



**A computer package of serial computing programs for
epistasis testing in genome-wide association studies**

USER MANUAL VERSION 2.0

**LI MA, DANIEL DVORKIN, JOHN R. GARBE
H. BIRALI RUNESHA, YANG DA**

**Department of Animal Science and Supercomputer Institute
University of Minnesota**

March 27, 2008

AUTHOR CONTRIBUTIONS

Li Ma is the author of the EPISNP and CPUHD programs, and is the main author of the EPISNPmpi program that is the basis of the current EPISNP program.

Daniel Dvorkin is the author of the EPISNPLOT program.

John R. Garbe is the author of the EPINET program.

H. Birali Runesha directed the development of the parallel computing coding, and did a portion of the coding of the EPISNPmpi program that is the basis of the current EPISNP program, and compiled the current EPISNP and CPUHD programs for various operation systems.

Yang Da is the project leader and the lead writer of this user manual.

RELEASE HISTORY

epiSNP version 1.0w: First release by L. Ma, D. Dvorkin and Y. Da, September 1 2006

epiSNP version 1.1w: Second release by L. Ma, D. Dvorkin, J.R. Garbe and Y. Da, March 19 2007.

The EPISNPLOT program adjusted the sample size range so that sample size data points generally do not overlap with data points of the most significant effects.

The EPINET program was added for making figures of epistasis networks.

The EPISNP program improved computing speed and fixed a few bugs.

epiSNP version 2.0: Third release by L. Ma, D. Dvorkin, J.R. Garbe, H.B. Runesha and Y. Da, March 27 2008.

The EPINET program now is based on the EPISNPmpi parallel computing code that was improved over the previous EPISNP code. The current EPISNP and CPUHD programs are compiled for Windows, Unix and Linux operation systems, while EPISNPLOT and EPINET remains to be Windows programs.

1	Introduction	4
1.1	Functionality and Applicability of epiSNP	4
1.2	Executable Programs of epiSNP.....	4
1.2.1	The CPUHD program.....	4
1.2.2	The EPISNP program.....	4
1.2.3	The EPISNP PLOT program.....	5
1.2.4	The EPINET program.....	5
1.2.5	Downloadable programs.....	5
2	Input files.....	6
2.1	Parameter File (parameter.dat)	6
2.2	SNP genotype Files	7
2.3	Phenotype (quantitative trait) File	8
3	Running the programs	9
3.1	Windows Operation System	9
3.1.1	Running the CPUHD program	9
3.1.2	Running the EPISNP program.....	10
3.1.3	Running the EPISNP PLOT program.....	12
3.1.4	Running the EPINET program	17
3.2	Linux and Unix Operation Systems.....	19
FIGURE 1	14
FIGURE 2	15
FIGURE 3	16
FIGURE 3	18
4	References	19

1. INTRODUCTION

1.1 Functionality and applicability of epiSNP

epiSNP is a computer package designed for genome-wide SNP candidate gene analysis of quantitative traits with emphasis on testing SNP epistasis effects. Main features of epiSNP are:

- Estimates of running time and storage requirements
- Pairwise statistical test for SNP epistasis effects on a quantitative trait
- Statistical test for single SNP effects on a quantitative trait
- Chromosome view of significance results and sample size for single SNP tests
- Graphical view of epistasis network

The epiSNP computer package is applicable to all bi-allelic loci of diploid species. The X chromosome loci can be analyzed but only females can be used. Similarly, the Z chromosome loci in birds can be analyzed but only male individuals can be used. The program assumes male to be the heterogametic sex, which is the case for mammals. For birds, the user can flip the gender definitions in the parameter file described in **2.1 Parameter file (parameter.dat)**, i.e., in the parameter.dat file, male birds are defined as females and female birds are defined males.

1.2 Executable programs of epiSNP

The epiSNP computer package has three executable programs: CPUHD, EPISNP, and EPISNPLOT. The CPUHD and EPINET programs are compiled for Windows, Unix and Linux operation systems, while the EPISNPLOT and EPINET are Windows programs.

1.2.1 The CPUHD program

The purpose of CPUHD is to estimate CPU time required to complete the data analysis using EPISNP program and the total storage space required to store the output files. These estimates will allow the user to determine whether the computer being used is capable of completing the data analysis within an acceptable time frame and whether the computer has enough disk storage space to hold the output files. It is highly recommended to run CPUHD before running EPISNP.

The CPUHD program requires one input file, episnp.par, which is the parameter file with user specified controls for running EPISNP program.

1.2.2 The EPISNP program

The EPISNP program conducts single-SNP tests and pairwise SNP epistasis tests. The single-SNP test provides three significance tests for each SNP: the overall effect, additive effect, and dominance effect of the SNP. The pairwise test provides five significance tests for each pair of SNPs: the interaction between the two SNPs, additive \times additive, additive \times dominance, dominance \times additive, and dominance \times dominance epistasis effects.

