

This tutorial supplements the manuscript submitted to Plant Journal November 13, 2007:

***Toward systems genetic analyses in barley: Integration of phenotypic, expression and genotype data into GeneNetwork***

Arnis Druka, Ilze Druka, Arthur G. Centeno, Hongqiang Li, Zhaohui Sun, William T. B. Thomas, Nicola Bonar, Brian J. Steffenson, Steven E. Ullrich, Andris Kleinhofs, Roger P. Wise, Timothy J. Close, Elena Potokina, Zewei Luo, Carola Wagner, Günther F. Schweizer, David F. Marshall, Michael J. Kearsey, Robert W. Williams and Robbie Waugh

**The tutorial demonstrates how to:**

- 1) query GeneNetwork for barley related information;
- 2) map barley mRNA abundance and higher-order traits;
- 3) associate different traits using correlation analysis;
- 4) enter your own data;
- 5) map genes and Mendelian (binary) traits.

As an example, one of the lignin pathway genes, cinnamyl alcohol dehydrogenase (CAD) will be used throughout the tutorial.

Please open the Bookmark panel in the Acrobat Reader to navigate through the tutorial.

**[http://barleygenetics.net/GN\\_barley\\_tutorial.html](http://barleygenetics.net/GN_barley_tutorial.html)**

## How to query GeneNetwork for barley related information?

1) go to the [www.genenetwork.org](http://www.genenetwork.org)

The screenshot shows the GeneNetwork website search interface. The browser address bar shows <http://www.genenetwork.org/>. The page title is "The GeneNetwork" and the subtitle is "From the University of Tennessee: www.genenetwork.org". The navigation menu includes Home, Search, Help, News, References, Policies, Accounts, and Links. The main content area is titled "Find Records" and contains a search form with the following fields and options:

- Choose Species: Barley
- Group: SXM
- Type: Phenotypes
- Database: SXM Published Phenotypes
- ANY: cinnamyl
- ALL: (empty)

Buttons for Search, Advanced Search, and Set To Default are visible. A footer section contains information about the service and funding sources.

2) select 'Barley'

3) select the data set type 'Phenotypes' - to query higher order traits; 'Embryo mRNA' or 'Leaf' mRNA - to query genes.

4) select the database. INFO file describes available options

5) write or copy&paste your query in the ANY and/or ALL field. Queries containing multiple strings should be space separated.

To look up the contents (all entries) of the 'Phenotypes' database type \* .

For the mRNA databases query string can contain Barley1 GeneChip probe set ID (eg Contig20\_at), or part of the gene annotation text (eg histone) or multiple space delimited strings.

Type 'cinnamyl' and click the 'Search' button

### Difference between the fields ANY and ALL

ANY will find all the entries that match to any entered space separated string. For example, 'cinnamyl dehydrogenase' typed in ANY will identify 254 records. If 'cinnamyl dehydrogenase' is queried from the ALL field, only five records are retrieved. ALL looks for the records where all query strings are present.

## Search Results

GeneNetwork searched the **Barley1 Embryo gcrMA SCRI (Dec06) Database** for all records that match the term cinnamyl.

GeneNetwork found a total of **5** records. To study any one of these records, click on its text below. To add one or more record to your Selection window, use the checkbox and then click the **Add to Collection** button.

Sort By

1.  **ProbeSet/Contig19854\_at: cinnamyl alcohol dehydrogenase**
2.  **ProbeSet/HVSMEh0081I20r2\_x\_at: cinnamyl alcohol dehydrogenase**
3.  **ProbeSet/EBpi07\_SQ002\_J15\_at: Cinnamyl-alcohol dehydrogenase**
4.  **ProbeSet/Contig4346\_at: Cinnamyl-alcohol dehydrogenase**
5.  **ProbeSet/X92754\_at: Cinnamyl-alcohol dehydrogenase**

**There are two analysis options to choose:**

### 1) Multiple trait analysis

By comparing side by side QTL profiles of these 5 probe sets it may be possible to evaluate their performance (or see if there is a QTL at all) and may be determine what they actually represent - a single gene or different members of the family ?

1) click 'Select All';

2) then click 'Add to Collection'.

