varg/(varg+vare)
```

```
summary(h2)
```

#Geração dos arquivos de genótipos e fenótipos para CROSS VALID - GS3

```
M=read.table("M.txt") #lendo matriz genótipos (mesma BayesA e B)

M1=M[1:500,]

id_novo_CV=format(seq(1,nrow(M1)),digits=3,width=3) # ID's
renumerados (1 até numero IDs) e FORMATADOS

y1=read.table("y1.txt") #lendo arquivo com fenótipos
y11=y1[1:500,]

write.table(M1,"Mgs3_01.txt",col.names=FALSE,row.names=FALSE,sep=" ",
quote=FALSE)

Mgs3_01=read.table("Mgs3_01.txt",colClasses="factor")

#inserindo col ID's na matriz Mgs3_0
Mgs31=cbind(id_novo_CV,Mgs3_01)

write.table(Mgs31,"Mgs31.txt",row.names=FALSE,col.names=FALSE,quote=
FALSE) #exportando a matrix Mgs31

fen=cbind(id_novo_CV,y11,rep(1,nrow(M1))) # 3 cols: id, fen e col
1's
write.table(fen,"y1gs31.txt",row.names=FALSE,col.names=FALSE,quote=F
ALSE)
```

BayesCπ: código GS3 – cross valid

DATAFILE

y1gs31.txt

PEDIGREE FILE

Arquivos de
validação

GENOTYPE FILE

Mgs31.txt

NUMBER OF LOCI (might be 0)

2500

METHOD (BLUP/MCMCBLUP/VCE/PREDICT)

VCE

SIMULATION

F

GIBBS SAMPLING PARAMETERS

NITER

3000

BURNIN

1000

THIN

2

CONV_CRIT (MEANINGFULL IF BLUP)

1d-4

CORRECTION (to avoid numerical problems)

1000

VARIANCE COMPONENTS SAMPLES

bayesCpi_var

SOLUTION FILE

bayesCpi_sol

TRAIT AND WEIGHT COLUMNS

2 0 #weight

NUMBER OF EFFECTS

2

POSITION IN DATA FILE TYPE OF EFFECT NUMBER OF LEVELS

3 cross 1

4 add_SNP 0

VARIANCE COMPONENTS (fixed for any BLUP, starting values for VCE)

vara

0.001 2

vard

0 2

varg

1.08 2

varp

0 2

vare

3.44 2

RECORD ID

1

CONTINUATION (T/F)

F

MODEL (T/F for each effect)

T T T T T

A PRIORI a

1 1

a PRIORI D

1 1

USE MIXTURE (BAYES Cpi)

T

BayesCPi : lendo resultados do software GS3 no R – CROSS VALID

```
M=read.table("M.txt") #lendo matriz genótipos (mesma BayesA e B)
M2=as.matrix(M[501:634,]) #genótipos população validação
```

```
#lendo arquivo de efeitos SNPs
```

```
sol_cv=read.table("bayesCpi_sol_cv",h=T)
a_bayC_cv=sol_cv$solution[-1]
```

```
u_bayC_cv=M2%*%a_bayC_cv #EGBV para populacao de valid
```

```
y1=read.table("y1.txt") #lendo arquivo com fenótipos
y12=y1[501:634,] #fenotipo para populacao de valid
```