The EPISNP program allows an arbitrary number of fixed non-genetic effects and covariables such as gender, age, birth weight, and treatment and living conditions, and allows arbitrary numbers of SNPs, chromosomes, and quantitative traits. The method of statistical test for pairwise epistasis effects is based on an extended Kempthorne model (Mao et al., 2006). The EPISNP program requires three input files: a parameter file (episnp.par), a user-named SNP genotype data file, and a user-named phenotypic file with observations of the quantitative trait(s).

1.2.3 The EPISNP PLOT program

The purpose of EPISNP PLOT is to plot the chromosome view figures, where each figure shows the significance of each of the three single SNP effects and the sample size for all SNPs on each chromosome. Output files from EPISNP for plotting the chromosome view figures are used as the input files. No user input file is required to run EPISNP PLOT.

1.2.4 The EPINET program

The purpose of EPINET is to draw figures of epistasis networks of interactive SNPs affecting a phenotype, using the output file of significant epistasis effects from EPISNP as the input file.

1.2.5 Downloadable programs

In Version 2.0 of epiSNP, EPISNP and CPUHD programs are supported for various operation systems, including Microsoft Windows, Linux, and Unix, while EPISNP PLOT and EPINET are supported for Microsoft Windows only. All executables can be downloaded from:

<http://animalgene.umn.edu/episnp/index.html>

After downloading, the compressed file ending with “.tar.gz”, e.g., epiSNP_2.0_SGI_Irix_Mips.tar.gz, can be decompressed using the following commands:

```
gunzip epiSNP_2.0_SGI_Irix_Mips.tar.gz
tar xvf epiSNP_2.0_SGI_Irix_Mips.tar
```

The above commands produce the following files:

<i>Operation system</i>	<i>Compiler</i>	<i>Processor</i>	<i>Binary</i>
Windows	Intel	Intel/AMD	epiSNP_2.0_Windows.zip
Irix	SGI	MIPS	epiSNP_2.0_SGI_Irix_Mips.tar.gz
Linux (SUSE)	Intel	AMD	epiSNP_2.0_intel_suse_AMD.tar.gz
Linux (SUSE)	Intel	Intel (EM64T)	epiSNP_2.0_intel_suse_EM64T.tar.gz
Linux	Portland	Intel (32bit)	epiSNP_2.0_Portland_Linux_Intel.tar.gz
Linux (SUSE)	Pathscale	AMD	epiSNP_2.0_Pathscale_suse_AMD.tar.gz
Unix (AIX)	XLF	Power4	epiSNP_2.0_xlf_AIX_Power.tar.gz
Unix (Solaris)	SUN	SPARC	epiSNP_2.0_Sun_Solaris_Sparc.tar.gz

The file epiSNP_Windows.zip has four executables for Windows: EPINET, CPUHD, EPISNPLOT, and EPINET. Each of the other files has two executables: EPINET and CPUHD. Output files of EPISNP running on Linux or Unix system can be used for EPISNPLOT and EPINET on a Windows computer.

Note: To request a build of an executable on a system architecture not covered, please send an email to Dr. Yang Da at yda@umn.edu and Dr. H. Birali Runesha at runesha@msi.umn.edu.

2. INPUT FILES

2.1 Parameter file (parameter.dat)

A parameter file with the name parameter.dat is required to run EPISNPmpi. The parameter file provides various user-specified controls and must have the name 'parameter.dat'. The following is an example of the parameter.dat file:

```
3 # number of traits
4 # starting position of traits in the trait file
3 1 2 # number of column for gender information and codes
for male and female
-1 # code for missing traits, non-genetic factors and
covariables
295 # number of individuals
2 # number of non-genetic factors
2 3 # positions of the non-genetic factors
0 # number of covariables
4 5 # positions of covariables in the trait file
3 # number of chromosomes
3 # sex chromosome number
1 # 1: 0=A1A1, 1=A1A2, 2=A2A2 and others=missing used for
coding; 2: A1/A1, A1/A2 and A2/A2 used for coding, 0/0 used for
missing
4 # starting position of SNPs in the SNP data file
6 chr14.dat # number of SNPs and file name of the first
Chromosome data
10 chr19.dat # number of SNPs and file name of the second
Chromosome data
3 chr22.dat # number of SNPs and file name of the third
Chromosome data
sample_data.txt # file name of the phenotype data
10 # number of most significant results for single SNP tests
to be printed in the output file
15 # number of most significant results for pairwise tests to
be printed in the output file
```

Table 1: Example of parameter.dat

WARNINGS:

a) Adding or deleting any line except adding chromosome files creates errors and is not allowed. The user may change the parameter values but may not add or delete any line in the parameter file provided by the program. The number of lines for chromosome files must not exceed the number of chromosomes.

b) The recommended largest number of ‘most significant results’ to be saved in the output file of significant results is 1000 for single locus analysis and 10,000 for pairwise analysis. Larger numbers will result in increased computing time.

