MiX99

Solving Large Mixed Model Equations

Genomic Models and MiX99

Esa Mäntysaari, MTT

Genomic modeling

Commonly Genomic Modeling by two alternative paths:

Multistep - evaluations

Three steps:

1. Genetic evaluations to get **EBV** or Daughter yield deviations (DYD)
   (or EBV are converted to deregressed proof **DRP**)
2. DYD or DRP are used to solve direct genomic values **DGV**
3. DGVs are combined with EBV to form Genomic Enhanced Breeding values **GEBV**

Single step evaluations

Evaluations include directly all animals:

  - Animals with genomic information
  - Animals without genomic information
- Genomic information is used:
  - directly in the model for phenotypic observations

  Or EBV of animals with records are converted to DRP and DRP are used as observations

- Yields directly the **GEBV**

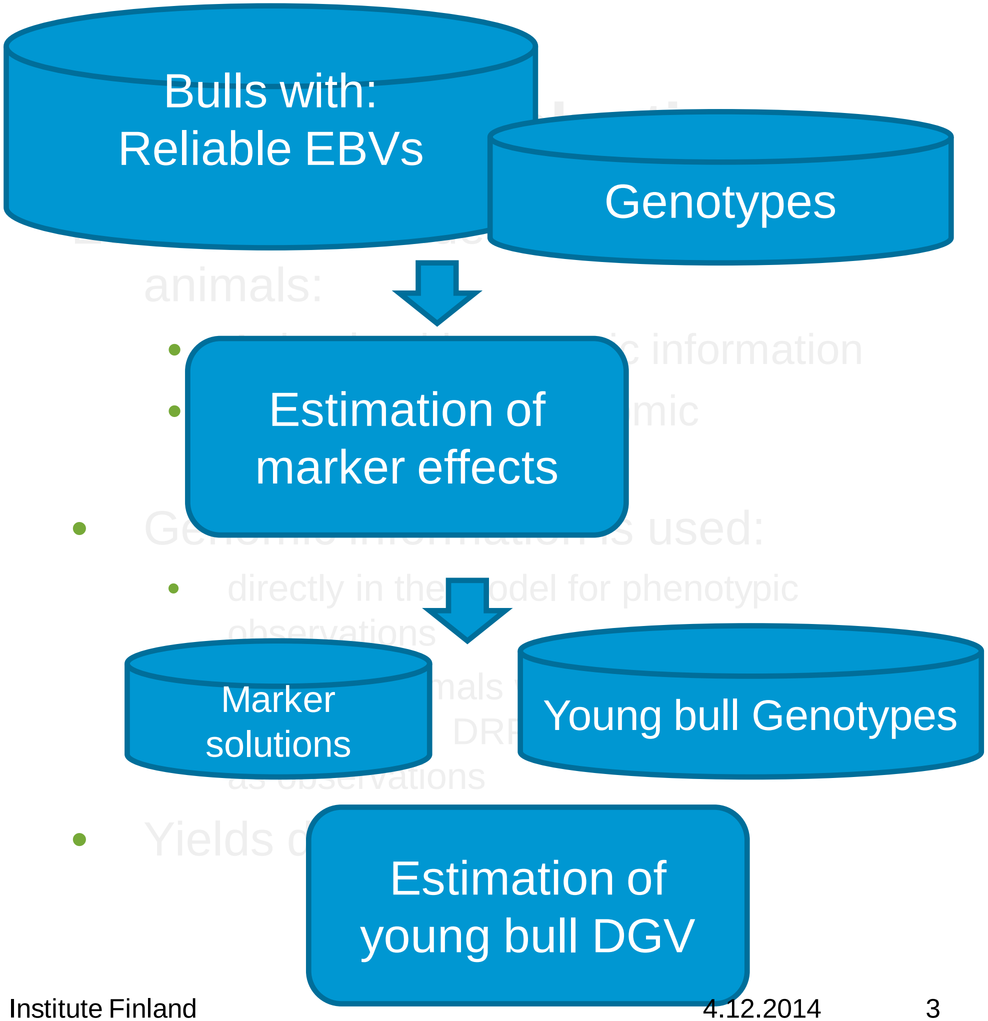
Genomic modeling

Commonly Genomic Modeling takes two alternative paths:

Multistep - evaluations

Three steps:

1. Genetic evaluations to get **EBV** or Daughter yield deviations (DYD)
   (or EBV are converted to deregressed proof **DRP**)
2. DYD or DRP are used to solve direct genomic values **DGV**
3. DGVs are combined with EBV to form Genomic Enhanced Breeding values **GEBV**



Bulls with:
Reliable EBVs

Genotypes

Estimation of marker effects

Marker solutions

Young bull Genotypes

Estimation of young bull DGV

animals:
- c information
- mic

- Genomic information is used:
  - directly in the odel for phenotypic observations
  - mals DRP as observations
- Yields c

Genomic modeling

Commonly Genomic Modeling by two alternative paths:

Multistep - evaluations

Three steps:

1. Genetic evaluations to get **EBV** or Daughter yield deviations (DYD)
   (or EBV are converted to deregressed proof **DRP**)
2. DYD or DRP are used to solve direct genomic values **DGV**
3. DGVs are combined with EBV to form Genomic Enhanced Breeding values **GEBV**

Single step evaluations

Evaluations include directly all animals:

  - Animals with genomic information
  - Animals without genomic information
- Genomic information is used:
  - directly in the model for phenotypic observations

  Or EBV of animals with records are converted to DRP and DRP are used as observations

- Yields directly the **GEBV**

Genomic modeling

- Frequently genomic BV models (multi-step) are simple
  - ”Precorrected” data is used
    DRPs (Deregressed genetic evaluations)
    DYDs (Daughter yield deviations – or yield deviations YD)
    EDCs or **ERC**s (Effective Daughter/Record Contributions) are used as weights
- Not large number of unknowns to solve
  - Dense nature of MME and bad model choice can lead into problems…

- Why use MiX99?
  - MiX99 treats genomic models same way as usual models:
    - can include fixed effects,
    - can include polygenic effects, or other random effects
    - can utilize multiple traits, when needed
  - MiX99 pipeline is less error prone than more diverse –tailored- applications
    - No extra renumbering or extra data edits
    - Matching utlilities: hginv, exa99, ApaX

Genomic Models available in MiX99

1. GBLUP Models
    - Generally: models that require specified variance structures e.g. marker assisted models and corresponding IBD matrices
    - Genomic relationships, RKHS* (Euclidean distance matrices)
    - Trait wise Genomic Relationship matrices

2. Marker effect models
    - SNP-BLUP with common variance
    - SNP-BLUP with heterogenous variances for SNP-effects
    - Haplotype –effect models
    - with or without other random effects (polygenic etc.)

3. Single-step models

*Reproducing kernel Hilbert spaces regression

Genomic model approach in MiX99

1. GBLUP Models
   - Generally:
     Solver does not know the structure of the matrix
   - User supplies the inverses of the variance structures
   -

2. Marker effect models
   - Generally:

     User supplies the genomic data design matrices – **regression matrix file**
   - Simplifies the use (less model instructions are needed)
   - Gives great freedom to specify the design matrix

3. Single-step models

From freedom came elegance

From freedom came responsibility !

Walk through example

SNP-BLUP

Model

Simple genomic marker effect linear model is

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

where

$\mathbf{y}$ is $n \times 1$ vector of observations,

$\mathbf{1}_n$ is $n \times 1$ vector of ones,

μ is the unknown general mean,

$\mathbf{Z}$ is a $n \times m$ matrix containing a column for each marker locus,

$\mathbf{g}$ is a $m \times 1$ vector of random SNP marker effects, and

$\mathbf{e}$ is a random residual vector.

Prerequisites for genomic evaluations

In the genomic model we need:

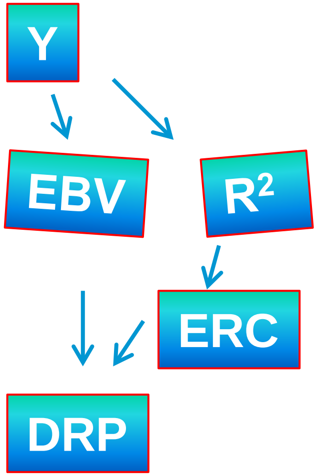
- High quality genotypes

  Typical step before genomic evaluation is extensive editing of genotypes: MAF, call rates, clones, high linkage
  - MiX99 uses data as -- animals in rows, SNPs in columns –

- Phenotype data
  - MiX99 data form as usual:
  - Columns:

    Fixed effects [random effects] observations [weights]