It will open 'SXM Trait Collection' window, which has several functions for multiple trait analysis.

### 2) Single trait analysis

If the best probe set for the query is known, or intention is to inspect the returned results individually, clicking directly on the record will open the 'Trait Data and Analysis form'.

***Collection** allows to create combined list of traits from several independent searches (mRNA abundance and higher order traits), so that they can be analyzed together.*

### SXM Trait Collection

Sort By

- ProbeSet/EBpi07\_SQ002\_J15\_at: Cinnamyl-alcohol dehydrogenase** --- FROM : **Barley1 Embryo gcrMA SCRI (Dec06) Database**
- ProbeSet/HVSMEh0081I20r2\_x\_at: cinnamyl alcohol dehydrogenase** --- FROM : **Barley1 Embryo gcrMA SCRI (Dec06) Database**
- ProbeSet/Contig4346\_at: Cinnamyl-alcohol dehydrogenase** --- FROM : **Barley1 Embryo gcrMA SCRI (Dec06) Database**
- ProbeSet/Contig19854\_at: cinnamyl alcohol dehydrogenase** --- FROM : **Barley1 Embryo gcrMA SCRI (Dec06) Database**
- ProbeSet/X92754\_at: Cinnamyl-alcohol dehydrogenase** --- FROM : **Barley1 Embryo gcrMA SCRI (Dec06) Database**

These functions are for multiple trait analysis

A quick way to obtain an overview on the QTLs for multiple traits is to use 'QTL Cluster Map' function:

- 1) click 'Select All';
- 2) click 'QTL Cluster Map' .

## QTL Cluster Map

### Cluster Tree

The upper part of the QTL Cluster output is a hierarchical cluster tree of the set of traits that were selected in the previous window. To generate this plot, distances between pairs of traits using  $(1 - r)$  where  $r$  is the Pearson product-moment correlation were computed. The hierarchy is assembled by successively linking traits and groups of traits.

### QTL heat map

The QTL heat map of all members of the Cluster Tree, extending from proximal Chr 1H at the top to distal Chr 7H at the bottom is generated below the tree. Each vertical column or stripe encodes the genome-wide p-value computed on the basis of 1000 permutations.