2.2 SNP genotype files (chrxxx files)

The SNP genotypes are in the chrxxx.dat files. Each chrxxx.dat file is for one chromosome. Each chromosome file must start with 'chr', and 'xxx' is a user specified chromosome number and will be used in the output files as the chromosome number. Alternatively, the SNP genotypes can be placed in one file with a maximum of 25 characters. However, the output files will not have the true chromosome numbers but rather, the file name except the first three characters is used as the chromosome name.

The coding for SNP genotypes has two options. The first option, with ‘1’ in the parameter file for the SNP genotype definitions, uses the following coding, 0 = AA, 1 = Aa, 2 = aa, i.e., '0' and '2' denote the two homozygous genotypes and '1' denotes the heterozygous genotype. A number of ‘3’ or greater denotes a missing SNP genotype.

The general format of the SNP genotype file for the first option is as follows:

```
{name of each column}
{family ID} {individual ID} {sex: 2 if female, 1 if male,
others if unknown} {locus1}{locus2}{locus3}{locus4}...
```

The second option for coding SNP genotype is that of Problem #1 of the Genetic Analysis Workshop 15 held in Tempa, Florida, November 12-15, 2006, where the three SNP genotypes of each locus are coded as A1/A1, A1/A2, and A2/A2, where A1 and A2 are nonzero integer numbers. A missing genotype is coded as 0/0.

fam_ID	ind_ID	Sex	M1	M2	M3	M4	M5	M6
1	31	1	0	2	0	1	0	0
1	32	1	1	1	0	2	0	0
1	33	1	1	2	0	2	1	0
1	34	1	2	2	0	1	0	0
2	40	1	1	2	0	2	0	0
2	41	1	0	2	0	2	0	0
2	42	1	0	2	2	2	0	0
2	43	1	1	2	0	1	0	0

Table 2: Example of a SNP genotype input file

2.3 Phenotype (quantitative trait) file

The phenotype file contains quantitative traits, fixed non-genetic effects such as gender, herd, year, season, blocks, treatment and living conditions, and covariables such as body weight and age. The name of the phenotype file is determined by the user in the parameter file.

The general format of the SNP genotype file is as follows:

```
{name of each column}
{family ID} {individual ID} {fixed effects, one fixed effect
per colum} {covariables, one covariable per column}{trait
1}{trait 2}{trait 3} . . .
```

Table 3 gives an example of a phenotype input file

The family IDs and individual IDs must match those in each chrxxx.dat file. Note that a trait can be used as a covariable. For example, birth weight can be a trait and can also be a covariable for analyzing the weight at six month old. Table 3 is an example of the phenotype file where 'sex' is a non-genetic fixed effect:

ind_ID	fam_ID	Sex	trait1	trait2	trait3
29	1	1	1.29	4.15	4.61
30	1	1	1.22	4.28	4.7
31	1	1	1.21	3.99	4.42
32	1	1	1.12	4.29	4.78
33	1	1	1.13	4.24	4.73
34	1	1	1.06	3.82	4.21
35	1	1	0.95	3.71	4.15
36	1	2	1.27	4.35	4.83
37	1	2	1.06	3.75	4.07
38	1	2	1.34	4.98	5.36
39	1	2	1.17	4.18	4.57
40	2	1	1.31	4.12	6.62
41	2	1	1.17	3.37	5.4
42	2	1	1.23	4.27	6.4
43	2	1	1.17	3.78	5.5
44	2	1	1.05	3.4	5.51
45	2	1	0.93	3.36	5.51
46	2	1	0.6	1.6	2.87
47	2	1	0.59	2.05	3.25
48	2	2	1.35	3.87	5.84
49	2	2	1.25	3.81	5.52

Table 3: Example of a phenotype input file

3 RUNNING THE PROGRAMS

3.1 Windows Operation System

3.1.1 Running the CPUHD program

It is highly recommended to run CPUHD before running EPISNP.

To run CPUHD, the cpuhd.exe and parameter.dat files must be in the same directory. The program can be run in two ways.

Method 1

- 1) Open 'mycomputer' and click the folder where cpuhd.exe program and the parameter.dat file are located.
- 2) Double click the cpuhd.exe file.