- Model parameters
- (co)Variances for markers – (co)variances for residuals



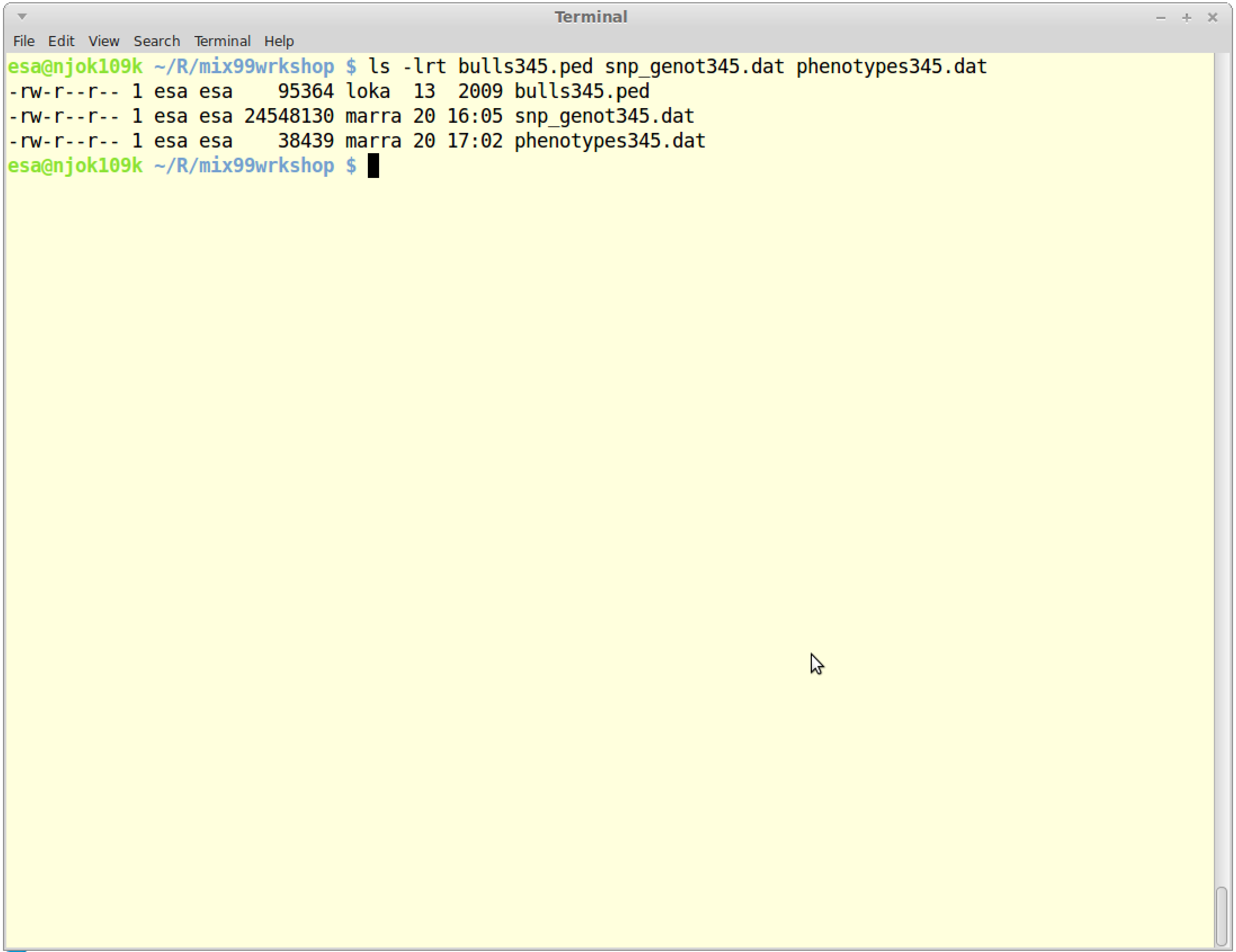
Y

EBV

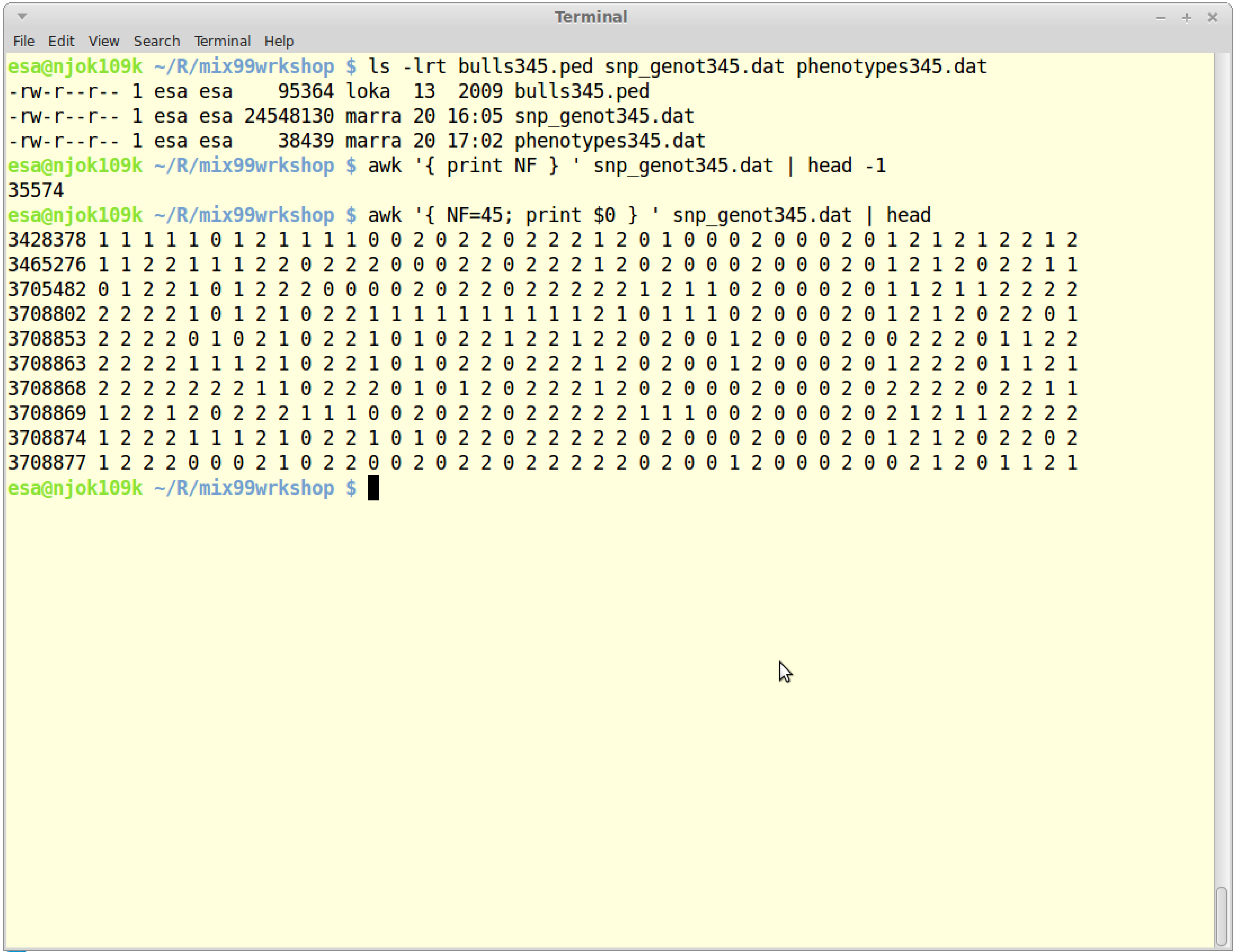
R^2

ERC

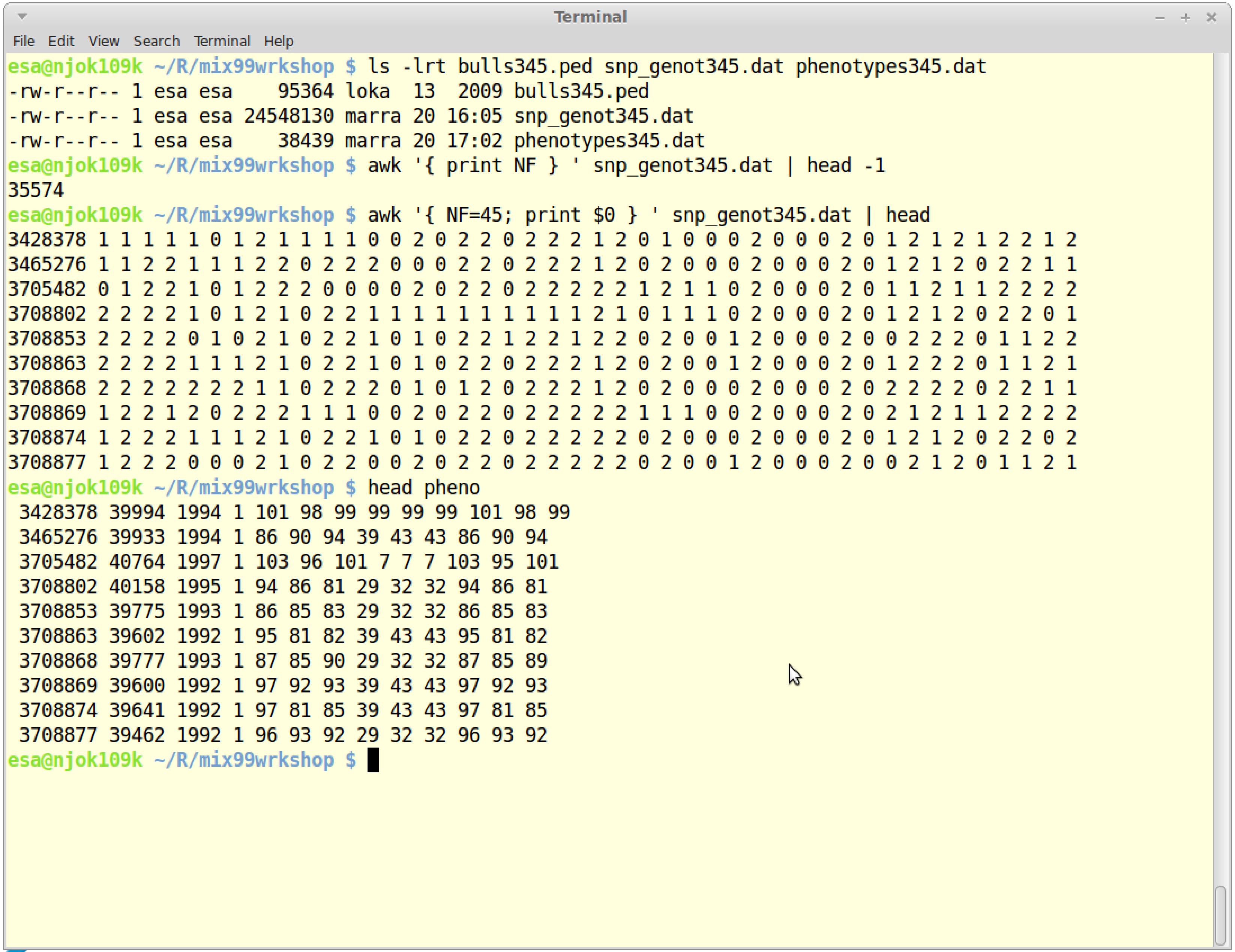
DRP



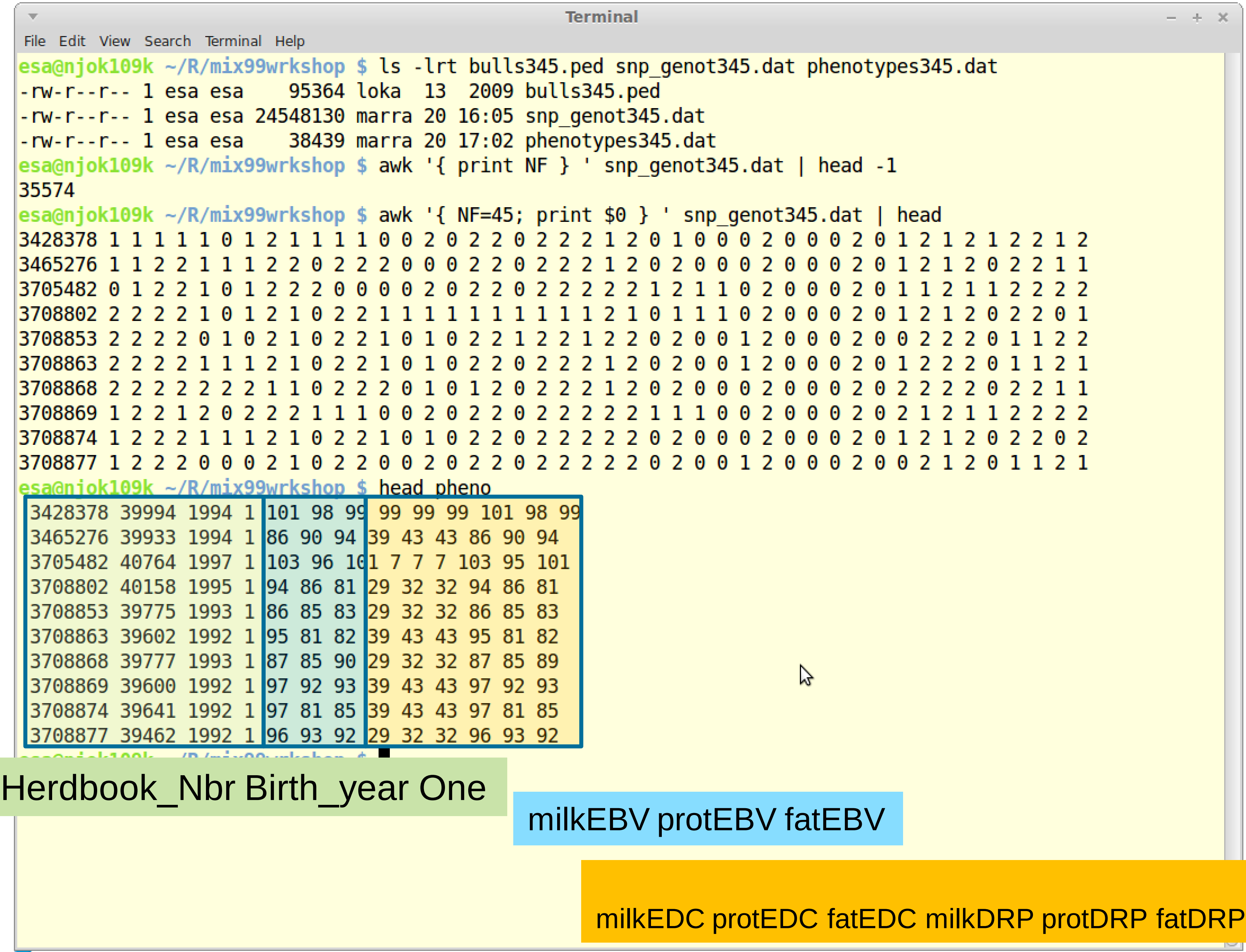
```
esa@njok109k ~/R/mix99wrkshop $ ls -lrt bulls345.ped snp_genot345.dat phenotypes345.dat
-rw-r--r-- 1 esa esa    95364 loka  13  2009 bulls345.ped
-rw-r--r-- 1 esa esa 24548130 marra 20 16:05 snp_genot345.dat
-rw-r--r-- 1 esa esa    38439 marra 20 17:02 phenotypes345.dat
esa@njok109k ~/R/mix99wrkshop $ 
```



```
esa@njok109k ~/R/mix99wrkshop $ ls -lrt bulls345.ped snp_genot345.dat phenotypes345.dat
-rw-r--r-- 1 esa esa    95364 loka  13  2009 bulls345.ped
-rw-r--r-- 1 esa esa 24548130 marra 20 16:05 snp_genot345.dat
-rw-r--r-- 1 esa esa    38439 marra 20 17:02 phenotypes345.dat
esa@njok109k ~/R/mix99wrkshop $ awk '{ print NF } ' snp_genot345.dat | head -1
35574
esa@njok109k ~/R/mix99wrkshop $ awk '{ NF=45; print $0 } ' snp_genot345.dat | head
3428378 1 1 1 1 1 0 1 2 1 1 1 1 0 0 2 0 2 2 0 2 2 2 1 2 0 1 0 0 0 2 0 0 0 2 0 1 2 1 2 1 2 2 1 2
3465276 1 1 2 2 1 1 1 2 2 0 2 2 2 0 0 0 2 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 1 1
3705482 0 1 2 2 1 0 1 2 2 2 0 0 0 0 2 0 2 2 0 2 2 2 2 2 1 2 1 1 0 2 0 0 0 2 0 1 1 2 1 1 2 2 2 2
3708802 2 2 2 2 1 0 1 2 1 0 2 2 1 1 1 1 1 1 1 1 1 1 2 1 0 1 1 1 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 1
3708853 2 2 2 2 0 1 0 2 1 0 2 2 1 0 1 0 2 2 1 2 2 1 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 2 2 0 1 1 2 2
3708863 2 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 1 2 0 2 0 0 1 2 0 0 0 2 0 1 2 2 2 0 1 1 2 1
3708868 2 2 2 2 2 2 2 1 1 0 2 2 2 0 1 0 1 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 2 2 2 2 0 2 2 1 1
3708869 1 2 2 1 2 0 2 2 2 1 1 1 0 0 2 0 2 2 0 2 2 2 2 2 1 1 1 0 0 2 0 0 0 2 0 2 1 2 1 1 2 2 2 2
3708874 1 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 2 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 2
3708877 1 2 2 2 0 0 0 2 1 0 2 2 0 0 2 0 2 2 0 2 2 2 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 1 2 0 1 1 2 1
esa@njok109k ~/R/mix99wrkshop $ 
```



```
esa@njok109k ~/R/mix99wrkshop $ ls -lrt bulls345.ped snp_genot345.dat phenotypes345.dat
-rw-r--r-- 1 esa esa    95364 loka  13  2009 bulls345.ped
-rw-r--r-- 1 esa esa 24548130 marra 20 16:05 snp_genot345.dat
-rw-r--r-- 1 esa esa    38439 marra 20 17:02 phenotypes345.dat
esa@njok109k ~/R/mix99wrkshop $ awk '{ print NF } ' snp_genot345.dat | head -1
35574
esa@njok109k ~/R/mix99wrkshop $ awk '{ NF=45; print $0 } ' snp_genot345.dat | head
3428378 1 1 1 1 1 0 1 2 1 1 1 1 0 0 2 0 2 2 0 2 2 2 1 2 0 1 0 0 0 2 0 0 0 2 0 1 2 1 2 1 2 2 1 2
3465276 1 1 2 2 1 1 1 2 2 0 2 2 2 0 0 0 2 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 1 1
3705482 0 1 2 2 1 0 1 2 2 2 0 0 0 0 2 0 2 2 0 2 2 2 2 2 1 2 1 1 0 2 0 0 0 2 0 1 1 2 1 1 2 2 2 2
3708802 2 2 2 2 1 0 1 2 1 0 2 2 1 1 1 1 1 1 1 1 1 1 2 1 0 1 1 1 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 1
3708853 2 2 2 2 0 1 0 2 1 0 2 2 1 0 1 0 2 2 1 2 2 1 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 2 2 0 1 1 2 2
3708863 2 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 1 2 0 2 0 0 1 2 0 0 0 2 0 1 2 2 2 0 1 1 2 1
3708868 2 2 2 2 2 2 2 1 1 0 2 2 2 0 1 0 1 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 2 2 2 2 0 2 2 1 1
3708869 1 2 2 1 2 0 2 2 2 1 1 1 0 0 2 0 2 2 0 2 2 2 2 2 1 1 1 0 0 2 0 0 0 2 0 2 1 2 1 1 2 2 2 2
3708874 1 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 2 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 2
3708877 1 2 2 2 0 0 0 2 1 0 2 2 0 0 2 0 2 2 0 2 2 2 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 1 2 0 1 1 2 1
esa@njok109k ~/R/mix99wrkshop $ head pheno
 3428378 39994 1994 1 101 98 99 99 99 99 101 98 99
 3465276 39933 1994 1 86 90 94 39 43 43 86 90 94
 3705482 40764 1997 1 103 96 101 7 7 7 103 95 101
 3708802 40158 1995 1 94 86 81 29 32 32 94 86 81
 3708853 39775 1993 1 86 85 83 29 32 32 86 85 83
 3708863 39602 1992 1 95 81 82 39 43 43 95 81 82
 3708868 39777 1993 1 87 85 90 29 32 32 87 85 89
 3708869 39600 1992 1 97 92 93 39 43 43 97 92 93
 3708874 39641 1992 1 97 81 85 39 43 43 97 81 85
 3708877 39462 1992 1 96 93 92 29 32 32 96 93 92
esa@njok109k ~/R/mix99wrkshop $ 
```



