

Today, you will work on the computational cluster located in Novosibirsk, Russia. There are few differences between the Novosibirsk cluster and the local server you used before. Notably, `pico` editor is not available and `jpico`, a very similar program, is to be used instead. You will all log in using the same user name, and then you should change to your own directory. Since you will share login with other users, be extremely careful in order not to delete other users' files!

When you produce a file (graphs) and want to retrieve it with WinSCP, do not forget to refresh the directory content (F5).

---

Throughout these instructions we use following notations:

The command is to be issued in command line (bash):

```
bash$ command
```

The command is to be issued in R statistical analysis package:

```
> command
```

The command is to be issued in SOLAR

```
solar> command
```

The comments which explain what this or that command is supposed to do are presented as

```
# comment
```

---

Log into the `mga.bionet.nsc.ru` server using username **stunihes1** and password **a3Fd1\_5** (the character before the underscore is digit “one” not letter “L”)

Issue the command

```
bash$ cd /mnt/xfss/nihes
```

This will bring you into working directory. There are 16 directories named `sn1`, `sn2` ... or `sn16`; your directory corresponds to the number you were assigned with at the start of the course.

There are two directories present:

```
complextrait
bigped
```

We start with complex trait analysis, thus change to “`complextrait`”

Repeat the login procedure again, in order to have two telnet sessions running; one will be used to run SOLAR and other to make complementary analyses.

## Exercise 1

During this exercise you will estimate IBD using MERLIN and perform heritability and QTL linkage analysis under a range of different models with SOLAR.

In “`complextrait`” directory, four data files are present (use `ls` command to get the file list):

```
solar.ped    – SOLAR pedigree data file
solar.mrk    – SOLAR marker data file
solar.map    – SOLAR map file
solar.phe    – SOLAR phenotypic data file
```

The pedigrees are represented graphically in the file `peds.pdf` (use WinSCP to retrieve the file and do not forget that the files are located at `/mnt/xfs/nihes/sn[your-number]`).

Inspect these files and answer the questions:

```
How many pedigrees are available for analysis?      ___ 232 ___
The largest pedigree is pedigree #1. What is it's size? ___ 20 ___
What is the bit-size of the pedigree #1?           ___ 13 ___
Can this pedigree be analysed using exact methods (MERLIN, GH)? ___ Yes ___
How many markers are available for analysis?        ___ 406 ___
What is the length of chromosome 6?                 ___ 190 cM ___
How many traits are available for analysis?         ___          ___
What are the names of these traits?                 ___          ___
```

For analysis 2 quantitative traits are available (`qt.cvr`); binary trait “gender” and two other QTs, named `male_qt` and `female_qt`, which represent the trait `qt`, stratified by gender

We are first going to explore the phenotypic data using R.

Start R and apply following commands:

```
# load the phenotypic data into R
> df <- read.csv("solar.phe")
# inspect variable names
> names(df)
# attach the data in order that “qt” may be used instead of df$qt
> attach(df)
```

The trait of interest is named “qt”. We want to describe this trait and it’s relation with sex and covariate named “cvr”.

To explore these relations graphically, use the following commands to produce graphs:

```
# direct graphic output to the file
> pdf(file="graphs.pdf")
# plot qt vs. cvr, using blue points
```

```

> plot(qt, cvr, col="blue", pch=19)
# make points corresponding to male red
> points(qt[gender == 1], cvr[gender == 1], col="red", pch=19)
# next use 2 graphs per page
> par(mfcol=c(2,1))
# draw histogram of qt for male; x-axis is limited to (-4,2)
> hist(qt[gender == 1],xlim=c(-4,2))
# draw histogram of qt for female
> hist(qt[gender == 2],xlim=c(-4,2))
# turn off graphic output
> dev.off()

```

Use WinSCP to inspect graphs produced by R (graphs.pdf)

Do you see any notable difference between sexes? Yes  
 Do you see any relation between qt and cvr? \_\_\_\_\_

**Positive correlation for men and no correlation for women**

Return to R and answer following questions (look at "Useful R commands" first)

What is mean trait value? -0.078  
 What is trait's variance? 0.736  
 Is the trait distributed normally? No (P-value = 0.0002)  
 What is mean trait value in male? -0.45  
 What is trait's variance in male? 0.86  
 Is the trait distributed normally in male? Yes (P-value = 0.21)  
 What is mean trait value in female? 0.223  
 What is trait's variance in female? 0.43  
 Is the trait distributed normally in female? Yes (P-value = 0.69)  
 Is there significant difference between male and female in terms of the QT mean?  
Yes (P-value = <0.0001)?  
 How large is this difference? 0.674  
 What is the regression coefficient of qt onto cvr? 0.145 (P-value = <0.0001)

Leave R and start SOLAR by using `solar` command

```

# load pedigree data into SOLAR; kinship matrix will be computed automatically
solar> load ped solar.ped
# load phenotypic data
solar> load pheno solar.phe