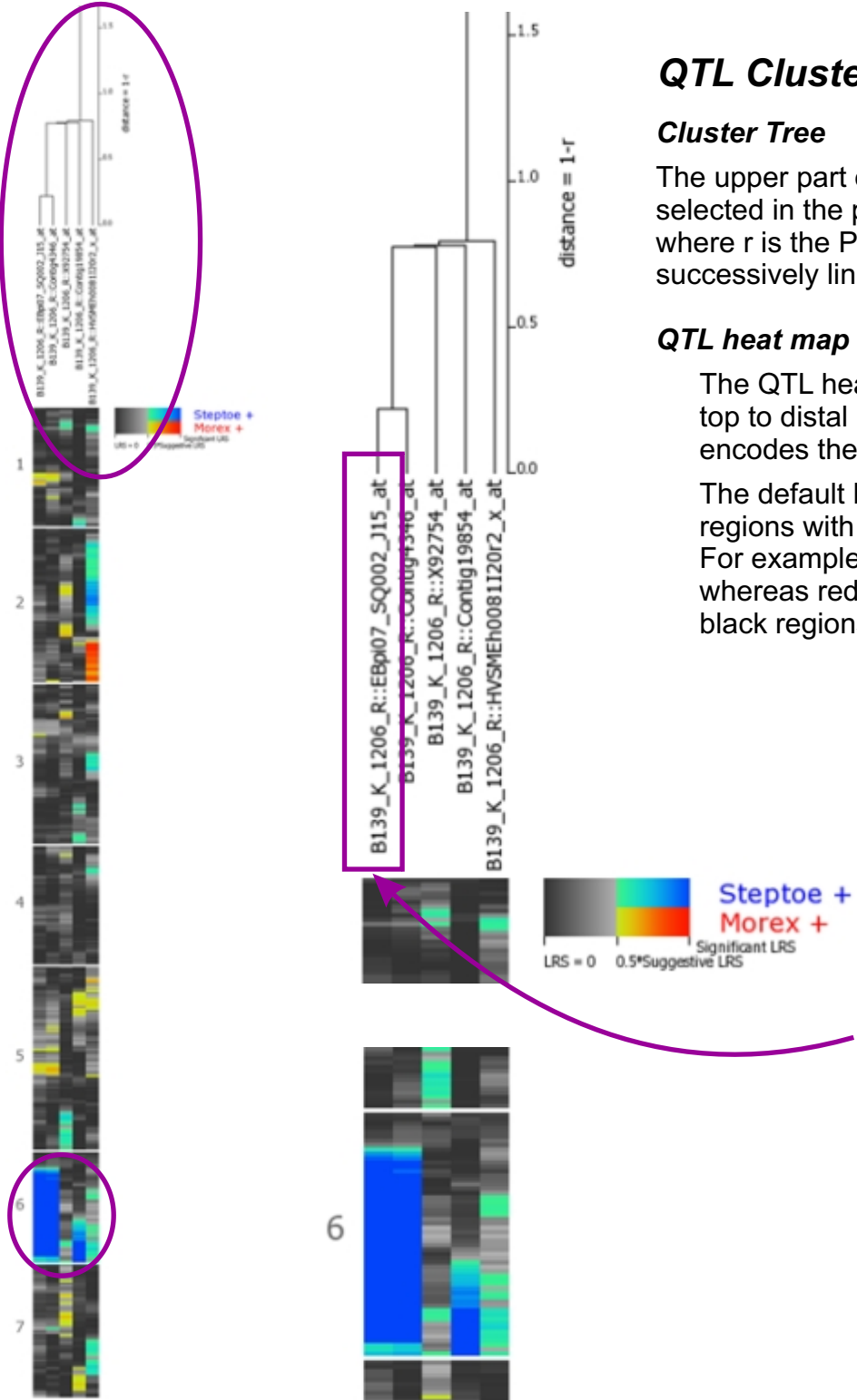
The default heat map is "Grey + Blue + Red" in which more intense colors mark chromosomal regions with comparatively high linkage statistics and the spectrum encodes the allelic effect. For example, blue regions are those in which Steptoe is associated with higher trait values, whereas red regions are those in which Morex is associated with higher trait values. Grey and black regions have insignificant linkage to trait variance.

Strong QTL on chromosome 6H is supported by two of the probe sets, while others either have single, weaker QTLs at different locations or have none at all. Therefore, the first two can be prioritized for further investigation.

However, thorough, independent, bioinformatics-based analysis or even lab work may be required to determine the cause of differential performance of the probe-sets and to decide which one is the best representative of the CAD gene.

Here, it is assumed that the first one is the best.

**To proceed, click the trait ID, it will open the 'Trait Data and Analysis Form'.**



## Trait Data and Analysis Form

**Trait ID** EBpi07\_SQ002\_J15\_at from [Barley1 Embryo gcRMA SCRI \(Dec06\)](#)

**Gene Symbol:** *Not available*  
**Description:** Cinnamyl-alcohol dehydrogenase  
**Location:** Not available [Info](#)

### Analysis Tools:

To analyze a trait, select appropriate options and one or more function buttons (Basic Statistics, Trait Correlations, Pair-Scan, etc.). New windows will open to display results and provide you access to a series of additional analysis tools. To review and edit data, scroll down to the Trait Data section.

Basic Statistics

Probe Tool

Add to Collection

Reset

Trait Correlations

Trait Correlations compares the values listed below with those of all other records in the database that you select to the right. You can edit values before initiating the analysis.

**Choose Database:**

Barley1 Embryo gcRMA SCRI (Dec06)

**Calculate:**

Pearson's Product-Moment

**Return:**

top 500

Interval Mapping

Interval Mapping computes linkage maps for the entire genome or single chromosomes. Use permutation and bootstrap tests to assess strength and consistency of linkage for single traits.

**Chromosome:**

All

**Options:**

- Permutation test  
 Bootstrap test

Marker Regression

Marker Regression plots permutation results, lists those markers linked to trait variation, and provides access to composite mapping functions.