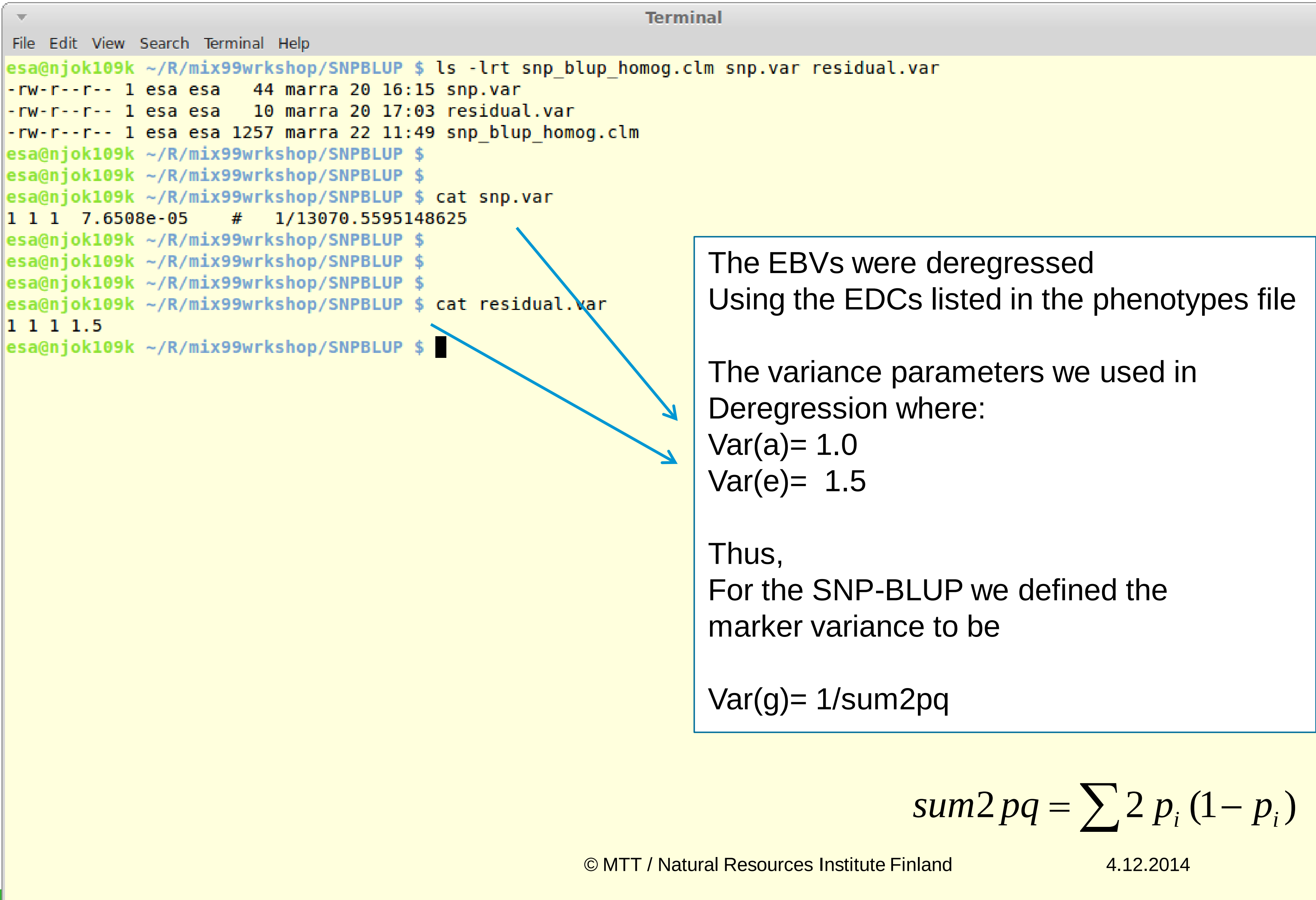
```
esa@njok109k ~/R/mix99wrkshop $ ls -lrt bulls345.ped snp_genot345.dat phenotypes345.dat
-rw-r--r-- 1 esa esa    95364 loka  13  2009 bulls345.ped
-rw-r--r-- 1 esa esa 24548130 marra 20 16:05 snp_genot345.dat
-rw-r--r-- 1 esa esa    38439 marra 20 17:02 phenotypes345.dat
esa@njok109k ~/R/mix99wrkshop $ awk '{ print NF } ' snp_genot345.dat | head -1
35574
esa@njok109k ~/R/mix99wrkshop $ awk '{ NF=45; print $0 } ' snp_genot345.dat | head
3428378 1 1 1 1 1 0 1 2 1 1 1 1 0 0 2 0 2 2 0 2 2 2 1 2 0 1 0 0 0 2 0 0 0 2 0 1 2 1 2 1 2 2 1 2
3465276 1 1 2 2 1 1 1 2 2 0 2 2 2 0 0 0 2 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 1 1
3705482 0 1 2 2 1 0 1 2 2 2 0 0 0 0 2 0 2 2 0 2 2 2 2 2 1 2 1 1 0 2 0 0 0 2 0 1 1 2 1 1 2 2 2 2
3708802 2 2 2 2 1 0 1 2 1 0 2 2 1 1 1 1 1 1 1 1 1 1 2 1 0 1 1 1 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 1
3708853 2 2 2 2 0 1 0 2 1 0 2 2 1 0 1 0 2 2 1 2 2 1 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 2 2 0 1 1 2 2
3708863 2 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 1 2 0 2 0 0 1 2 0 0 0 2 0 1 2 2 2 0 1 1 2 1
3708868 2 2 2 2 2 2 2 1 1 0 2 2 2 0 1 0 1 2 0 2 2 2 1 2 0 2 0 0 0 2 0 0 0 2 0 2 2 2 2 0 2 2 1 1
3708869 1 2 2 1 2 0 2 2 2 1 1 1 0 0 2 0 2 2 0 2 2 2 2 2 1 1 1 0 0 2 0 0 0 2 0 2 1 2 1 1 2 2 2 2
3708874 1 2 2 2 1 1 1 2 1 0 2 2 1 0 1 0 2 2 0 2 2 2 2 2 0 2 0 0 0 2 0 0 0 2 0 1 2 1 2 0 2 2 0 2
3708877 1 2 2 2 0 0 0 2 1 0 2 2 0 0 2 0 2 2 0 2 2 2 2 2 0 2 0 0 1 2 0 0 0 2 0 0 2 1 2 0 1 1 2 1
esa@njok109k ~/R/mix99wrkshop $ head pheno
3428378 39994 1994 1 101 98 99 99 99 99 101 98 99
3465276 39933 1994 1 86 90 94 39 43 43 86 90 94
3705482 40764 1997 1 103 96 101 7 7 7 103 95 101
3708802 40158 1995 1 94 86 81 29 32 32 94 86 81
3708853 39775 1993 1 86 85 83 29 32 32 86 85 83
3708863 39602 1992 1 95 81 82 39 43 43 95 81 82
3708868 39777 1993 1 87 85 90 29 32 32 87 85 89
3708869 39600 1992 1 97 92 93 39 43 43 97 92 93
3708874 39641 1992 1 97 81 85 39 43 43 97 81 85
3708877 39462 1992 1 96 93 92 29 32 32 96 93 92
esa@njok109k ~/R/mix99wrkshop $
```

ID Herdbook_Nbr Birth_year One

milkEBV protEBV fatEBV

milkEDC protEDC fatEDC milkDRP protDRP fatDRP

MiX99 Instructions and parameter files:



```
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ ls -lrt snp_blup_homog.clm snp.var residual.var
-rw-r--r-- 1 esa esa   44 marra 20 16:15 snp.var
-rw-r--r-- 1 esa esa   10 marra 20 17:03 residual.var
-rw-r--r-- 1 esa esa 1257 marra 22 11:49 snp_blup_homog.clm
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ cat snp.var
1 1 1  7.6508e-05    #   1/13070.5595148625
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ cat residual.var
1 1 1 1.5
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ 
```