```

Now, we are going to estimate heritability of the trait "qt"

```

# select the trait
solar> trait qt
# estimate heritability
solar> polyg

```

What is qt heritability estimate (s.e.)? 0.31 (s.e. = 0.06)  
 How significant is deviation from zero? <0.0001

In R analysis, we have seen that qt value is different between genders. Let us include sex as a covariate into polygenic analysis:

```
# tells SOLAR that we are going to define new model
solar> model new
# select the trait
solar> trait qt
# add sex as a covariate into the model
solar> cov sex
# estimate heritability
solar> polyg
```

What is qt heritability estimate (s.e.) when adjustment for **sex** is made? 0.37 (s.e. = 0.06)  
How significant is deviation from zero? <0.0001

Repeat heritability analysis using cvr as covariate

What is qt heritability estimate (s.e.) when adjustment for **cvr** is made? 0.29 (s.e. = 0.06)  
How significant is deviation from zero? <0.0001

Repeat heritability analysis using both sex and cvr as covariates

What is qt heritability estimate (s.e.) when adjustment for **sex and cvr** is made? 0.33 (s.e. = 0.06)  
How significant is deviation from zero? <0.0001

From this analysis, what would be a **sure guess** for covariates to use in linkage?

- sex
- cvr
- sex and cvr
- none

As we have seen in R analysis, there is notable difference between qt values in male and female, both in terms of the mean and the variance. This may indicate that the control of the QT is different in different sexes. To test this hypothesis, we will perform bi-variate heritability analysis. To allow such analysis, the trait information must be partitioned into two variables, the first having qt values for one sex and the second for the other. In the solar.phe file, this partitioning was already done (traits named male\_qt and female\_qt)

```
# tells SOLAR that we are going to define new model
solar> model new
# two traits for bi-variate analysis
solar> trait male_qt female_qt
# specify polygenic model
solar> polymod
# fix environmental correlation as 0
solar> fix rhoe
# maximise the model
solar> maxi
# perform complete bi-variate heritability analysis
```

```
solar> polyg
```

What is female heritability estimate (s.e.)?	<u>0.69</u>	(s.e. = <u>0.11</u> )
What is male heritability estimate (s.e.)?	<u>0.44</u>	(s.e. = <u>0.11</u> )
What is the estimate of genetic correlation (RhoG)?	<u>0.15</u>	

Now we will start linkage analysis of the trait qt.

First of all, using SOLAR we are going to prepare files for IBD estimation; second, we will use MERLIN to estimate IBD; third, we will import these IBD estimates into SOLAR and finally we will do QTL linkage.

To prepare file for IBD estimation:

```
# make directory where the IBD matrices will be stored
solar> mkdir mibd6
# tell SOLAR that mibd6 directory is used for storage of IBD matrices
solar> mibddir mibd6
# tell what chromosome will be analysed
solar> chrom 6
# load genetic marker map
solar> load map solar.map
# load marker data
solar> load marker solar.mrk
# estimate marker allelic frequencies
solar> freq mle
# initially, scan at 10cM spacing. If some point will show LOD>0.59, these regions will be
# re-scanned using 1cM interval
solar> interval 10
# prepare files for MERLIN IBD estimation
solar> mibd prep merlin
```

Now, use the second telnet session to compute IBD using MERLIN. To do this you will run the command file `mlmibd.cmd`, which runs MERLIN

```
bash$ sh mlmibd.cmd
```

You can actually inspect the content of `mlmibd.cmd` (e.g. `less mlmibd.cmd`) to see what the MERLIN was requested to do.

When IBD analysis is finished, you may inspect the resulting IBD file (e.g. `less merlin.ibd`); then return to the SOLAR session and import the IBD matrices using

```
# tells SOLAR to import IBD matrices produced by MERLIN for chromosome 6
solar> mibd import merlin 6
```

Now we are ready to do linkage analysis. First, we need to specify and estimate the heritability model again and then multipoint analysis is easily performed.

```
# delete all old model information
solar> model new
# analysis of the trait qt, no covariates
```

```
solar> trait qt
# estimate the model
solar> polyg
# start multipoint QTL linkage analysis
solar> mul -ov
```

Maximum LOD score of 1.17 occurs at the position 70 cM

You can produce and save LOD score graphs by switching to other session and

```
bash$ perl lodgraph.pl qt/multipoint1.out qt-nocovar
```

which will convert the LOD score data stored in qt/multipoint1.out into a pdf file named qt-nocovar.pdf (or any other for your choice!)

Repeat the analysis using sex as a covariate. Act as you did when you estimated heritability and then, after “polyg”, use “mul -ov”. If you want to save the LOD graphs, do not forget produce them before you start new analysis!