Method 2

Open the C-prompt window, and type

```
>chdir c:\epiSNP_location
```

```
>cpuhd
```

The output file is named WARNINGS.OUT and contains a summary of key statistics in the parameter file, calculation of numbers of tests, and estimates and warnings of CPU time and disk space required to finish the running of the mainprogram, episnp.exe. An example of WARNINGS.OUT is given below:

```
KEY PARAMETERS
=====
TOTAL NUMBER OF TRAITS:          11
TOTAL NUMBER OF CHROMOSOMES:    2
TOTAL NUMBER OF LOCI:           50003

NUMBER OF TESTS
=====
NUMBER OF SINGLE LOCUS TESTS:    550033
NUMBER OF PAIRWISE TESTS:       13.75138 BILLION

ESTIMATED CPU TIME AND DISK SPACE REQUIRED
=====
ESTIMATED COMPUTING TIME:        1.043190 YEARS
DISK SPACE FOR OUTPUT FILES:     10.04267 GB

WARNINGS
=====
WARNING I:   1.043190 YEARS OF CPU TIME ARE REQUIRED
WARNING II:  10.04267 GB DISK SPACE IS REQUIRED

ADVICES
=====
THE LARGEST THRESHOLD P-VALUE(P1) SHOULD BE < 1.4028286E-03
FOR THE OUTPUT FILES TO BE < 10 GB
THE LARGEST THRESHOLD P-VALUE(P1) SHOULD BE < 7.0144911E-03
FOR THE OUTPUT FILES TO BE < 50 GB
```

3.1.2 Running the EPISNP program

To run EPISNP, the episnp.exe, parameter.dat, phnotype and SNP genotype (chromosome) files must be in the same directory. The program can be run in two ways.

Method 1

- 1) Open 'mycomputer' and click the folder where episnp.exe and its input files are located
- 2) Double click the episnp.exe file.

Method 2

Open the C-prompt window, and type

```
>chdir c:\epiSNP_location
```

```
>episnp
```

The EPISNP produces several output files, including two summary files, three files of significant pairwise epistasis results corresponding to three user specified p-values, three files of significant single SNP effects corresponding to three user specified p-values, and one file for plotting the chromosome view of significance and sample size for single locus analysis using EPISNPLOT.

3.2.1.1 Screen output

Screen output is designed to monitor the statuses of the execution of EPISNPmpi. Version 1.0 of EPISNPmpi tested for 128 processors is shown in Table 4.

```
Pairwise analysis ...
Single locus analysis ...
Pairwise analysis ...
Number of tests =          171
Number of tests =          171
      1 Number of tests per processor =          86
      0 Number of tests per processor =          85
The CPU time for running is 3.2002000000000002E-002
seconds
Cleaning up all processes ...
done.
```

Table 4: Screen output of EPISNPmpi Version 1.0

3.2.1.2 Output files of test results

The results of statistical tests for single-locus effects are saved in the output file named “single_locus_sig.out”. Single-locus effects include the marker genotypic effect, and additive and dominant effects. Table 5 shows an example of the single-locus test results in “single_locus_sig.out”.

7 MOST SIGNIFICANT RESULTS OF SINGLE LOCUS ANALYSIS				
=====				
m = overall marker effect				
a = additive effect, d = dominance effect				

Chr	Locus	Trait	Test	P_value

01	rs2017	trait1	d	0.113E+00
	D	12 21	11	
	Estimate	4.132 -3.487	-5.273	
01	rs2017	trait1	m	0.243E+00
01	rs2840	trait1	a	0.291E+00
	A	2 1		
	Estimate	0.604 -4.655		
01	rs2477	trait1	a	0.121E+00
	A	2 1		
	Estimate	4.725 -1.968		
01	rs7703	trait1	m	0.202E+00
01	rs7703	trait1	d	0.371E+00
	D	12 11	21	
	Estimate	2.585 -0.559	-6.230	
01	rs4999	trait1	a	0.243E+00
	A	1 2		
	Estimate	2.439 -1.912		

Table 5: Example of single_locus_sig.out

The results of statistical tests for pairwise epistasis effects are saved in the output file named “pairwise_sig.out”. Pairwise epistasis effects include the two-locus interaction effects, additive × additive, additive × dominance, dominance × additive, and dominance × dominance effects. Table 6 shows an example of the pairwise test results in pairwise_sig.out.

5 MOST SIGNIFICANT RESULTS OF PAIRWISE EPISTASIS ANALYSIS										
Chr1	SNP1	Chr2	SNP2				Trait		Effect	p-value
5	M7	19	M12				Trait1		I	0.345E-06
5	M7	19	M12				Trait1		AA ^a	0.793E-07
	i_k ^a	2_1	1_2	1_1	2_2					
	Estimate	0.388	0.314	-0.305	-0.365					
8	M8	22	M93				Trait1		AD ^b	0.158E-07
	i_kl ^b	1_11	2_12	1_22	2_22	1_12	2_11			
	Estimate	2.335	1.049	0.238	-0.303	-0.868	-3.937			
9	M9	14	M612				Trait2		DA ^c	0.383E-08
	ij_k ^c	11_1	13_1	12_2	22_2	11_2	12_1			
	Estimate	4.892	3.344	0.874	-0.259	-1.871	-2.074			
19	M112	23	M182				Trait3		DD ^d	0.793E-07
	ij_kl ^d	22_11	11_11	11_22	12_12	22_22	12_22	22_12	11_12	12_11
	Estimate	2.969	1.839	1.719	1.543	0.616	-0.374	-0.691	-1.446	-2.118

Table 6: Example of pairwise_sig.out

The second output file of single-locus tests is used as the input file of the EPISNPLOT program to draw figures of chromosome view of significant results, and the second output file of pairwise epistasis tests is used as the input file of the EPINET program to draw figures of epistasis network. The EPISNPLOT and EPINET are compiled for *Windows* and are distributed as the epiSNP package, which is freely available at <http://animalgene.umn.edu/episnp/index.html>.