Display LRS greater than

OR display all LRS

These are the functions to map QTLs more precisely  
**Click 'Interval Mapping' button**



## Profile of the genome-wide QTL scan obtained using the 'Interval Mapping' function.

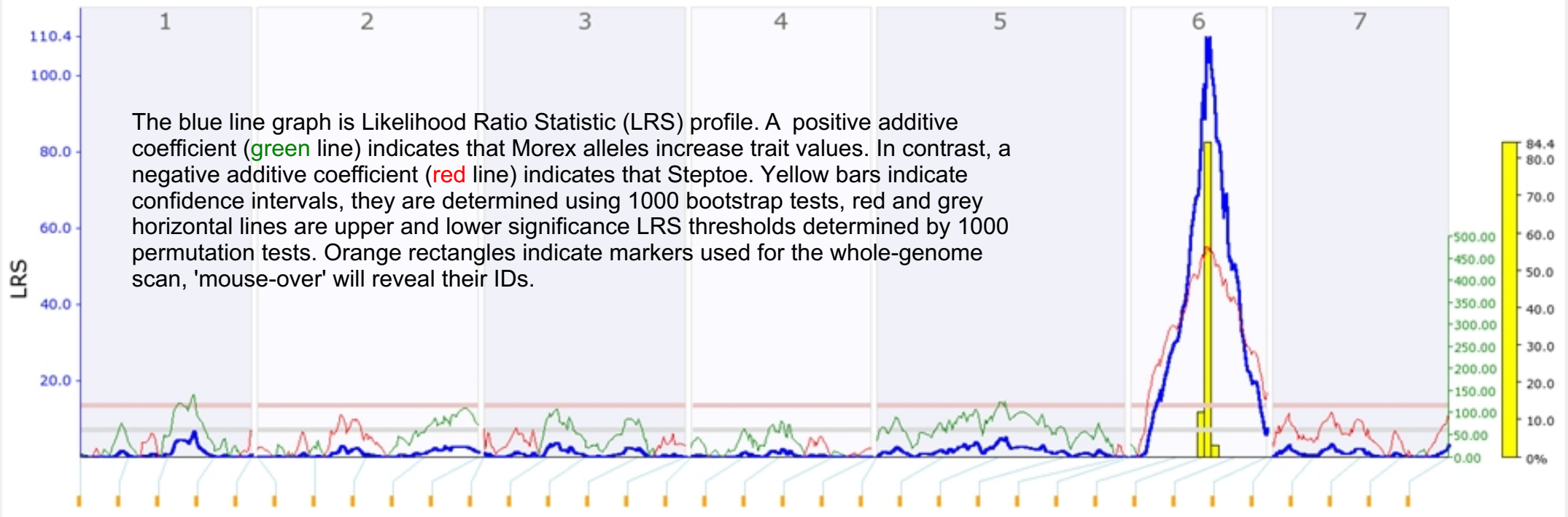
### Map Viewer: Whole Genome

Population: Barley EXM  
Database: Barley1 Embryo gcRNA SCRI (Dec06)  
Trait ID: EBpi07\_SQ002\_115\_at  
Gene Symbol: None

[Download](#) results in tab-delimited text format.

Chr:     Permutation Test   
 Bootstrap Test   
 Allele Effects   
 Legend  
 ZX Plot

View:  to   
Units:     
 units on Y-axis (0 for default)  
Width:  pixels (minimum=1400)  
\* only apply to single chromosome physical mapping



This graph reveals the following information about the CAD gene:

- 1) number of significant eQTLs (single in this case);
- 2) location of the eQTL (chromosome 6H). Scanning only chromosome 6H will reveal more precise location of the QTL peak;
- 3) allelic effect (Steptoe);
- 4) significance level (LRS=110, high).

Based on 1) and 4), it can be concluded that the CAD gene itself also maps to the same location on chromosome 6H.

**Next:**

- 1) minimize this window;
- 2) go back to the 'Trait Data and Analysis form'

## Trait Correlations

Which higher order traits correlate to CAD mRNA abundance? Can any suggestive associations be revealed?

