```
include("/home/nicole/Jupyter/SSBRnoJ/src/SSBR.jl")
In [1]:
        using SSBR
In [2]: function getPos(ped,IDs)
            posAi = Array(Int64, size(IDs, 1))
            for (i,id) = enumerate(IDs[:,1])
                posAi[i] = ped.idMap[id].seqID
            end
            return posAi
        end
Out[2]: getPos (generic function with 1 method)
In [3]: ; cd Data/0.5g2k/G/10
        /home/nicole/Jupyter/JG3/Data/0.5g2k/G/10
In [4]:
        ;ls
        Correlation.G5.G.J.txt
        Correlation.G5.G.JC.txt
        Correlation.G5.G.PBLUP.txt
        G0.Genotype.ID
        G0.ID
        G0.noGenotype.ID
        G1.Genotype.ID
        G1.ID
        G1.noGenotype.ID
        G2.Genotype.ID
        G2.ID
        G2.noGenotype.ID
        G3.Genotype.ID
        G3.ID
        G3.noGenotype.ID
        G4.Genotype.ID
        G4.ID
        G4.noGenotype.ID
        G5.Genotype.ID
```

```
In [5]: |;awk '{print $1}' PedAll.txt | sort -b > all.ID
 In [6]: | ;awk '{print $1}' GenNF.txt | sort -b > genotype.ID
 In [7]: |;join -v1 all.ID genotype.ID > noGenotype.ID
 In [8]: | ;awk '{print $1,$2}' Phe.txt > sim.phenotype
 In [9]: | ;awk '{print $1,$3}' PheAll.txt > sim.bv
In [10]: ; awk 'NR >=1 && NR <=8000 {print $1}' PedAll.txt | sort -b > G0.ID
In [11]: ; awk 'NR >=8001 && NR <=16000 {print $1}' PedAll.txt | sort -b > G1.ID
In [12]: ; awk 'NR >=16001 && NR <=24000 {print $1}' PedAll.txt | sort -b > G2.ID
In [13]: ; awk 'NR >=24001 && NR <=32000 {print $1}' PedAll.txt | sort -b > G3.ID
In [14]: ; awk 'NR >=32001 && NR <=40000 {print $1}' PedAll.txt | sort -b > G4.ID
In [15]: ; awk 'NR >=40001 && NR <=48000 {print $1}' PedAll.txt | sort -b > G5.ID
In [16]: ;join GO.ID genotype.ID > GO.Genotype.ID
In [17]: |;join G1.ID genotype.ID > G1.Genotype.ID
In [18]: ;join G2.ID genotype.ID > G2.Genotype.ID
In [19]: ;join G3.ID genotype.ID > G3.Genotype.ID
In [20]: |;join G4.ID genotype.ID > G4.Genotype.ID
In [21]: ;join G5.ID genotype.ID > G5.Genotype.ID
```

```
In [22]: |;join -v1 G0.ID genotype.ID > G0.noGenotype.ID
In [23]: ;join -v1 G1.ID genotype.ID > G1.noGenotype.ID
In [24]: ;join -v1 G2.ID genotype.ID > G2.noGenotype.ID
         ; join -v1 G3.ID genotype.ID > G3.noGenotype.ID
In [25]:
         ; join -v1 G4.ID genotype.ID > G4.noGenotype.ID
In [26]:
         ;join -v1 G5.ID genotype.ID > G5.noGenotype.ID
In [27]:
         ;wc G0.Genotype.ID;wc G1.Genotype.ID;wc G2.Genotype.ID;wc G3.Genotype.ID;wc G4.Genotype.ID;wc G5.Genotype
In [28]:
          200 200 1200 GO.Genotype.ID
          200 200 1200 G1.Genotype.ID
          200 200 1200 G2.Genotype.ID
          200 200 1200 G3.Genotype.ID
          200 200 1200 G4.Genotype.ID
          8000 8000 48000 G5.Genotype.ID
         ;wc G0.noGenotype.ID;wc G1.noGenotype.ID;wc G2.noGenotype.ID;wc G3.noGenotype.ID;wc G4.noGenotype.ID;wc G
In [29]:
          7800 7800 46800 G0.noGenotype.ID
          7800 7800 46800 Gl.noGenotype.ID
          7800 7800 46800 G2.noGenotype.ID
          7800 7800 46800 G3.noGenotype.ID
          7800 7800 46800 G4.noGenotype.ID
         0 0 0 G5.noGenotype.ID
In [30]:
         ped,A Mats,numSSBayes = calc Ai("PedAll.txt", "genotype.ID", calculateInbreeding=false)
         nothing
         df
                = read genotypes("GenNF.txt", numSSBayes)
                                                                                 # with centering
         M Mats = make MMats(df, A Mats, ped, center=true);
         y Vecs = make yVecs("sim.phenotype",ped,numSSBayes);
         Z Mats = make ZMats(ped,y Vecs,numSSBayes)