3.1.3 Running the EPISNPLOT program

To run EPISNPLOT, the episnpplot.exe and output files from episnp.exe must be in the same directory. Then, open the C-prompt window, and type

```
>chdir c:\epiSNP_location
>episnpplot -f input_file_name
```

The -f switch is actually optional; if it is omitted (i.e., typing simply ">episnpplot") then EPISNPLOT will attempt to find an input file called "episnpplot.dat", and notify you if it cannot find the file. Note that by default, the

EPISNP output file which EPISNPLOT uses as input is called “figure_data1.out”, so the usual command-line invocation will be:

```
>epispnpplot -f figure_data1.out
```

Figure 1 is an example of the figures produced by the above command. Other command-line switches besides `-f` are available to offer options for plotting figures. The user may choose to plot only specific traits and/or chromosomes with the `-t` and `-c` switches. For example, if you type the following:

```
>epispnpplot -f input_file_name -t TRAIT2 -c 14
```

EPISNPLOT will generate only one plot file, that showing markers on Chromosome 14 for Trait TRAIT2. (By default, EPISNPLOT generate one .eps plot file for each chromosome-trait combination in the input file.) The `-t` and `-c` switches may be used independently: “>epispnpplot -f input_file_name -c 14” will create one plot file for each trait on Chromosome 14, while “>epispnpplot -f input_file_name -t TRAIT2” will create one plot file for TRAIT2 on each chromosome.

Chromosome names are a special case in that if they have a trailing “.dat” in their names in the input file, this will be removed by EPISNPLOT for internal use (The input file is not modified). For example, if the input file contains the chromosome name “chr14.dat”, then to plot this chromosome you would use the `-c chr14` switch, not `-c chr14.dat`.

The user may also choose to leave the markers in the order they appear in the input file, or may sort them with the `-s` switch, where valid sort options are ‘m’ for overall marker significance, ‘a’ for additive significance, ‘d’ for dominance significance, and ‘n’ for number of observations. For example, if you type:

```
>epispnpplot -f input_file_name -s d
```

EPISNPLOT will, for each chromosome-trait combination, sort the markers in order of ascending dominance significance. **Figure 2** is an example of the figure produced by the ‘-s a’ command while **Figure 3** is an example by the ‘-s d’ command for a different data set with much denser marker coverage.

EPISNPLOT’s output files are named as follows:

```
input_file_name.chromosome_name.trait_name.sort_type.eps.
```

For example, the file “epispnpplot.dat.14.TRAIT2.sm.eps” is the plot for data from the file “epispnpplot.dat” for markers on Chromosome 14 for Trait TRAIT2, sorted by overall marker significance. If sort_type is “s0”, this means the markers are unsorted; i.e., they are shown on the plot in the order they appear in the input file.

Finally, if the number of markers in a given chromosome-trait combination is greater than 21, then not all markers will be denoted by labels and hashmarks along the

bottom axis. Furthermore, these markers' values will be plotted as part of the line on the plot rather than having plot symbols (squares for significance levels, circles for number of observations.) If this is the case, then the optimum number of markers (between 11 and 21 inclusive) will have labels and hashmarks on the bottom axis to make interpretation of the plot lines as straightforward as possible. You can observe this in the following figures, which are examples of EPISNPLOT output.