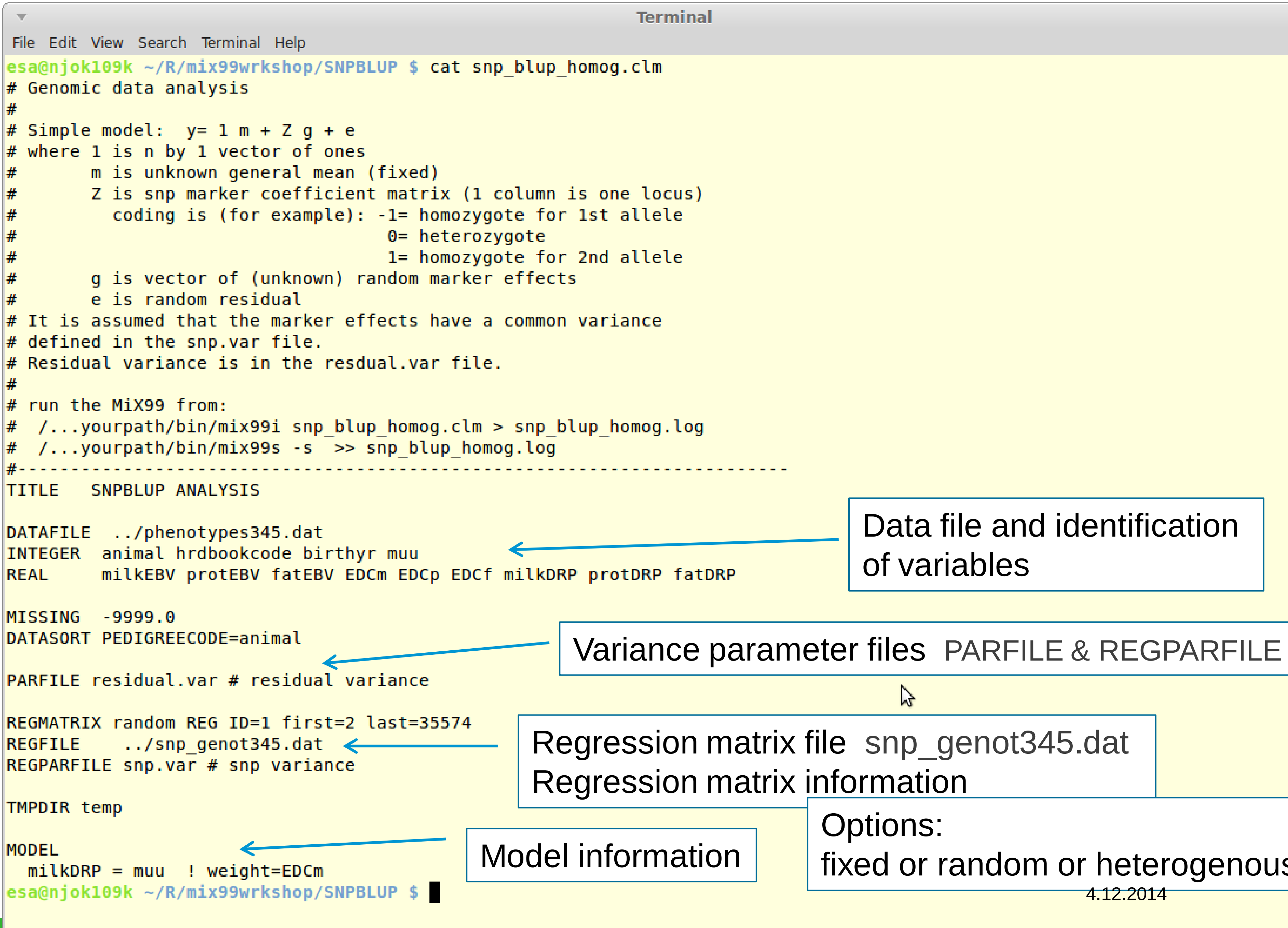
The EBVs were deregressed
Using the EDCs listed in the phenotypes file

The variance parameters we used in
Deregression where:
Var(a)= 1.0
Var(e)= 1.5

Thus,
For the SNP-BLUP we defined the
marker variance to be

Var(g)= 1/sum2pq

$$sum2pq = \sum 2\, p_i\,(1 - p_i)$$



```
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ cat snp_blup_homog.clm
# Genomic data analysis
#
# Simple model:  y= 1 m + Z g + e
# where 1 is n by 1 vector of ones
#       m is unknown general mean (fixed)
#       Z is snp marker coefficient matrix (1 column is one locus)
#         coding is (for example): -1= homozygote for 1st allele
#                                   0= heterozygote
#                                   1= homozygote for 2nd allele
#       g is vector of (unknown) random marker effects
#       e is random residual
# It is assumed that the marker effects have a common variance
# defined in the snp.var file.
# Residual variance is in the resdual.var file.
#
# run the MiX99 from:
#  /...yourpath/bin/mix99i snp_blup_homog.clm > snp_blup_homog.log
#  /...yourpath/bin/mix99s -s  >> snp_blup_homog.log
#-------------------------------------------------------------------------
TITLE   SNPBLUP ANALYSIS

DATAFILE  ../phenotypes345.dat
INTEGER  animal hrdbookcode birthyr muu
REAL     milkEBV protEBV fatEBV EDCm EDCp EDCf milkDRP protDRP fatDRP

MISSING  -9999.0
DATASORT PEDIGREECODE=animal

PARFILE residual.var # residual variance

REGMATRIX random REG ID=1 first=2 last=35574
REGFILE    ../snp_genot345.dat
REGPARFILE snp.var # snp variance

TMPDIR temp

MODEL
  milkDRP = muu  ! weight=EDCm
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
```

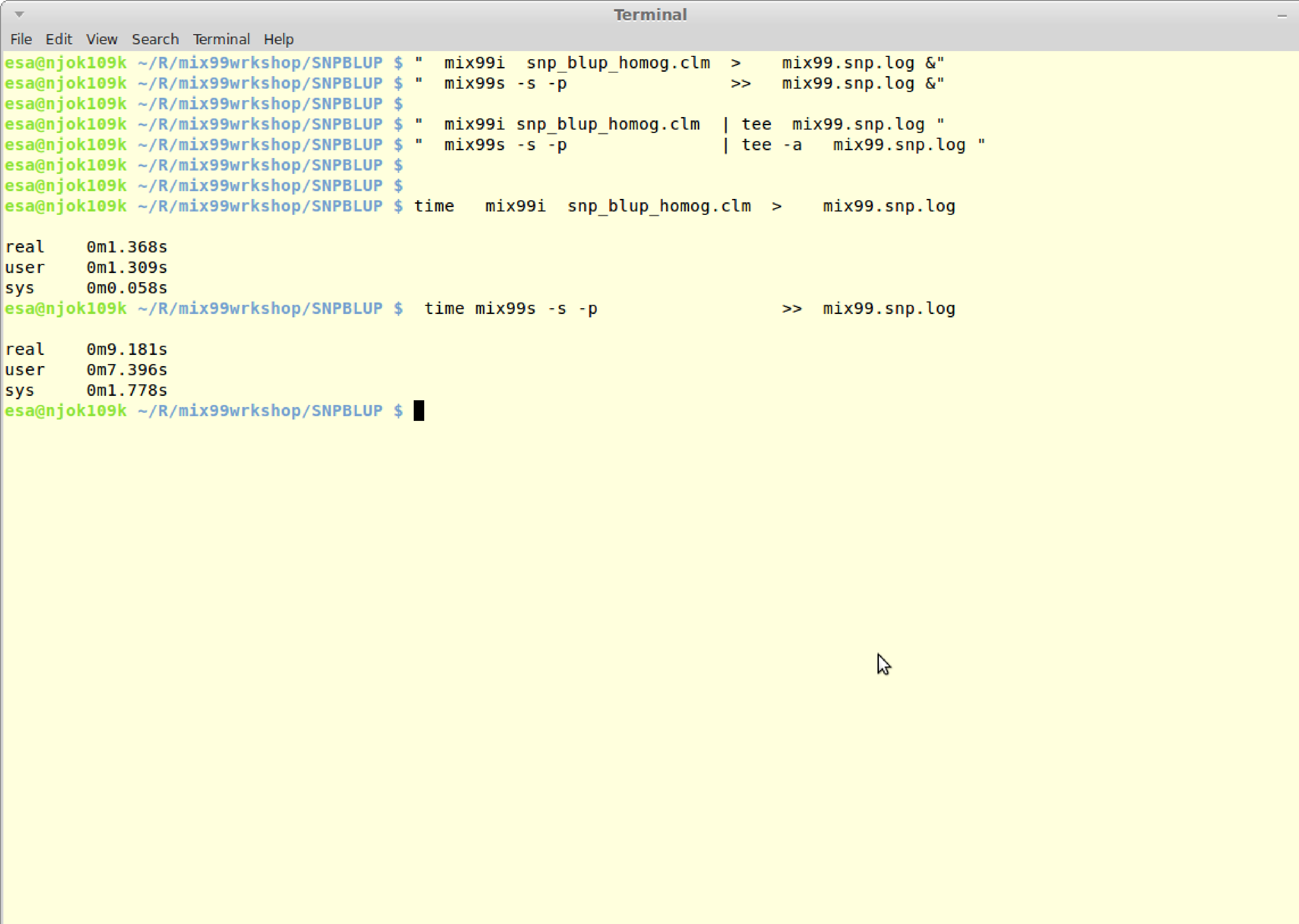
Data file and identification of variables

Variance parameter files PARFILE & REGPARFILE

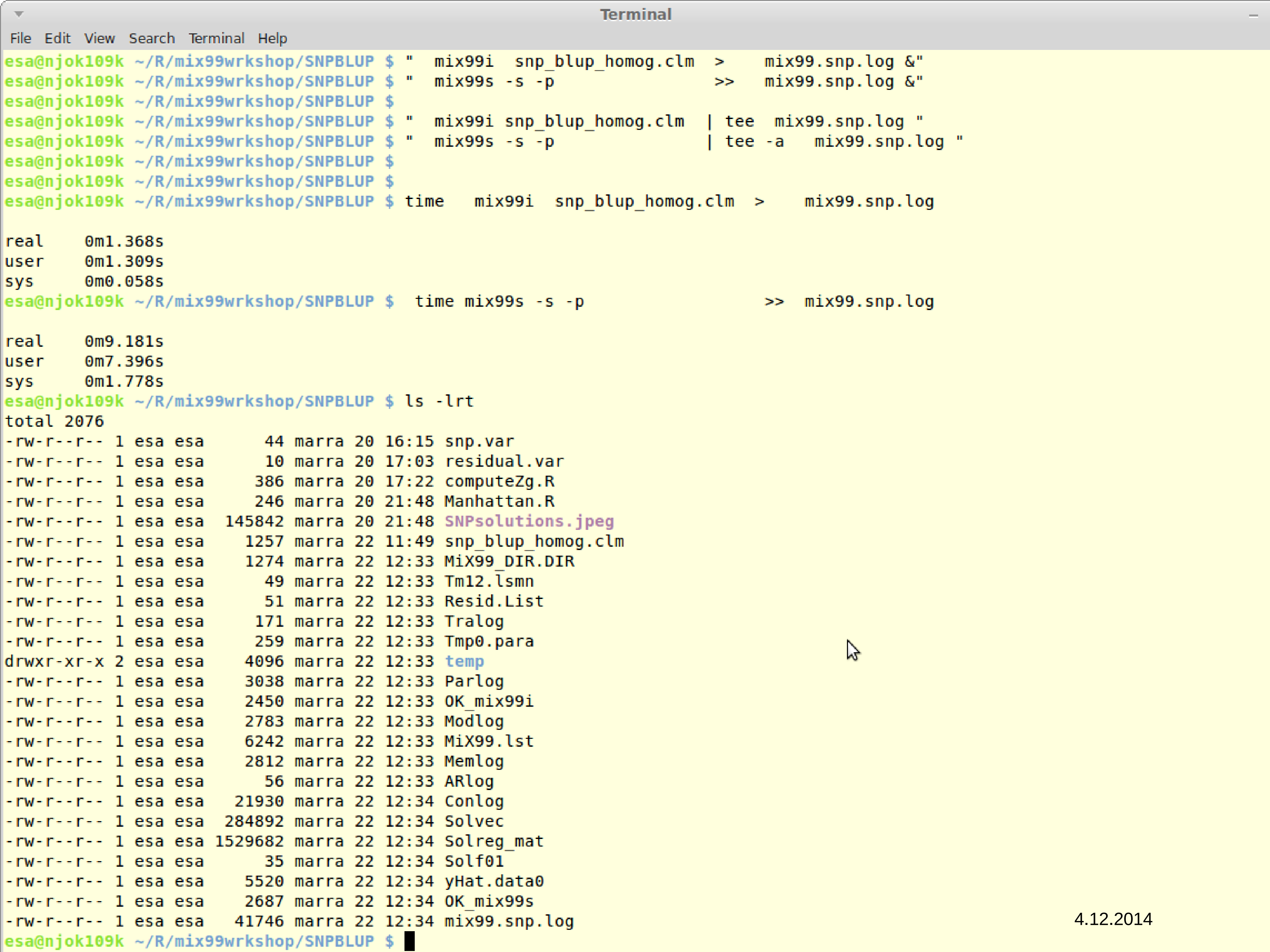
Regression matrix file snp_genot345.dat
Regression matrix information