### Trait Data and Analysis Form

**Trait ID** EBpi07\_SQ002\_J15\_at from [Barley1 Embryo gcRNA SCRI \(Dec06\)](#)

**Gene Symbol:** *Not available*  
**Description:** Cinnamyl-alcohol dehydrogenase  
**Location:** Not available [Info](#)

#### Analysis Tools:

To analyze a trait, select appropriate options and one or more function buttons (Basic Statistics, Trait Correlations, Pair-Scan, etc.). New windows will open to display results and provide you access to a series of additional analysis tools. To review and edit data, scroll down to the Trait Data section.

Basic Statistics

Probe Tool

Add to Collection

Reset

Trait Correlations

Trait Correlations compares the values listed below with those of all other records in the database that you select to the right. You can edit values before initiating the analysis.

**Choose Database:**

Barley SMP Published Phenotypes

**Calculate:**

Pearson's Product-Moment

**Return:**

top 500

1) select 'Barley SXM Published Phenotypes'

2) click 'Trait Correlations'



## Results of the correlation analysis

### Correlation Table

Values of Record EBpi07\_SQ002\_J15\_at in the [Barley1 Embryo gcrMA SCRI \(Dec06\)](#) database were compared to all 211 records in the [Barley SMP Published Phenotypes](#) database. The top 211 correlations ranked by the Pearson's product-moment correlation are displayed. You can resort this list using the small arrowheads in the top row.

Clicking on the record ID will open the published phenotype data for that publication. Click on the correlation to see a scatter plot of the trait data.

Multiple Mapping

QTL Cluster Map

Download Table

Select All

Invert

Clear

Add to Collection

Display strain names in correlation plot

Display fit line in correlation plot

	Record ID	Phenotype	Authors	Year	Correlation	N Cases	p Value
1 <input type="checkbox"/>	49890	Emergence of the second leaf - single leaf frequency (SCRI glasshouse 2004)	<i>Druka A</i>	2006	-0.4046	139	5.59e-07
2 <input type="checkbox"/>	49951	Grain width F4 (Okayama 2006)	<i>Druka A</i>	2006	-0.3497	139	2.06e-05
3 <input type="checkbox"/>	49877	"Lodging (Outlook, Saskatchewan 1992)"	<i>Druka A</i>	2006	0.3443	139	2.83e-05
4 <input type="checkbox"/>	49784	"Yield - MT/ha (Aberdeen, Idaho 1991)"	<i>Druka A</i>	2006	-0.3223	139	9.72e-05
5 <input type="checkbox"/>	49932	Grain width F7 (SCRI 2003)	<i>Druka A</i>	2006	0.3189	139	0.00012
6 <input type="checkbox"/>	49768	Thousand grain weight (SCRI 2005)	<i>Druka A</i>	2006	0.3162	139	0.00013

In this particular example, one of the top correlates Lodging seems to be suggestive CAD response trait. One can interpret that variation in the CAD mRNA may cause differential enzyme level and consecutively variation in the lignin composition (CAD is lignin biosynthetic gene). Lignins, as a constituents of the plant secondary cell walls can determine mechanical properties of the plant organs. Conceivably, lodging may depend on the cell wall mechanical properties determined by the lignins.

Correlation between lodging trait values and CAD mRNA is significant, but not very strong. This suggests that other factors (environmental or genetic) can determine lodging. Mapping of the lodging trait may provide clues on what they are.

**Click on the Lodging Record ID.**

## Interval Mapping of the higher order traits

**Trait Data and Analysis Form**

Trait ID 49877 from [Barley SMP Published Phenotypes](#)

**Phenotype:** "Lodging (Outlook, Saskatchewan 1992)"  
**Authors:** Druka A  
**Title:** Confidential

**Analysis Tools:**

To analyze a trait, select appropriate options and one or more function buttons (Basic Statistics, Trait Correlations, Pair-Scan, etc.). New windows will open to display results and provide you access to a series of additional analysis tools. To review and edit data, scroll down to the Trait Data section.

[Basic Statistics](#) [Add to Collection](#) [Reset](#)

---

[Trait Correlations](#)

Choose Database:   
Calculate:  Return:

Trait Correlations compares the values listed below with those of all other records in the database that you select to the right. You can edit values before initiating the analysis.

