```
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/6
        /home/nicole/Jupyter/JG3/Data/0.5g2k/G/6
In [4]:
        ;ls
        GEIINT . LAL
        PedAll.txt
        Phe.txt
        PheAll.txt
        Regression.G5.G.J.txt
        Regression.G5.G.JC.txt
        Regression.G5.G.PBLUP.txt
        all.ID
        alphaEstimates
        genotype.ID
        meanOfSNPGAll
        meanOfSNPGG0
        meanOfSNPGG1
        meanOfSNPGG2
        meanOfSNPGG3
        meanOfSNPGG4
        meanOfSNPGG5
        noGenotype.ID
        sim.bv
        sim.phenotype
```

```
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
In [29]: ;wc G0.noGenotype.ID;wc G1.noGenotype.ID;wc G2.noGenotype.ID;wc G3.noGenotype.ID;wc G4.noGenotype.ID;wc G
          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
                = 0.710
                = 0.710
         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
         7062.691952 seconds (23.88 G allocations: 737.464 GB, 2.80% gc time)
In [32]: betaHat
Out[32]: 1-element Array{Float64,1}:
          33.9502
In [33]: using DataFrames
         df = readtable("sim.bv", eltypes =[UTF8String, Float64], separator = ' ',header=false)
In [34]:
         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 = 0.919
         SSBRJC from Gibbs - all.ID : regression of TBV on GEBV = 1.002
Out[35]: 0.9187566460909963
In [36]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[36]: 0.15927333663736037
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 = 0.925
         SSBRJC from Gibbs - genotype.ID : regression of TBV on GEBV = 1.135
Out[37]: 0.9252488664703576
In [38]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[38]: 1.5033196035095704
```

```
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 = 0.883
         SSBRJC from Gibbs - noGenotype.ID : regression of TBV on GEBV = 0.977
Out[39]: 0.8832195673923228
In [40]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[40]: -0.15089118648699565
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 = 0.708
         SSBRJC from Gibbs - G0.ID: regression of TBV on GEBV = 0.868
Out[41]: 0.7076726253075455
In [42]: GEBV = aHat1[posAi]
         G1GEBV=mean(GEBV)
Out[42]: -1.3528495178881854
```

```
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 = 0.782
         SSBRJC from Gibbs - G1.ID: regression of TBV on GEBV = 1.002
Out[43]: 0.7815182308727646
In [44]: GEBV = aHat1[posAi]
         G1GEBV=mean(GEBV)
Out[44]: -0.6773662549866445
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 = 0.776
         SSBRJC from Gibbs - G2.ID : regression of TBV on GEBV = 1.025
Out[45]: 0.776022323330401
In [46]: GEBV = aHat1[posAi]
         G2GEBV=mean(GEBV)
Out[46]: -0.07815337841098088
```

```
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 = 0.773
         SSBRJC from Gibbs - G3.ID : regression of TBV on GEBV = 1.057
Out[47]: 0.7730941782767329
In [48]: GEBV = aHat1[posAi]
         G3GEBV=mean(GEBV)
Out[48]: 0.4784388531861695
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 = 0.779
         SSBRJC from Gibbs - G4.ID : regression of TBV on GEBV = 1.054
Out[49]: 0.7792434322643231
In [50]: GEBV = aHat1[posAi]
         G4GEBV=mean(GEBV)
Out[50]: 1.0257865255212935
```

```
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 = 0.911
         SSBRJC from Gibbs - G5.ID: regression of TBV on GEBV = 1.131
Out[51]: 0.9114168971364179
In [52]: GEBV = aHat1[posAi]
         G5GEBV=mean(GEBV)
Out[52]: 1.5597837924025098
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 = 0.939
         SSBRJC from Gibbs - G0.Genotype.ID : regression of TBV on GEBV = 1.117
Out[53]: 0.939403499178937
In [54]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[54]: 0.059223925449283545
```

```
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 = 0.938
         SSBRJC from Gibbs - G1.Genotype.ID : regression of TBV on GEBV = 1.091
Out[55]: 0.9384132232096764
In [56]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[56]: 0.5780450985087076
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 = 0.933
         SSBRJC from Gibbs - G2.Genotype.ID : regression of TBV on GEBV = 1.056
Out[57]: 0.9325254707994003
In [58]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[58]: 1.0601320056953711
```

```
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 = 0.937
         SSBRJC from Gibbs - G3.Genotype.ID : regression of TBV on GEBV = 1.051
Out[59]: 0.9366707493342554
In [60]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[60]: 1.5429379339352443
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 = 0.862
         SSBRJC from Gibbs - G4.Genotype.ID : regression of TBV on GEBV = 1.055
Out[61]: 0.8616145815605176
In [62]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[62]: 2.0176914982416636
```

```
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 = 0.911
         SSBRJC from Gibbs - G5.Genotype.ID : regression of TBV on GEBV = 1.131
Out[63]: 0.9114168971364179
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]: 1.5597837924025098
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 = 0.688
         SSBRJC from Gibbs - G0.noGenotype.ID: regression of TBV on GEBV = 0.876
Out[67]: 0.6881796749627768
In [68]: GEBV = aHat1[posAi]
         mean(GEBV)
Out[68]: -1.3890565292558126
```

```
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 = 0.763
         SSBRJC from Gibbs - G1.noGenotype.ID : regression of TBV on GEBV = 1.004
Out[69]: 0.7630601559970338
In [70]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[70]: -0.7095562896916534
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 = 0.759
         SSBRJC from Gibbs - G2.noGenotype.ID: regression of TBV on GEBV = 1.025
Out[71]: 0.7585230978321187
In [72]: | GEBV = aHat1[posAi]
         mean (GEBV)
Out[72]: -0.10734018313165658
```

```
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 = 0.755
         SSBRJC from Gibbs - G3.noGenotype.ID : regression of TBV on GEBV = 1.055
Out[73]: 0.7551229142235282
In [74]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[74]: 0.4511440049618343
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 = 0.764
         SSBRJC from Gibbs - G4.noGenotype.ID: regression of TBV on GEBV = 1.049
Out[75]: 0.764345930549447
In [76]: GEBV = aHat1[posAi]
         mean (GEBV)
Out[76]: 1.0003530646823096
In [77]: numSSBayes
Out[77]: SSBR.NumSSBayes(54946,45946,9000,40000,39000,1000,2000)
```