Model information

Options:
fixed or random or heterogenous

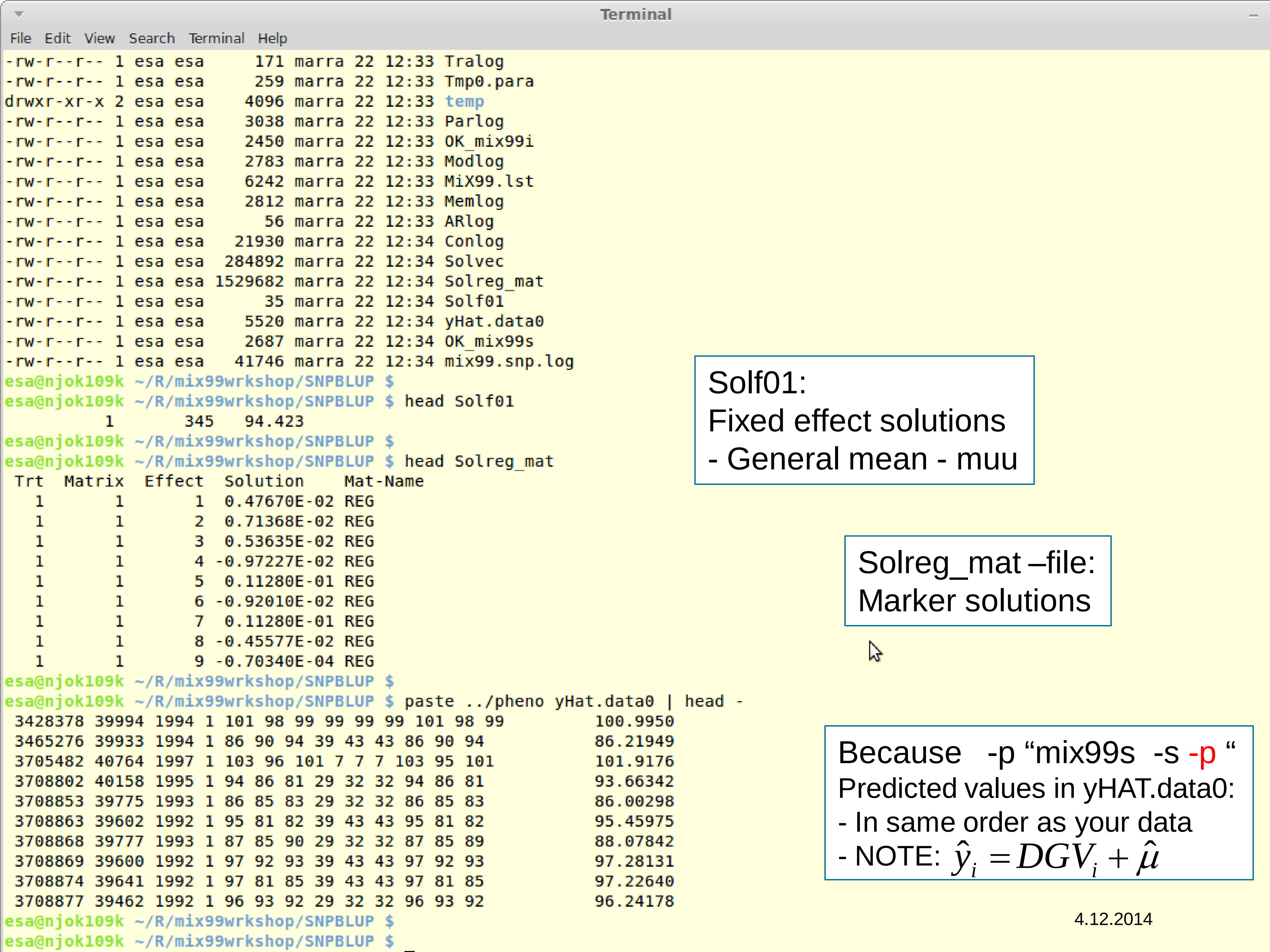


```
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99i  snp_blup_homog.clm  >    mix99.snp.log &"
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99s -s -p                >>   mix99.snp.log &"
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99i snp_blup_homog.clm  | tee  mix99.snp.log "
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99s -s -p               | tee -a   mix99.snp.log "
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ time   mix99i  snp_blup_homog.clm  >    mix99.snp.log

real    0m1.368s
user    0m1.309s
sys     0m0.058s
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $  time mix99s -s -p                  >>  mix99.snp.log

real    0m9.181s
user    0m7.396s
sys     0m1.778s
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ 
```



```
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99i  snp_blup_homog.clm  >    mix99.snp.log &"
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99s -s -p                >>   mix99.snp.log &"
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99i snp_blup_homog.clm  | tee  mix99.snp.log "
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ "  mix99s -s -p               | tee -a   mix99.snp.log "
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ time   mix99i  snp_blup_homog.clm  >    mix99.snp.log

real    0m1.368s
user    0m1.309s
sys     0m0.058s
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $  time mix99s -s -p                  >>  mix99.snp.log

real    0m9.181s
user    0m7.396s
sys     0m1.778s
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ ls -lrt
total 2076
-rw-r--r-- 1 esa esa      44 marra 20 16:15 snp.var
-rw-r--r-- 1 esa esa      10 marra 20 17:03 residual.var
-rw-r--r-- 1 esa esa     386 marra 20 17:22 computeZg.R
-rw-r--r-- 1 esa esa     246 marra 20 21:48 Manhattan.R
-rw-r--r-- 1 esa esa  145842 marra 20 21:48 SNPsolutions.jpeg
-rw-r--r-- 1 esa esa    1257 marra 22 11:49 snp_blup_homog.clm
-rw-r--r-- 1 esa esa    1274 marra 22 12:33 MiX99_DIR.DIR
-rw-r--r-- 1 esa esa      49 marra 22 12:33 Tm12.lsmn
-rw-r--r-- 1 esa esa      51 marra 22 12:33 Resid.List
-rw-r--r-- 1 esa esa     171 marra 22 12:33 Tralog
-rw-r--r-- 1 esa esa     259 marra 22 12:33 Tmp0.para
drwxr-xr-x 2 esa esa    4096 marra 22 12:33 temp
-rw-r--r-- 1 esa esa    3038 marra 22 12:33 Parlog
-rw-r--r-- 1 esa esa    2450 marra 22 12:33 OK_mix99i
-rw-r--r-- 1 esa esa    2783 marra 22 12:33 Modlog
-rw-r--r-- 1 esa esa    6242 marra 22 12:33 MiX99.lst
-rw-r--r-- 1 esa esa    2812 marra 22 12:33 Memlog
-rw-r--r-- 1 esa esa      56 marra 22 12:33 ARlog
-rw-r--r-- 1 esa esa   21930 marra 22 12:34 Conlog
-rw-r--r-- 1 esa esa  284892 marra 22 12:34 Solvec
-rw-r--r-- 1 esa esa 1529682 marra 22 12:34 Solreg_mat
-rw-r--r-- 1 esa esa      35 marra 22 12:34 Solf01
-rw-r--r-- 1 esa esa    5520 marra 22 12:34 yHat.data0
-rw-r--r-- 1 esa esa    2687 marra 22 12:34 OK_mix99s
-rw-r--r-- 1 esa esa   41746 marra 22 12:34 mix99.snp.log
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
```

4.12.2014



```
-rw-r--r-- 1 esa esa     171 marra 22 12:33 Tralog
-rw-r--r-- 1 esa esa     259 marra 22 12:33 Tmp0.para
drwxr-xr-x 2 esa esa    4096 marra 22 12:33 temp
-rw-r--r-- 1 esa esa    3038 marra 22 12:33 Parlog
-rw-r--r-- 1 esa esa    2450 marra 22 12:33 OK_mix99i
-rw-r--r-- 1 esa esa    2783 marra 22 12:33 Modlog
-rw-r--r-- 1 esa esa    6242 marra 22 12:33 MiX99.lst
-rw-r--r-- 1 esa esa    2812 marra 22 12:33 Memlog
-rw-r--r-- 1 esa esa      56 marra 22 12:33 ARlog
-rw-r--r-- 1 esa esa   21930 marra 22 12:34 Conlog
-rw-r--r-- 1 esa esa  284892 marra 22 12:34 Solvec
-rw-r--r-- 1 esa esa 1529682 marra 22 12:34 Solreg_mat
-rw-r--r-- 1 esa esa      35 marra 22 12:34 Solf01
-rw-r--r-- 1 esa esa    5520 marra 22 12:34 yHat.data0
-rw-r--r-- 1 esa esa    2687 marra 22 12:34 OK_mix99s
-rw-r--r-- 1 esa esa   41746 marra 22 12:34 mix99.snp.log
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ head Solf01
          1       345    94.423
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ head Solreg_mat
 Trt   Matrix   Effect   Solution     Mat-Name
   1        1        1   0.47670E-02 REG
   1        1        2   0.71368E-02 REG
   1        1        3   0.53635E-02 REG
   1        1        4  -0.97227E-02 REG
   1        1        5   0.11280E-01 REG
   1        1        6  -0.92010E-02 REG
   1        1        7   0.11280E-01 REG
   1        1        8  -0.45577E-02 REG
   1        1        9  -0.70340E-04 REG
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ paste ../pheno yHat.data0 | head -
 3428378 39994 1994 1 101 98 99 99 99 99 101 98 99          100.9950
 3465276 39933 1994 1 86 90 94 39 43 43 86 90 94            86.21949
 3705482 40764 1997 1 103 96 101 7 7 7 103 95 101           101.9176
 3708802 40158 1995 1 94 86 81 29 32 32 94 86 81            93.66342
 3708853 39775 1993 1 86 85 83 29 32 32 86 85 83            86.00298
 3708863 39602 1992 1 95 81 82 39 43 43 95 81 82            95.45975
 3708868 39777 1993 1 87 85 90 29 32 32 87 85 89            88.07842
 3708869 39600 1992 1 97 92 93 39 43 43 97 92 93            97.28131
 3708874 39641 1992 1 97 81 85 39 43 43 97 81 85            97.22640
 3708877 39462 1992 1 96 93 92 29 32 32 96 93 92            96.24178
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ _
```

Solf01:
Fixed effect solutions
- General mean - muu

Solreg_mat –file:
Marker solutions

Because -p “mix99s -s -p “
Predicted values in yHAT.data0:
- In same order as your data
- NOTE: $\hat{y}_i = DGV_i + \hat{\mu}$

4.12.2014

More advanced SNP-BLUP …

- Add polygenic effect to the model

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Wa} + \mathbf{Zg} + \mathbf{e}$$

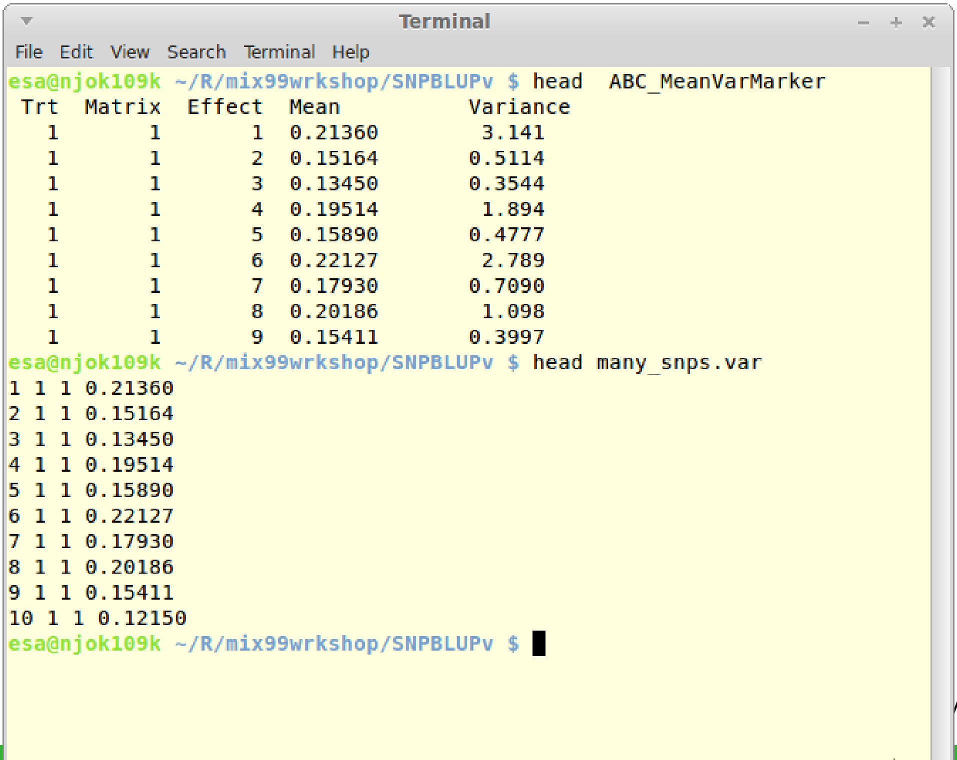
  - simply add PEDFILE, PEDIGREE and the animal into model

- Use multiple trait models
  - simply add more model lines, and respesify the PARFILE
  - Especially useful with zero correlations,
    because it allows to run multiple runs simultaneously