---

[Interval Mapping](#)

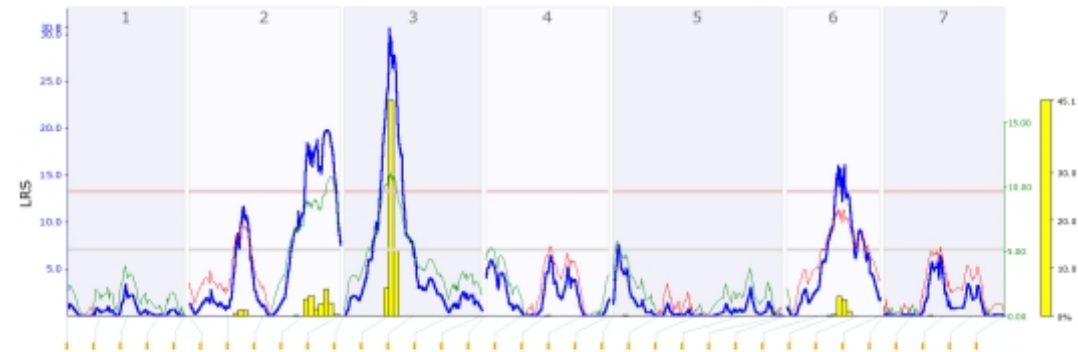
Interval Mapping computes linkage maps for the entire genome or single chromosomes. Use permutation and bootstrap tests to assess strength and consistency of linkage for single traits.

Chromosome:   
Options:  
 Permutation test  
 Bootstrap test

[Marker Regression](#)

Marker Regression plots permutation results, lists those markers linked to trait variation, and provides access to composite mapping functions.

Display LRS greater than   
OR display all LRS



### Mapping of the lodging trait provides the following information:

- 1) lodging locus on chr 6H coincides with that of CAD mRNA abundance;
- 2) they both have the same allelic effects;
- 3) major lodging QTL is located on chromosome 3H, and there are at least two additional loci on chromosome 2H.

*The CAD mRNA has a strong QTL on chromosome 6H. What additional loci can be revealed if major CAD mRNA QTL is considered as a background and only residual variation used for mapping?*

The **Composite Interval Mapping** function of the GeneNetwork can do this.

**Open the CAD Trait Data and Analysis Form that was minimized before.**

## Composite Interval Mapping

**Trait Data and Analysis Form**

Trait ID EBpi07\_SQ002\_J15\_at from Barley1 Embryo gcrMA SCRI (Dec06)

Gene Symbol: Not available  
Description: Cinnamyl-alcohol dehydrogenase  
Location: Not available [Info](#)

**Analysis Tools:**

To analyze a trait, select appropriate options and one or more function buttons (Basic Statistics, Trait Correlations, Pair-Scan, etc.). New windows will open to display results and provide you access to a series of additional analysis tools. To review and edit data, scroll down to the Trait Data section.

[Basic Statistics](#) [Probe Tool](#) [Add to Collection](#) [Reset](#)

---

[Trait Correlations](#)

Trait Correlations compares the values listed below with those of all other records in the database that you select to the right. You can edit values before initiating the analysis.

Choose Database: Barley1 Embryo gcrMA SCRI (Dec06)

Calculate: Pearson's Product-Moment  Return: top 500

---

[Interval Mapping](#)

Interval Mapping computes linkage maps for the entire genome or single chromosomes. Use permutation and bootstrap tests to assess strength and consistency of linkage for single traits.

Chromosome: All

Options:  
 Permutation test  
 Bootstrap test

[Marker Regression](#)

Marker Regression plots permutation results, lists those markers linked to trait variation, and provides access to composite mapping functions.

Display LRS greater than   
OR display all LRS

Click 'Marker Regression' button.  
This will automatically identify the best associated with CAD mRNA abundance marker.