         X Mats, W Mats = make_XWMats(Z_Mats,M_Mats,numSSBayes)
                                                                                 # no J
         nothing
```

```
In [31]:
         vG
                 = 0.858
                = 0.858
         vRes
         nIter = 50000
         @time aHat1,alphaHat,betaHat,epsiHat =
         ssGibbs(M Mats,y Vecs, Z Mats, X Mats, W Mats, A Mats, numSSBayes, vRes, vG, nIter, outFreq=5000);
         nothing
         This is iteration 5000
         This is iteration 10000
         This is iteration 15000
         This is iteration 20000
         This is iteration 25000
         This is iteration 30000
         This is iteration 35000
         This is iteration 40000
         This is iteration 45000
         This is iteration 50000
         7103.608576 seconds (23.86 G allocations: 736.886 GB, 2.79% gc time)
In [32]: betaHat
Out[32]: 1-element Array{Float64,1}:
          33.9412
In [33]: using DataFrames
In [34]: | df = readtable("sim.bv", eltypes =[UTF8String, Float64], separator = ' ',header=false)
         a = Array(Float64, numSSBayes.num ped)
         for (i,ID) in enumerate(df[:,1])
             j = ped.idMap[ID].seqID
             a[j] = df[i,2]
         end
```

```
In [35]: IDs = readtable("all.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor1 = cor(a[posAi],aHat1[posAi])[1,1]
         reg1 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - all.ID : correlation = %6.3f\n", cor1 ) # with epsilon
         @printf("SSBRJC from Gibbs - all.ID : regression of TBV on GEBV = %6.3f\n", reg1)
         JCAll = cor1
         SSBRJC from Gibbs - all.ID : correlation = 0.912
         SSBRJC from Gibbs - all.ID : regression of TBV on GEBV = 0.981
Out[35]: 0.9121902427393044
In [36]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[36]: 0.16540829799163542
In [37]: IDs = readtable("genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor2 = cor(a[posAi],aHat1[posAi])[1,1]
         reg2 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - genotype.ID : correlation = %6.3f\n", cor2 ) # with epsilon
         @printf("SSBRJC from Gibbs - genotype.ID : regression of TBV on GEBV = %6.3f\n", reg2)
         JCAll = cor2
         SSBRJC from Gibbs - genotype.ID : correlation = 0.913
         SSBRJC from Gibbs - genotype.ID : regression of TBV on GEBV = 1.076
Out[37]: 0.9133195233274184
In [38]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[38]: 1.441777147086522
```

```
In [39]: IDs = readtable("noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor3 = cor(a[posAi],aHat1[posAi])[1,1]
         reg3 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - noGenotype.ID : correlation = %6.3f\n", cor3 ) # with epsilon
         @printf("SSBRJC from Gibbs - noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg3)
         JCAll = cor3
         SSBRJC from Gibbs - noGenotype.ID : correlation = 0.881
         SSBRJC from Gibbs - noGenotype.ID : regression of TBV on GEBV = 0.965
Out[39]: 0.8813439551936759
In [40]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[40]: -0.1291383594917998
In [41]: IDs = readtable("G0.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor4 = cor(a[posAi],aHat1[posAi])[1,1]
         reg4 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G0.ID : correlation = %6.3f\n", cor4 ) # with epsilon
         @printf("SSBRJC from Gibbs - G0.ID : regression of TBV on GEBV = %6.3f\n", req4)
         JCAll = cor4
         SSBRJC from Gibbs - G0.ID : correlation = 0.750