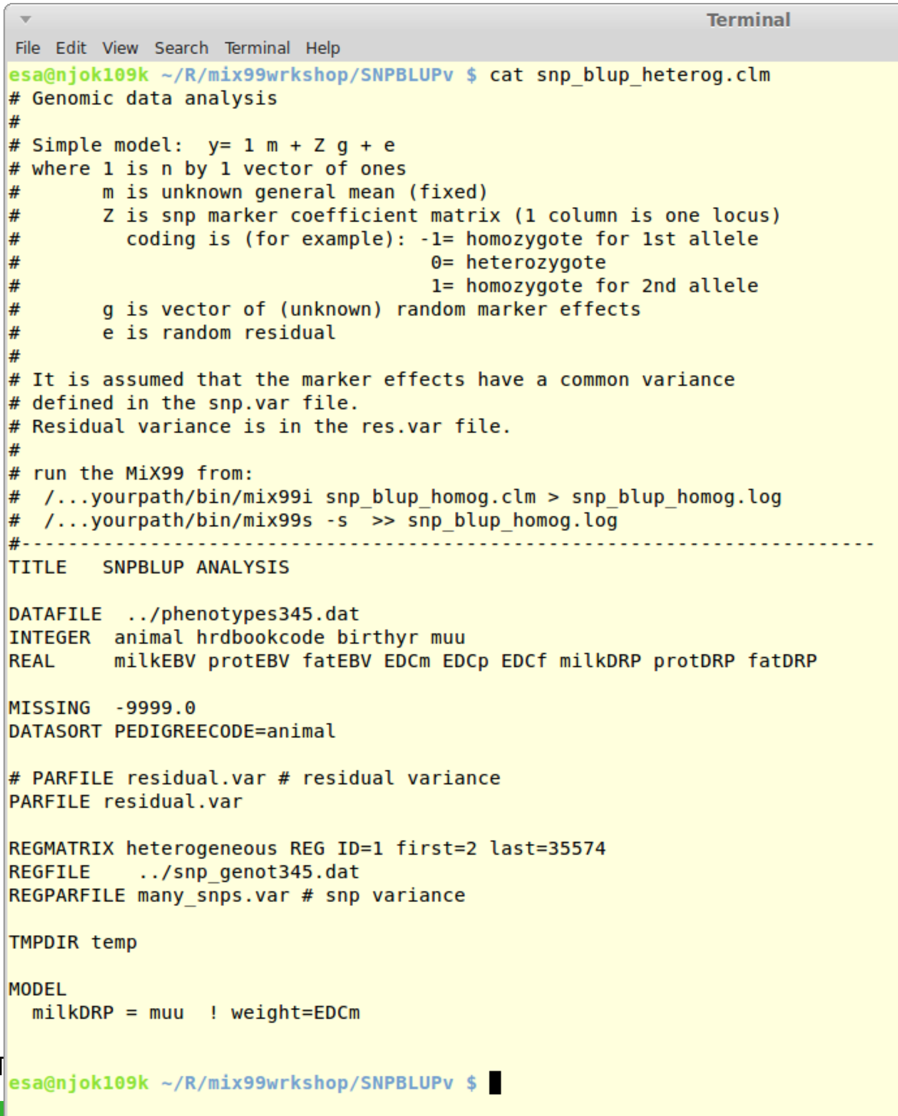
- Use marker specific variances
  - Specify REGMATRIX type as HETEROGENOUS
  - respesify REGPARFILE to span all m markers

“Fast BayesB MAP SNP BLUP”

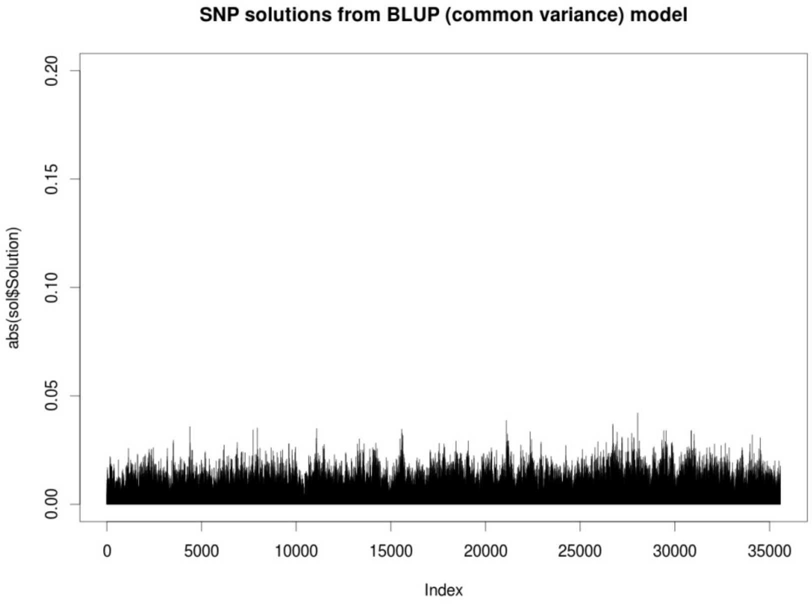
- Estimate the individual marker variances
  - eg. abc_mix <options.abc



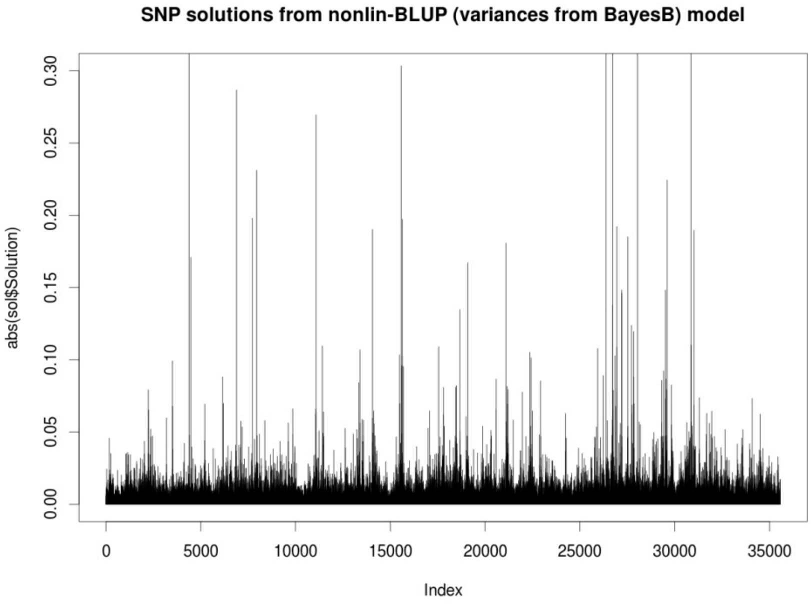
```
esa@njok109k ~/R/mix99wrkshop/SNPBLUPv $ head  ABC_MeanVarMarker
 Trt   Matrix  Effect  Mean           Variance
   1        1       1  0.21360         3.141
   1        1       2  0.15164        0.5114
   1        1       3  0.13450        0.3544
   1        1       4  0.19514         1.894
   1        1       5  0.15890        0.4777
   1        1       6  0.22127         2.789
   1        1       7  0.17930        0.7090
   1        1       8  0.20186         1.098
   1        1       9  0.15411        0.3997
esa@njok109k ~/R/mix99wrkshop/SNPBLUPv $ head many_snps.var
1 1 1 0.21360
2 1 1 0.15164
3 1 1 0.13450
4 1 1 0.19514
5 1 1 0.15890
6 1 1 0.22127
7 1 1 0.17930
8 1 1 0.20186
9 1 1 0.15411
10 1 1 0.12150
esa@njok109k ~/R/mix99wrkshop/SNPBLUPv $ 
```



```
esa@njok109k ~/R/mix99wrkshop/SNPBLUPv $ cat snp_blup_heterog.clm
# Genomic data analysis
#
# Simple model:  y= 1 m + Z g + e
# where 1 is n by 1 vector of ones
#       m is unknown general mean (fixed)
#       Z is snp marker coefficient matrix (1 column is one locus)
#         coding is (for example): -1= homozygote for 1st allele
#                                   0= heterozygote
#                                   1= homozygote for 2nd allele
#       g is vector of (unknown) random marker effects
#       e is random residual
#
# It is assumed that the marker effects have a common variance
# defined in the snp.var file.
# Residual variance is in the res.var file.
#
# run the MiX99 from:
#  /...yourpath/bin/mix99i snp_blup_homog.clm > snp_blup_homog.log
#  /...yourpath/bin/mix99s -s  >> snp_blup_homog.log
#------------------------------------------------------------------------
TITLE   SNPBLUP ANALYSIS

DATAFILE  ../phenotypes345.dat
INTEGER  animal hrdbookcode birthyr muu
REAL     milkEBV protEBV fatEBV EDCm EDCp EDCf milkDRP protDRP fatDRP

MISSING  -9999.0
DATASORT PEDIGREECODE=animal

# PARFILE residual.var # residual variance
PARFILE residual.var

REGMATRIX heterogeneous REG ID=1 first=2 last=35574
REGFILE    ../snp_genot345.dat
REGPARFILE many_snps.var # snp variance

TMPDIR temp

MODEL
  milkDRP = muu  ! weight=EDCm


esa@njok109k ~/R/mix99wrkshop/SNPBLUPv $ 
```



SNP solutions from BLUP (common variance) model



SNP solutions from nonlin-BLUP (variances from BayesB) model

Walk through example

GBLUP

GBLUP Model

G-BLUP can be derived easily by substituting $\mathbf{a}_t = \mathbf{Z}_t\mathbf{g}$:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{a}_t + \mathbf{e}$$

$\Rightarrow \mathrm{Var}(\mathbf{a}_t) = \mathrm{Var}(\mathbf{Z}_t\mathbf{g}) = \mathbf{Z}_t\,\mathrm{Var}(\mathbf{g})\mathbf{Z}_t' = \mathbf{Z}_t\mathbf{Z}_t'\sigma_g^2$
when $\mathrm{Var}(\mathbf{g}) = \mathbf{I}\sigma_g^2$.

$\Rightarrow \mathbf{a}_t \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$ where genomic matrix is $\mathbf{G} = \mathbf{Z}_t\mathbf{Z}_t'$.
and correspondingly $\mathbf{y} \sim N(\mu, (\mathbf{R} + \mathbf{G}\sigma_g^2))$

GBLUP vs. SNP-BLUP

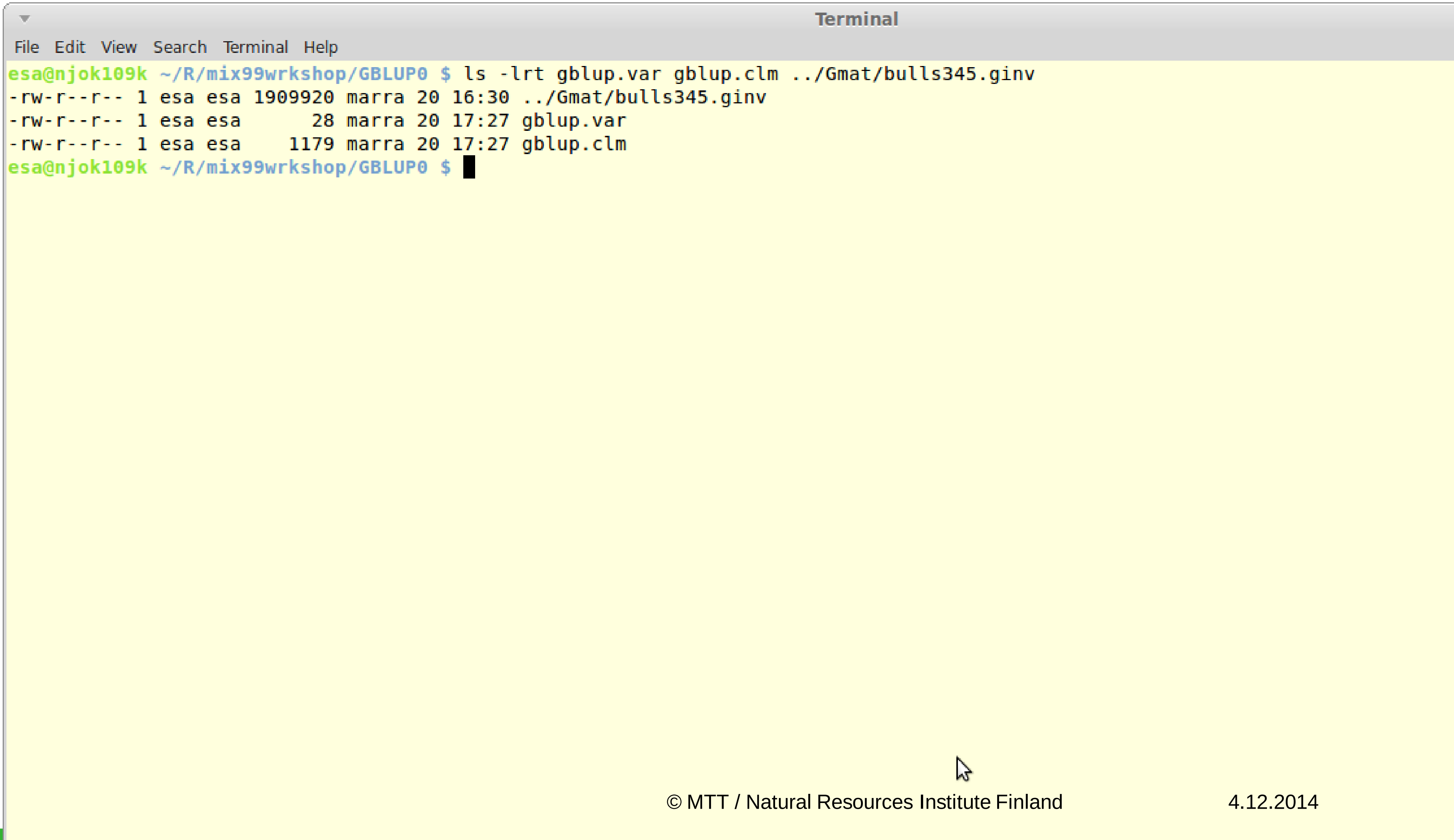
SNP-BLUP and G-BLUP give the same $\mathbf{Z}\hat{\mathbf{g}}$, i.e. the same DGVs