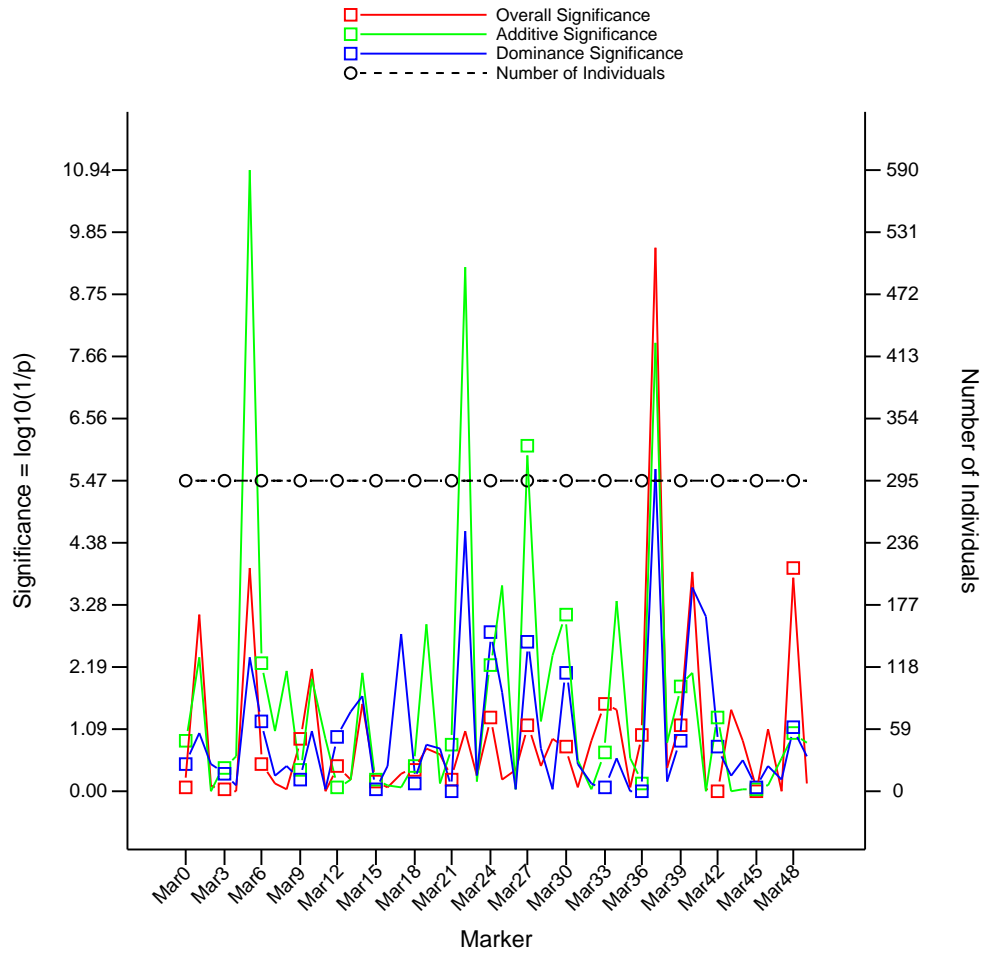


Figure 1. Example of EPISNPLOT output file based on the original marker order from the input file.