         SSBRJC from Gibbs - G0.ID: regression of TBV on GEBV = 0.937
Out[41]: 0.7496529278463033
In [42]: GEBV = aHat1[posAi]
         G1GEBV=mean(GEBV)
Out[42]: -1.409157052691542
```

```
In [43]: IDs = readtable("G1.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor4 = cor(a[posAi],aHat1[posAi])[1,1]
         reg4 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G1.ID : correlation = %6.3f\n", cor4 ) # with epsilon
         @printf("SSBRJC from Gibbs - G1.ID : regression of TBV on GEBV = %6.3f\n", req4)
         JCAll = cor4
         SSBRJC from Gibbs - G1.ID : correlation = 0.776
         SSBRJC from Gibbs - G1.ID: regression of TBV on GEBV = 0.994
Out[43]: 0.7761437931991035
In [44]: GEBV = aHat1[posAi]
         G1GEBV=mean(GEBV)
Out[44]: -0.6463578054477354
In [45]: IDs = readtable("G2.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor5 = cor(a[posAi],aHat1[posAi])[1,1]
         reg5 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G2.ID : correlation = %6.3f\n", cor5 ) # with epsilon
         @printf("SSBRJC from Gibbs - G2.ID : regression of TBV on GEBV = %6.3f\n", reg5)
         JCAll = cor5
         SSBRJC from Gibbs - G2.ID : correlation = 0.764
         SSBRJC from Gibbs - G2.ID : regression of TBV on GEBV = 0.996
Out[45]: 0.764034962538284
In [46]: GEBV = aHat1[posAi]
         G2GEBV=mean(GEBV)
Out[46]: -0.005660384389752501
```

```
In [47]: IDs = readtable("G3.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor6 = cor(a[posAi],aHat1[posAi])[1,1]
         reg6 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G3.ID : correlation = %6.3f\n", cor6 ) # with epsilon
         @printf("SSBRJC from Gibbs - G3.ID : regression of TBV on GEBV = %6.3f\n", reg6)
         JCAll = cor6
         SSBRJC from Gibbs - G3.ID : correlation = 0.755
         SSBRJC from Gibbs - G3.ID : regression of TBV on GEBV = 0.996
Out[47]: 0.7546225325294572
In [48]: GEBV = aHat1[posAi]
         G3GEBV=mean(GEBV)
Out[48]: 0.5198095323723287
In [49]: IDs = readtable("G4.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor7 = cor(a[posAi],aHat1[posAi])[1,1]
         reg7 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G4.ID : correlation = %6.3f\n", cor7 ) # with epsilon
         @printf("SSBRJC from Gibbs - G4.ID : regression of TBV on GEBV = %6.3f\n", req7)
         JCAll = cor7
         SSBRJC from Gibbs - G4.ID : correlation = 0.756
         SSBRJC from Gibbs - G4.ID : regression of TBV on GEBV = 0.987
Out[49]: 0.7556953604165728
In [50]: GEBV = aHat1[posAi]
         G4GEBV=mean(GEBV)
Out[50]: 1.0470918814121346
```

```
In [51]: IDs = readtable("G5.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor8 = cor(a[posAi],aHat1[posAi])[1,1]
         reg8 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G5.ID : correlation = %6.3f\n", cor8 ) # with epsilon
         @printf("SSBRJC from Gibbs - G5.ID : regression of TBV on GEBV = %6.3f\n", reg8)
         JCAll = cor8
         SSBRJC from Gibbs - G5.ID : correlation = 0.900
         SSBRJC from Gibbs - G5.ID: regression of TBV on GEBV = 1.062
Out[51]: 0.899576474588086
In [52]: GEBV = aHat1[posAi]
         G5GEBV=mean(GEBV)
Out[52]: 1.4867236166943794
In [53]: IDs = readtable("G0.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor9 = cor(a[posAi],aHat1[posAi])[1,1]
         reg9 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G0.Genotype.ID : correlation = %6.3f\n", cor9 ) # with epsilon
         @printf("SSBRJC from Gibbs - G0.Genotype.ID : regression of TBV on GEBV = %6.3f\n", reg9)