Part of the screen that comes up after hitting 'Marker Regression' button

**Composite Analysis**

Choose a locus as control background from the following menu for the composite analysis

Choose background: snp\_0003

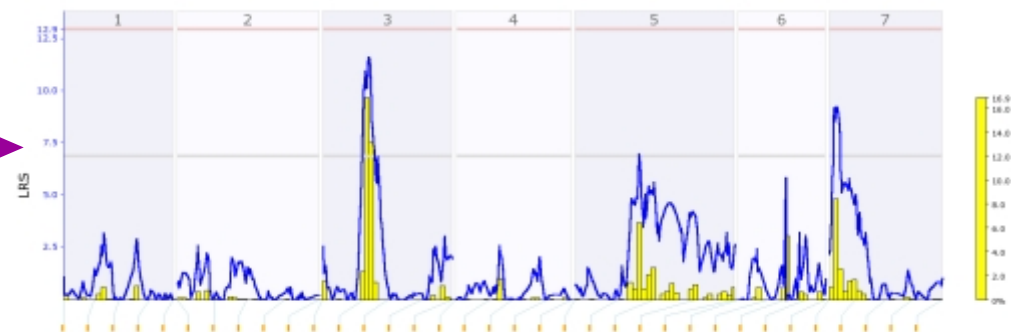
[Composite Regression](#)

Chr: All  [Composite Interval Mapping](#)

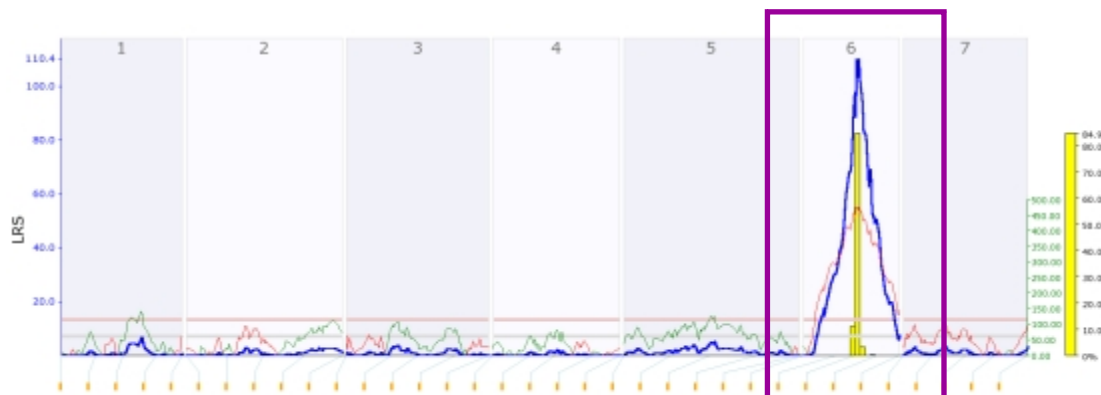
Permutation test  
 Bootstrap test

### Conclusion

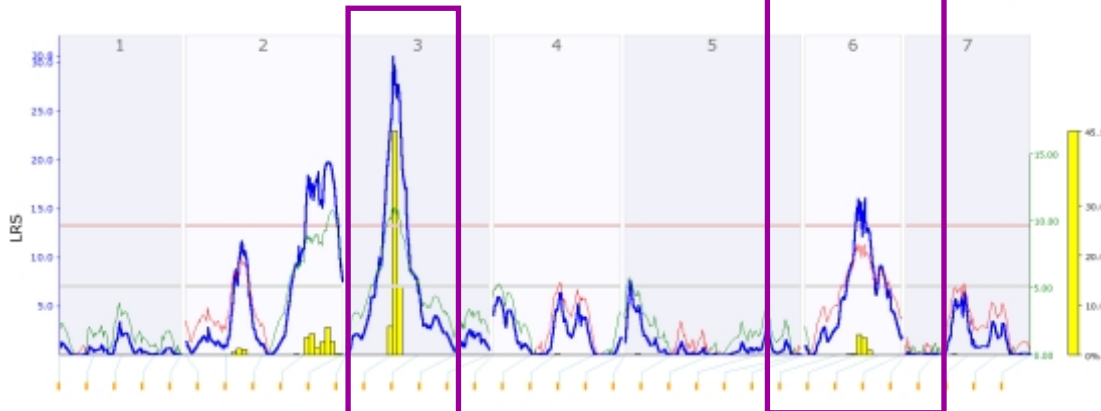
The mRNA accumulation of the CAD gene also seems to be controlled by the locus on chromosome 3H