SNP-BLUP has $m + 1$ unknowns

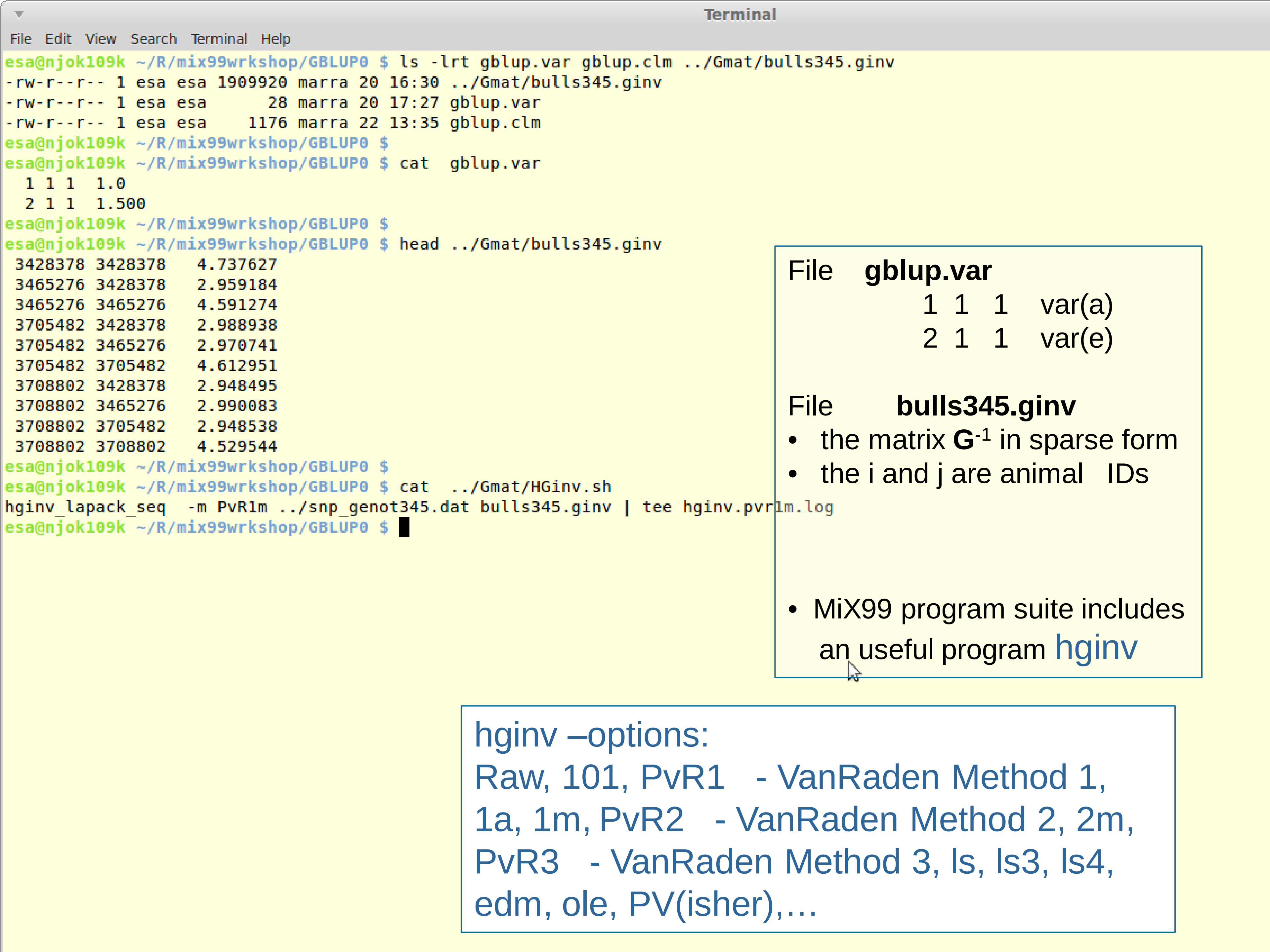
G-BLUP has $n + 1$ unknowns.

- Often currently number of marker effects $m >> n$
- the G-BLUP requires building and inverting the **G** matrix.
- G-BLUP is similar to traditional pedigree relationship matrix based animal model BLUP: **A** replaced by $\mathbf{G} = \mathbf{Z}\mathbf{Z}'$.

Running the GBLUP in MiX99



```
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ ls -lrt gblup.var gblup.clm ../Gmat/bulls345.ginv
-rw-r--r-- 1 esa esa 1909920 marra 20 16:30 ../Gmat/bulls345.ginv
-rw-r--r-- 1 esa esa      28 marra 20 17:27 gblup.var
-rw-r--r-- 1 esa esa    1179 marra 20 17:27 gblup.clm
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ 
```



```
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ ls -lrt gblup.var gblup.clm ../Gmat/bulls345.ginv
-rw-r--r-- 1 esa esa 1909920 marra 20 16:30 ../Gmat/bulls345.ginv
-rw-r--r-- 1 esa esa      28 marra 20 17:27 gblup.var
-rw-r--r-- 1 esa esa    1176 marra 22 13:35 gblup.clm
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ cat  gblup.var
  1 1 1  1.0
  2 1 1  1.500
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ head ../Gmat/bulls345.ginv
 3428378 3428378   4.737627
 3465276 3428378   2.959184
 3465276 3465276   4.591274
 3705482 3428378   2.988938
 3705482 3465276   2.970741
 3705482 3705482   4.612951
 3708802 3428378   2.948495
 3708802 3465276   2.990083
 3708802 3705482   2.948538
 3708802 3708802   4.529544
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ cat  ../Gmat/HGinv.sh
hginv_lapack_seq  -m PvR1m ../snp_genot345.dat bulls345.ginv | tee hginv.pvr1m.log
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $
```

File **gblup.var**

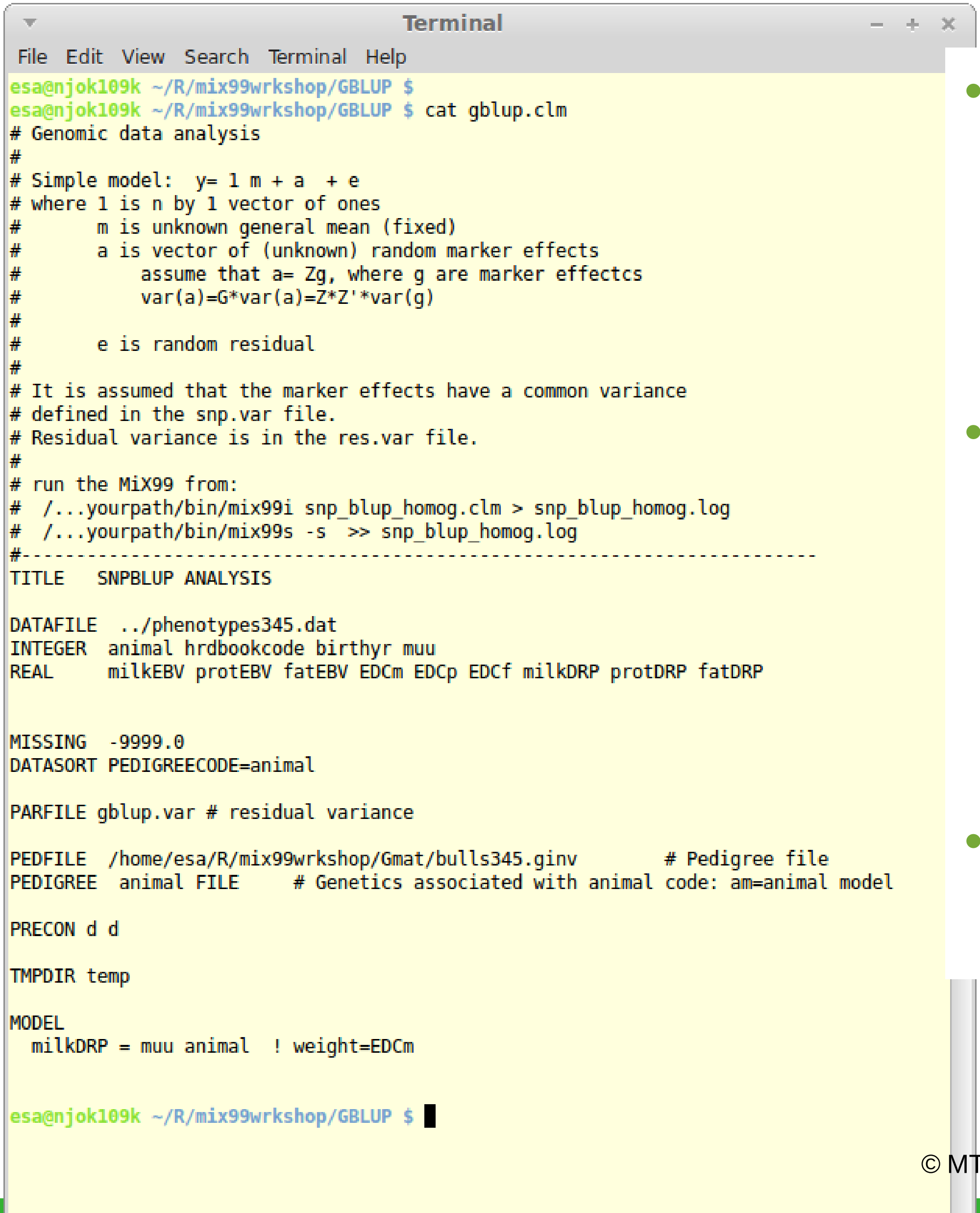
1 1 1 var(a)
2 1 1 var(e)