         JCAll = cor9
         SSBRJC from Gibbs - G0.Genotype.ID : correlation = 0.922
         SSBRJC from Gibbs - G0.Genotype.ID: regression of TBV on GEBV = 1.064
Out[53]: 0.9219704156434453
In [54]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[54]: 0.18567102948500128
```

```
In [55]: IDs = readtable("G1.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor9 = cor(a[posAi],aHat1[posAi])[1,1]
         reg9 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G1.Genotype.ID : correlation = %6.3f\n", cor9 ) # with epsilon
         @printf("SSBRJC from Gibbs - G1.Genotype.ID : regression of TBV on GEBV = %6.3f\n", reg9)
         JCAll = cor9
         SSBRJC from Gibbs - G1.Genotype.ID : correlation = 0.936
         SSBRJC from Gibbs - G1.Genotype.ID : regression of TBV on GEBV = 1.084
Out[55]: 0.935568149429011
In [56]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[56]: 0.6737968936118078
In [57]: IDs = readtable("G2.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor10 = cor(a[posAi],aHat1[posAi])[1,1]
         reg10 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G2.Genotype.ID : correlation = %6.3f\n", cor10 ) # with epsilon
         @printf("SSBRJC from Gibbs - G2.Genotype.ID : regression of TBV on GEBV = %6.3f\n", req10)
         JCAll = cor10
         SSBRJC from Gibbs - G2.Genotype.ID : correlation = 0.928
         SSBRJC from Gibbs - G2.Genotype.ID : regression of TBV on GEBV = 1.064
Out[57]: 0.9275233310455199
In [58]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[58]: 1.0879893354615933
```

```
In [59]: IDs = readtable("G3.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor11 = cor(a[posAi],aHat1[posAi])[1,1]
         reg11 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G3.Genotype.ID : correlation = %6.3f\n", cor11 ) # with epsilon
         @printf("SSBRJC from Gibbs - G3.Genotype.ID : regression of TBV on GEBV = %6.3f\n", req11)
         JCAll = cor11
         SSBRJC from Gibbs - G3.Genotype.ID : correlation = 0.930
         SSBRJC from Gibbs - G3.Genotype.ID : regression of TBV on GEBV = 1.095
Out[59]: 0.9295480357762214
In [60]: GEBV = aHat1[posAi]
         mean(GEBV)
Out[60]: 1.5824385159480179
In [61]: IDs = readtable("G4.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor12 = cor(a[posAi],aHat1[posAi])[1,1]
         reg12 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G4.Genotype.ID : correlation = %6.3f\n", cor12 ) # with epsilon
         @printf("SSBRJC from Gibbs - G4.Genotype.ID : regression of TBV on GEBV = %6.3f\n", reg12)
         JCAll = cor12
         SSBRJC from Gibbs - G4.Genotype.ID : correlation = 0.859
         SSBRJC from Gibbs - G4.Genotype.ID : regression of TBV on GEBV = 1.030
Out[61]: 0.8590216321834214
In [62]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[62]: 1.8811311766118775
```

```
In [63]: IDs = readtable("G5.Genotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor13 = cor(a[posAi],aHat1[posAi])[1,1]
         reg13 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G5.Genotype.ID : correlation = %6.3f\n", cor13 ) # with epsilon
         @printf("SSBRJC from Gibbs - G5.Genotype.ID : regression of TBV on GEBV = %6.3f\n", req13)
         JCAll = cor13
         SSBRJC from Gibbs - G5.Genotype.ID : correlation = 0.900
         SSBRJC from Gibbs - G5.Genotype.ID: regression of TBV on GEBV = 1.062
Out[63]: 0.899576474588086
In [64]: writedlm("Correlation.G5.G.C.txt",cor13)
In [65]: writedlm("Regression.G5.G.C.txt",reg13)