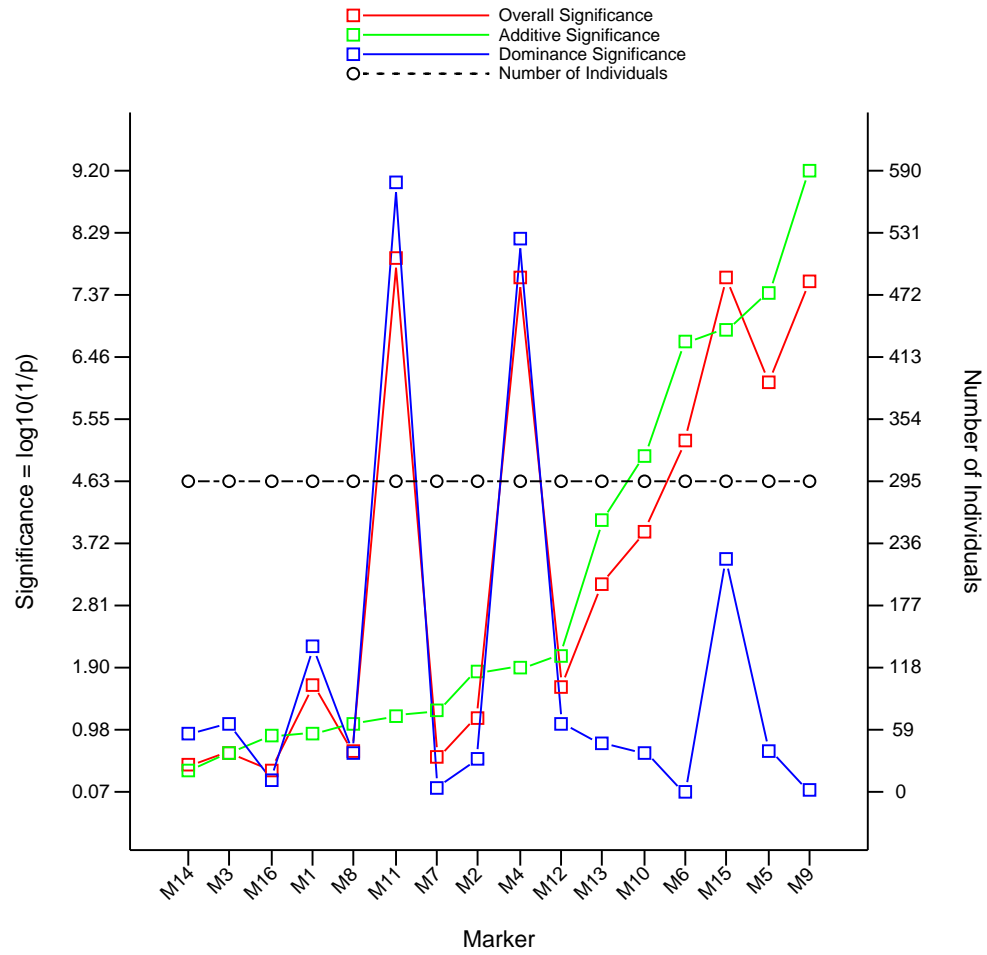


Figure 2. Example of EPISNPLOT output file sorted in ascending order of significant additive effects.

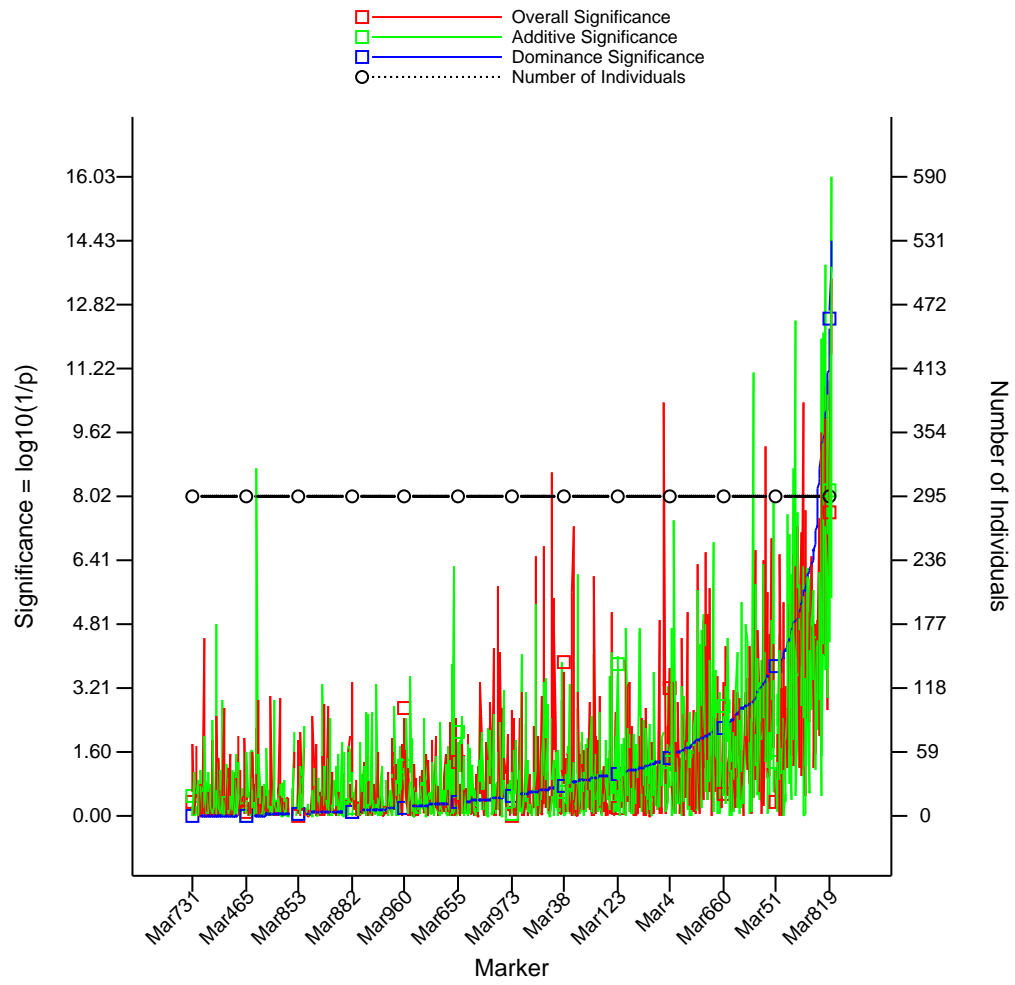


Figure 3. Example of EPISNPLOT output file sorted in ascending order of significant dominance effects.

3.1.4 Running the EPINET program

The program is run from the command line. The program requires two input files: the parameter file to specify four significance levels (p values) for selecting loci in the epistasis networks, and the effect file that contains epistasis testing results from the EPISNP program. The default input is to use 'effects.dat' as the input file and to print the 10 largest networks. Optionally, the user can specify the file name on the command line. If the input file is specified, the number of networks to print can also be specified, for example, the following command

```
epiNet.exe inputfile.dat 5 or epiNet inputfile.dat 5
```

will produce 5 epistasis networks using inputfile.dat as the input file. **Figure 4** shows two examples of such epistasis networks.