File **bulls345.ginv**

- the matrix **G**$^{-1}$ in sparse form
- the i and j are animal IDs

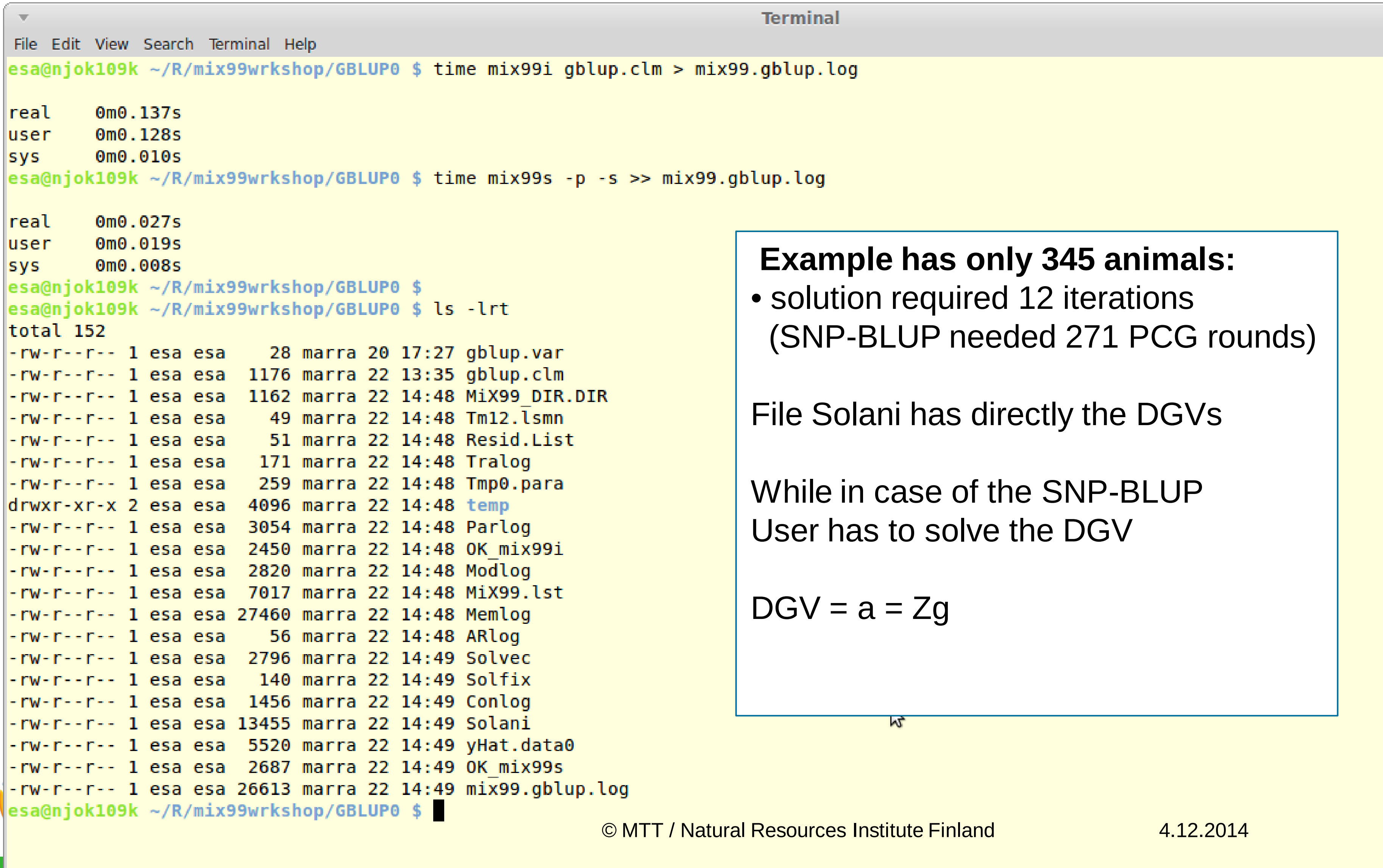
- MiX99 program suite includes an useful program hginv

hginv –options:
Raw, 101, PvR1 - VanRaden Method 1,
1a, 1m, PvR2 - VanRaden Method 2, 2m,
PvR3 - VanRaden Method 3, ls, ls3, ls4,
edm, ole, PV(isher),…



```
esa@njok109k ~/R/mix99wrkshop/GBLUP $
esa@njok109k ~/R/mix99wrkshop/GBLUP $ cat gblup.clm
# Genomic data analysis
#
# Simple model:  y= 1 m + a  + e
# where 1 is n by 1 vector of ones
#       m is unknown general mean (fixed)
#       a is vector of (unknown) random marker effects
#           assume that a= Zg, where g are marker effectcs
#           var(a)=G*var(a)=Z*Z'*var(g)
#
#       e is random residual
#
# It is assumed that the marker effects have a common variance
# defined in the snp.var file.
# Residual variance is in the res.var file.
#
# run the MiX99 from:
#  /...yourpath/bin/mix99i snp_blup_homog.clm > snp_blup_homog.log
#  /...yourpath/bin/mix99s -s  >> snp_blup_homog.log
#--------------------------------------------------------------------------
TITLE   SNPBLUP ANALYSIS

DATAFILE  ../phenotypes345.dat
INTEGER  animal hrdbookcode birthyr muu
REAL     milkEBV protEBV fatEBV EDCm EDCp EDCf milkDRP protDRP fatDRP


MISSING  -9999.0
DATASORT PEDIGREECODE=animal

PARFILE gblup.var # residual variance

PEDFILE  /home/esa/R/mix99wrkshop/Gmat/bulls345.ginv        # Pedigree file
PEDIGREE  animal FILE     # Genetics associated with animal code: am=animal model

PRECON d d

TMPDIR temp

MODEL
  milkDRP = muu animal  ! weight=EDCm


esa@njok109k ~/R/mix99wrkshop/GBLUP $ █
```

- Instructions not much different than before
  - the same data file
- Pedigree file is replaced by **G**$^{-1}$ –file
  - Specification FILE
- Model statement now includes also animal



```
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ time mix99i gblup.clm > mix99.gblup.log

real	0m0.137s
user	0m0.128s
sys	0m0.010s
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ time mix99s -p -s >> mix99.gblup.log

real	0m0.027s
user	0m0.019s
sys	0m0.008s
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ ls -lrt
total 152
-rw-r--r-- 1 esa esa    28 marra 20 17:27 gblup.var
-rw-r--r-- 1 esa esa  1176 marra 22 13:35 gblup.clm
-rw-r--r-- 1 esa esa  1162 marra 22 14:48 MiX99_DIR.DIR
-rw-r--r-- 1 esa esa    49 marra 22 14:48 Tm12.lsmn
-rw-r--r-- 1 esa esa    51 marra 22 14:48 Resid.List
-rw-r--r-- 1 esa esa   171 marra 22 14:48 Tralog
-rw-r--r-- 1 esa esa   259 marra 22 14:48 Tmp0.para
drwxr-xr-x 2 esa esa  4096 marra 22 14:48 temp
-rw-r--r-- 1 esa esa  3054 marra 22 14:48 Parlog
-rw-r--r-- 1 esa esa  2450 marra 22 14:48 OK_mix99i
-rw-r--r-- 1 esa esa  2820 marra 22 14:48 Modlog
-rw-r--r-- 1 esa esa  7017 marra 22 14:48 MiX99.lst
-rw-r--r-- 1 esa esa 27460 marra 22 14:48 Memlog
-rw-r--r-- 1 esa esa    56 marra 22 14:48 ARlog
-rw-r--r-- 1 esa esa  2796 marra 22 14:49 Solvec
-rw-r--r-- 1 esa esa   140 marra 22 14:49 Solfix
-rw-r--r-- 1 esa esa  1456 marra 22 14:49 Conlog
-rw-r--r-- 1 esa esa 13455 marra 22 14:49 Solani
-rw-r--r-- 1 esa esa  5520 marra 22 14:49 yHat.data0
-rw-r--r-- 1 esa esa  2687 marra 22 14:49 OK_mix99s
-rw-r--r-- 1 esa esa 26613 marra 22 14:49 mix99.gblup.log
esa@njok109k ~/R/mix99wrkshop/GBLUP0 $ 
```

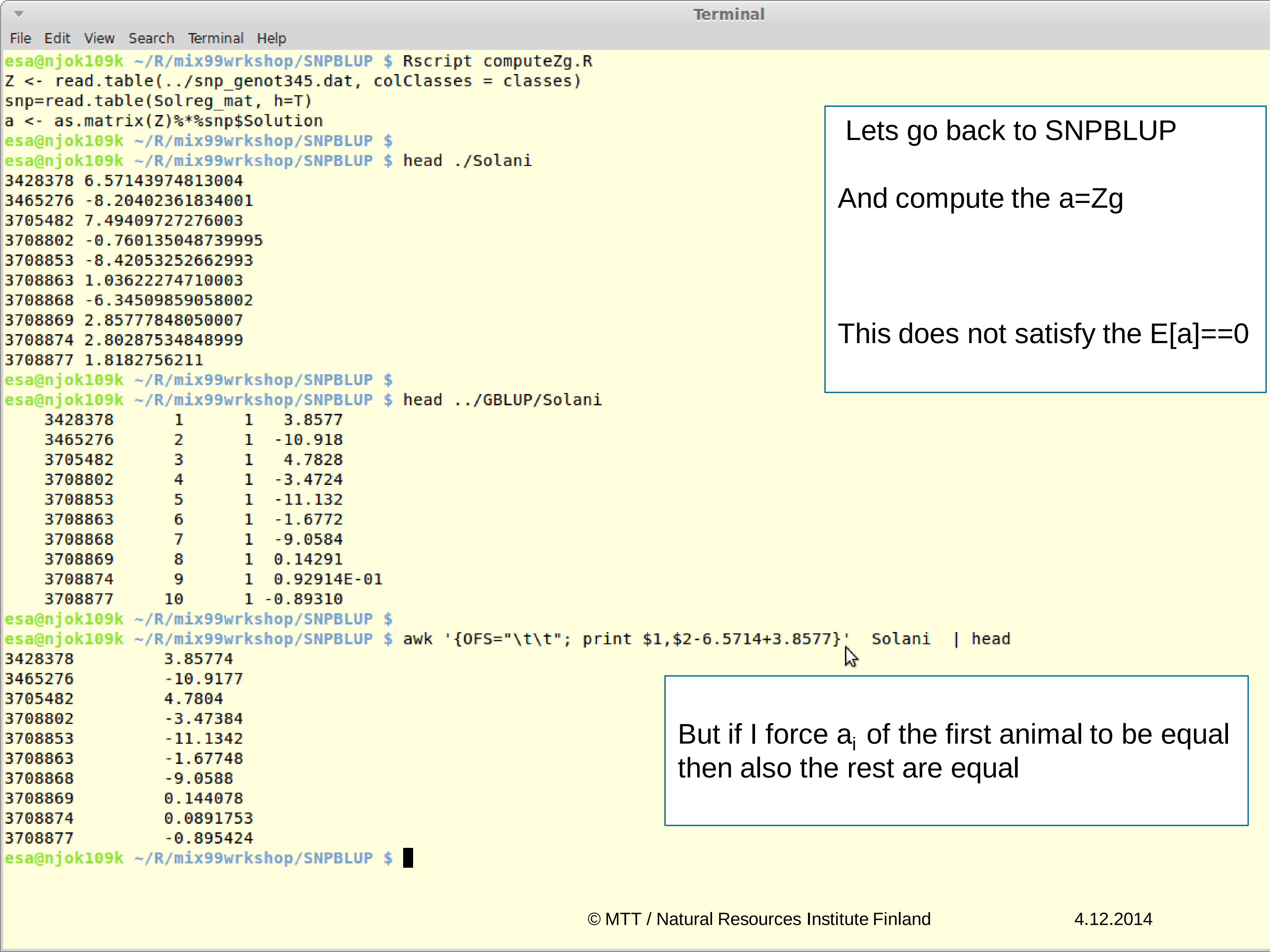
Example has only 345 animals:

- solution required 12 iterations (SNP-BLUP needed 271 PCG rounds)

File Solani has directly the DGVs

While in case of the SNP-BLUP
User has to solve the DGV

DGV = a = Zg



```
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ Rscript computeZg.R
Z <- read.table(../snp_genot345.dat, colClasses = classes)
snp=read.table(Solreg_mat, h=T)
a <- as.matrix(Z)%*%snp$Solution
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ head ./Solani
3428378 6.57143974813004
3465276 -8.20402361834001
3705482 7.49409727276003
3708802 -0.760135048739995
3708853 -8.42053252662993
3708863 1.03622274710003
3708868 -6.34509859058002
3708869 2.85777848050007
3708874 2.80287534848999
3708877 1.8182756211
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ head ../GBLUP/Solani
    3428378      1      1   3.8577
    3465276      2      1  -10.918
    3705482      3      1   4.7828
    3708802      4      1  -3.4724
    3708853      5      1  -11.132
    3708863      6      1  -1.6772
    3708868      7      1  -9.0584
    3708869      8      1  0.14291
    3708874      9      1  0.92914E-01
    3708877     10      1 -0.89310
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $ awk '{OFS="\t\t"; print $1,$2-6.5714+3.8577}'  Solani  | head
3428378         3.85774
3465276         -10.9177
3705482         4.7804
3708802         -3.47384
3708853         -11.1342
3708863         -1.67748
3708868         -9.0588
3708869         0.144078
3708874         0.0891753
3708877         -0.895424
esa@njok109k ~/R/mix99wrkshop/SNPBLUP $
```

Lets go back to SNPBLUP

And compute the a=Zg

This does not satisfy the E[a]==0

But if I force a_i of the first animal to be equal then also the rest are equal

More advanced GBLUP …

- Add polygenic effect to the model

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Wu} + \mathbf{a} + \mathbf{e}$$

  - You need PEDFILE; PEDIGREE; COVFILE and RANDOM instructions
  - MODEL : “y = muu genomic(animal) polygenic(animal)”
- Use multiple trait models
  - simply add more model lines, and respecify the PARFILE
  - Especially useful with zero correlations, because it allows to run multiple runs simultaneously
- To use marker specific variances
  - Specify the **G**-matrix as you like: **G=ZDZ**’, with any **D** (and/or **Z**)

Bottom line

- MiX99 package operates equally with pedigree and genomic models.
  - Usable utilities like exa99 and ApaX, plus few genomic specific tools like hginv, and SNP-based PEV programs.

- Simple GBLUP and SNP-BLUP are easy to run with MiX99
- More advanced models possible, read the CLIM manual

- Ambitious development plans:
  - Single-step evaluations for very large populations
  - Continuous evaluations etc.