In [66]: GEBVG5Gall = aHat1[posAi]
         GEBVG5G=mean(GEBVG5Gall)
Out[66]: 1.4867236166943794
In [67]: IDs = readtable("G0.noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor14 = cor(a[posAi],aHat1[posAi])[1,1]
         reg14 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G0.noGenotype.ID : correlation = %6.3f\n", cor14 ) # with epsilon
         @printf("SSBRJC from Gibbs - G0.noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg14)
         JCAll = cor14
         SSBRJC from Gibbs - G0.noGenotype.ID : correlation = 0.732
         SSBRJC from Gibbs - G0.noGenotype.ID : regression of TBV on GEBV = 0.952
Out[67]: 0.7322213297544249
In [68]: GEBV = aHat1[posAi]
         mean(GEBV)
Out[68]: -1.4500500804396588
```

```
In [69]: IDs = readtable("G1.noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor14 = cor(a[posAi],aHat1[posAi])[1,1]
         reg14 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G1.noGenotype.ID : correlation = %6.3f\n", cor14 ) # with epsilon
         @printf("SSBRJC from Gibbs - G1.noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg14)
         JCAll = cor14
         SSBRJC from Gibbs - G1.noGenotype.ID : correlation = 0.760
         SSBRJC from Gibbs - G1.noGenotype.ID: regression of TBV on GEBV = 1.002
Out[69]: 0.759967581111568
In [70]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[70]: -0.6802079259364414
In [71]: IDs = readtable("G2.noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor15 = cor(a[posAi],aHat1[posAi])[1,1]
         reg15 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G2.noGenotype.ID : correlation = %6.3f\n", cor15 ) # with epsilon
         @printf("SSBRJC from Gibbs - G2.noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg15)
         JCAll = cor15
         SSBRJC from Gibbs - G2.noGenotype.ID : correlation = 0.749
         SSBRJC from Gibbs - G2.noGenotype.ID: regression of TBV on GEBV = 0.996
Out[71]: 0.7488262115674177
In [72]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[72]: -0.033702684898761354
```

```
In [73]: IDs = readtable("G3.noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor16 = cor(a[posAi],aHat1[posAi])[1,1]
         reg16 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G3.noGenotype.ID : correlation = %6.3f\n", cor16 ) # with epsilon
         @printf("SSBRJC from Gibbs - G3.noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg16)
         JCAll = cor16
         SSBRJC from Gibbs - G3.noGenotype.ID : correlation = 0.737
         SSBRJC from Gibbs - G3.noGenotype.ID : regression of TBV on GEBV = 0.991
Out[73]: 0.7373724470303493
In [74]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[74]: 0.4925626353575675
In [75]: IDs = readtable("G4.noGenotype.ID", eltypes =[UTF8String], separator = ' ',header=false)
         posAi = getPos(ped,IDs)
         cor17 = cor(a[posAi],aHat1[posAi])[1,1]
         reg17 = linreg(aHat1[posAi], a[posAi])[2,1]
         @printf("SSBRJC from Gibbs - G4.noGenotype.ID : correlation = %6.3f\n", cor17 ) # with epsilon
         @printf("SSBRJC from Gibbs - G4.noGenotype.ID : regression of TBV on GEBV = %6.3f\n", reg17)
         JCAll = cor17
         SSBRJC from Gibbs - G4.noGenotype.ID : correlation = 0.744
         SSBRJC from Gibbs - G4.noGenotype.ID: regression of TBV on GEBV = 0.980
Out[75]: 0.7435081605607942
In [76]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[76]: 1.025706258458295
In [77]: numSSBayes
Out[77]: SSBR.NumSSBayes(54904,45904,9000,40000,39000,1000,2000)
```