The input file is the same format as the output file of the epiSNP program. Each line describes an interaction. Each line has 7 items:

- first chromosome number
- first marker name
- second chromosome number
- second marker name
- trait name
- effect type (I, AA, AD, DA, or DD)
- threshold p-value to declare significance.

To use user-specified threshold p-values for establishing epistasis network, four threshold p-values can be specified in a parameter file named epiNet.par. If this file is present, the four p values will be read from the file. The p values should be specified in order from largest to smallest. The p values can be specified in regular decimal notation, or in scientific notation in the form 2.3E-5 or 1E-8, etc.

Two output files are generated for each network, a png file and a postscript (.ps) file. The files are named 1.png, 2.png, 3.png, etc in descending order of network size. In the graph of a network a node is drawn for each marker with the marker name and chromosome number shown. The color of the marker node indicates the significance of the most significant effect for that marker.

- $p < 1E-8 = \text{red}$
- $p < 1E-7 = \text{blue}$
- $p < 1E-6 = \text{green}$
- $p < 1E-5 = \text{white}$

Lines connecting markers indicate significant effects. The color of the lines indicates the type of effect.

I = black
 AA = red
 AD = purple
 DA = blue
 DD = green

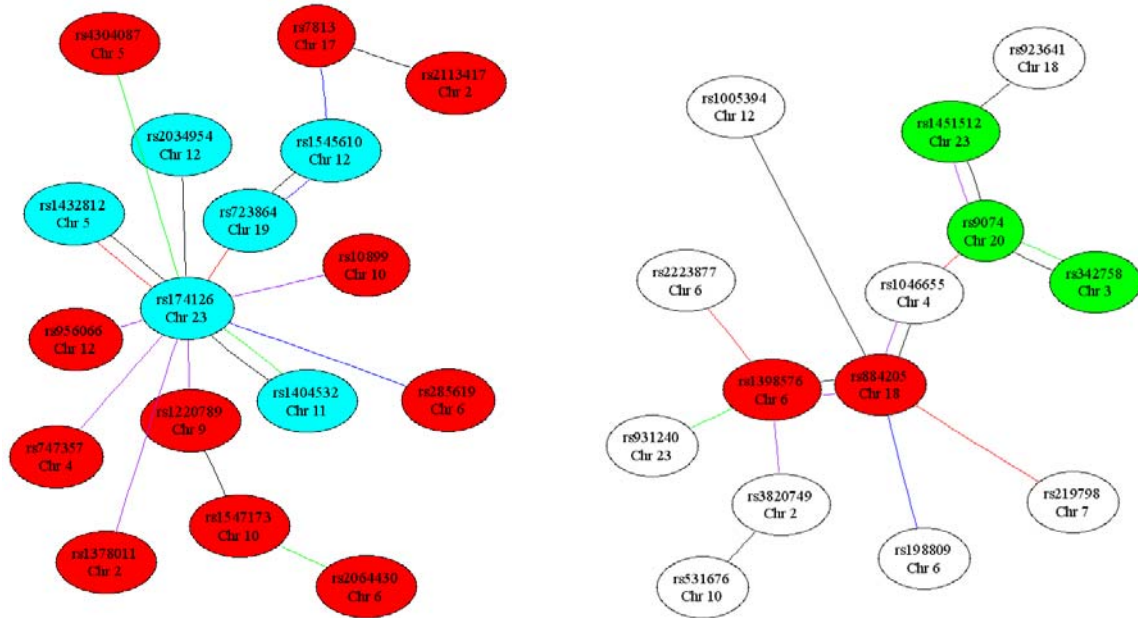


Figure 4. Examples of potential SNP epistasis network of a phenotype.

Line color: black = I effect, red = A×A, purple = A×D, blue = D×A, green = D×D. A pair of SNPs connected by two lines had two significant epistasis effects. Node color: red: $p < 10^{-8}$, cyan: $p < 10^{-7}$, green: $p < 10^{-6}$, white: $p < 10^{-5}$.

3.2 Linux and Unix Operation Systems

The commands for uncompress the .tar.gz files and to run the programs under Linux and Unix operating systems are the same for all platforms. For example, for the intel compiler with AMD processor, use the following commands to uncompress,

```
gunzip epiSNP_2.0_intel_suse_AMD.tar.gz  
untar the file: tar xvf epiSNP_2.0_intel_suse_AMD.tar
```

The command to run EPISNP is

```
./EPISNP
```

And the command to run CPUHD is

```
./CPUHD
```

To uncompress the .tar.gz file for a different platform, just use the file name corresponding to the platform.

4 REFERENCES

Yongcai Mao, Nicole R. London, Li Ma, Daniel Dvorkin and Yang Da (2007) Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. *Physiological Genomics* 28: 46-52.