When adjustment for sex is made, maximum LOD score of 2.06 occurs at the position 69 cM

When adjustment for cvr is made, maximum LOD score of 0.82 occurs at the position 102 cM

When bivariate analysis is performed maximum LOD score of 3.64 occurs at the position 73 cM

For this analysis, LOD score graph can be produced using the command

```
bash$ perl lodgraph.pl male_qt.female_qt/multipoint1.out qt-bivariate
```

Exit SOLAR by

```
solar> quit
```

Based on your results,

What can you tell about the type of relation between qt and sex? \_\_\_\_\_

If you inspect the male\_qt.female\_qt/multipoint1.out you will see that the QTL heritability (at the location of maximum) is about 70% for men and almost zero for female. This tells that the gene work in men, but not in women.

What can you tell about the type of relation between qt and cvr? \_\_\_\_\_

cvr may be independently controlled by the gene, or it could be an endophenotype for qt.

## Exercise 2

During this exercise, you will estimate IBD matrices for large pedigrees using LOKI and will perform QTL linkage analysis using SOLAR.

In both telnet sessions, change to the “bigped” directory by

```
bash$ cd ../bigped
```

The pedigrees are represented graphically in the file `peds.pdf` (use WinSCP to retrieve the file). Inspect these graphs and SOLAR files and answer

How many pedigrees are present? 5

What is the bit-size of pedigree #4? 122

Assume that on some fast computer MERLIN analysis of a pedigree of bit-size 18 takes 1 sec. Every additional bit increases the computation time by the factor of 4 (thus 20-bit pedigree will be analysed in 16 sec.). What will be the time to estimate IBD for pedigree #4 using MERLIN? \_\_\_\_\_

$4^{104}$  sec = 130447 x  $10^{50}$  years. The age of the universe is  $2 \times 10^{10}$  years ... It will take long time!

How many traits are present for analysis? 1

How many markers are present? 121

What is the length of chromosome 17? 81 cM

Start solar and estimate heritability of qt:

```
# load pedigree data into SOLAR; kinship matrix will be computed automatically
solar> load ped solar.ped
# load phenotypic data
solar> load pheno solar.phe
# select the trait
solar> trait qt
# estimate heritability
solar> polyg
```

What is qt heritability estimate (s.e.)? 0.55 (s.e. = 0.096)

How significant is deviation from zero? <0.0001

Now we are going to use LOKI to estimate IBD matrix. First, we will prepare LOKI files using SOLAR:

```
# make directory where the IBD matrices will be stored
solar> mkdir mibd17
# tell SOLAR that mibd17 directory is used for storage of IBD matrices
solar> mibddir mibd17
# tell what chromosome will be analysed
solar> chrom 17
# load genetic marker map
solar> load map solar.map
```

```
# load marker data
solar> load marker solar.mrk
# estimate marker allelic frequencies
solar> freq mle
# initially, scan at 10cM spacing. If some point will show LOD>0.59, these regions will be
# re-scanned using 1cM interval
solar> interval 10
# prepare files for LOKI IBD estimation
solar> mibd prep loki
```

Now, go to the second telnet window

```
# convert files into internal LOKI format
bash$ prep lkmibd.prep
```

We will first perform fast analysis using very small number of iterations.

Edit the file lkmibd.loki that the first three lines read:

```
iterations 1000
start output 500,1
output frequency 1,1
```

This instructs LOKI to use 1000 iterations in total (line 1). The output to the IBD file will start at iteration 500 and the one to the screen will start at iteration 1 (line 2). The output to both file and screen will happen on every iteration (line 3).

Run LOKI to estimate IBD:

```
bash$ loki lkmibd.loki
```

When LOKI finishes, return to SOLAR session and import IBD data

```
solar> mibd import loki 17
```

Compute LOD scores using

```
solar> mul -ov
```

The highest LOD score of \_\_\_\_\_ is observed at position \_\_\_\_\_ cM.

The LODs you observe vary between runs. Typically you will get LODs between 3 and 8 with mean of about 5; however, the location of the peak will be pretty close to the true location at 37 cM

Let us see how robust are the IBD estimates. For this, return to other telnet session and increase the number of iterates:

Edit the file lkmibd.loki that the first three lines read:

```
iterations 2000
```



```
start output 1000,1
output frequency 1,100
```

Run LOKI to estimate IBD:

```
bash$ loki lkmibd.loki
```

When LOKI finishes, return to SOLAR session and import IBD data

```
solar> mibd import loki 17
```

Compute LOD scores using

```
solar> mul -ov
```

The highest LOD score of \_\_\_\_\_ is observed at position \_\_\_\_\_ cM.

If the time permits, repeat the analysis using 10,000 iterates with 5000 burn-in period.

The highest LOD score of \_\_\_\_\_ is observed at position \_\_\_\_\_ cM.

The IBD matrices were computed using 100,000 iterates; these are stored at /mnt/xf/nihes/100K.ibd.gz. Use these matrices for QTL linkage analysis by

```
solar> mibd import loki -file /mnt/xf/nihes/100K.ibd.gz 17
```

The highest LOD score of \_\_\_\_\_ is observed at position \_\_\_\_\_ cM.