```
#calculando a acurácia do método BayesCPi
```

```
cor(y12, u_bayC_cv)
```


Improved Bayes LASSO: código GS3

DATAFILE

y1gs3.txt

PEDIGREE FILE

GENOTYPE FILE

Mgs3.txt

NUMBER OF LOCI (might be 0)

2500

METHOD (BLUP/MCMCBLUP/VCE/PREDICT)

VCE

SIMULATION

F

GIBBS SAMPLING PARAMETERS

NITER

3000

BURNIN

1000

THIN

2

CONV_CRIT (MEANINGFULL IF BLUP)

1d-4

CORRECTION (to avoid numerical problems)

1000

VARIANCE COMPONENTS SAMPLES

iblasso_var

SOLUTION FILE

iblasso_sol

TRAIT AND WEIGHT COLUMNS

2 0 #weight

Comp var

Efeitos SNPs

NUMBER OF EFFECTS

2

POSITION IN DATA FILE TYPE OF EFFECT NUMBER OF LEVELS

3 cross 1

4 add_SNP 0

VARIANCE COMPONENTS (fixed for any BLUP, starting values for VCE)

vara

0.001 2

vard

0 2

varg

1.08 2

varp

0 2

vare

3.44 2

RECORD ID

1

CONTINUATION (T/F)

F

MODEL (T/F for each effect)

T T T T T

A PRIORI a

1 1

a PRIORI D

1 1

USE MIXTURE (BAYES Cpi)

F

OPTION BayesianLasso Tibshirani

iblasso : lendo resultados do software GS3 no R

#lendo arquivo de efeitos SNPs

```
sol=read.table("iblasso_sol",h=T)
a_iblasso=sol$solution[-1]
```

#lendo arquivo de genótipos

```
M=as.matrix(read.table("M.txt"))
```

```
u_iblasso=M%*%a_iblasso #EGBV para bayesCPi
```

```
id=read.table("id_names.txt") #Id's originais
```

```
rownames(u_iblasso)=t(id)
```

```
plot(hist(u_iblasso)) #distribuição dos GEBVs
```

```
top10=quantile(u_iblasso[,1], probs = c(0.9)) #top10%
top10
```

indivíduos selecionados e respectivos GEBVs

```
sel_id=data.frame(u_iblasso[,1][u_iblasso[,1]>=top10])
colnames(sel_id)=c("GEBV")
```

calculando h2

```
var=read.table("iblasso_var",h=T) #lendo cadeias MCMC
```

```
varg=var$X2varapqpi
```

```
vare=var$vare
```

```
h2=varg/(varg+vare)
```

```
summary(h2)
```

Improved bayesian LASSO: código GS3 – cross valid

DATAFILE

y1gs31.txt

PEDIGREE FILE

Arquivos de
validação

GENOTYPE FILE

Mgs31.txt

NUMBER OF LOCI (might be 0)

2500

METHOD (BLUP/MCMCBLUP/VCE/PREDICT)

VCE

SIMULATION

F

GIBBS SAMPLING PARAMETERS

NITER

3000

BURNIN

1000

THIN

2

CONV_CRIT (MEANINGFULL IF BLUP)

1d-4

CORRECTION (to avoid numerical problems)

1000

VARIANCE COMPONENTS SAMPLES

bayesCpi_var

SOLUTION FILE

bayesCpi_sol

TRAIT AND WEIGHT COLUMNS

2 0 #weight

NUMBER OF EFFECTS

2

POSITION IN DATA FILE TYPE OF EFFECT NUMBER OF LEVELS

3 cross 1

4 add_SNP 0

VARIANCE COMPONENTS (fixed for any BLUP, starting values for VCE)

vara

0.001 2

vard

0 2

varg

1.08 2

varp

0 2

vare

3.44 2

RECORD ID

1

CONTINUATION (T/F)

F

MODEL (T/F for each effect)

T T T T T

A PRIORI a

1 1

a PRIORI D

1 1

USE MIXTURE (BAYES Cpi)

F

OPTION BayesianLasso Tibshirani

Iblasso: lendo resultados do software GS3 no R – CROSS VALID

```
M=read.table("M.txt") #lendo matriz genótipos (mesma BayesA e B)
M2=as.matrix(M[501:634,]) #genótipos população validação

#lendo arquivo de efeitos SNPs
sol_cv=read.table("iblasso_sol_cv",h=T)
a_iblasso_cv=sol_cv$solution[-1]

u_iblasso_cv=M2%*%a_iblasso_cv #EGBV para populacao de valid

y1=read.table("y1.txt") #lendo arquivo com fenótipos
y12=y1[501:634,] #fenotipo para populacao de valid

#calculando a acurácia do método IBLASSO

cor(y12, u_bayC_cv)
```