## Summary of the mapping results

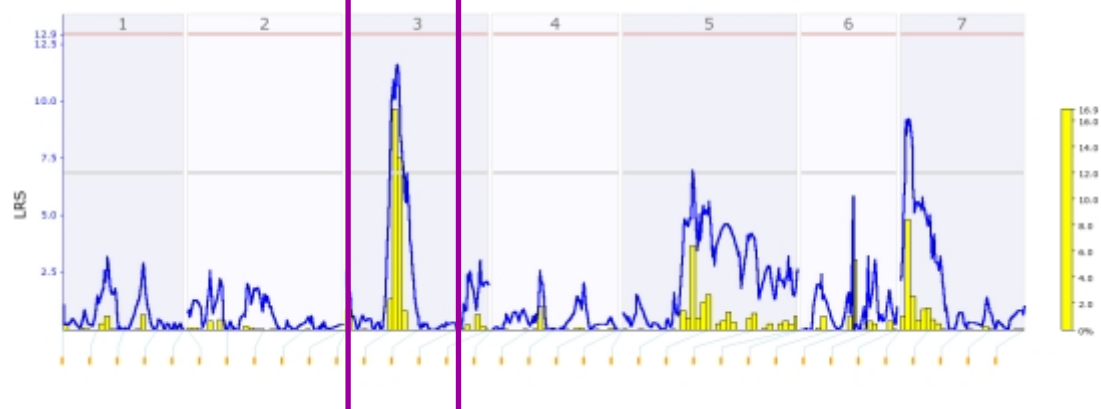


CAD mRNA cisQTL



lodging

Identification of the overlapping QTLs for different traits may suggest a common regulatory mechanism



CAD mRNA transQTL

## How to enter your own data set into the GeneNetwork?

### Format of the file to be entered into the GeneNetwork for analysis

In case this matrix is transposed, replace 'column' with 'row'

SM DHL IDs - original numbering (1-200) preceded by the 'SM'. Three digit number has to be used.

Numerical trait values

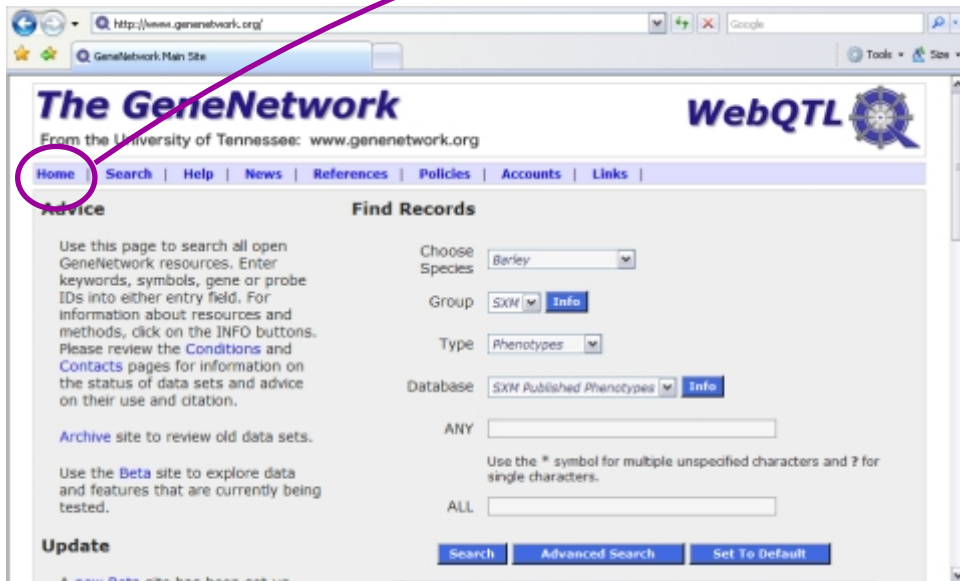
	@format=column	FP_95_Fargo_PD1	FP_95_Fargo_PD2	FP_95_Langdon_PD2
SM001		28	20	2
SM002		30	20	6
SM003		40	x	6
SM004		32	14	6
SM005		x	x	x
SM006		10	10	6
SM007		x	x	0
SM009		12	8	2
SM010		0	22	12
SM011		0	44	4
SM012		20	20	0
SM013		20	20	20
SM014		26	12	10
SM015		16	16	4

Trait IDs

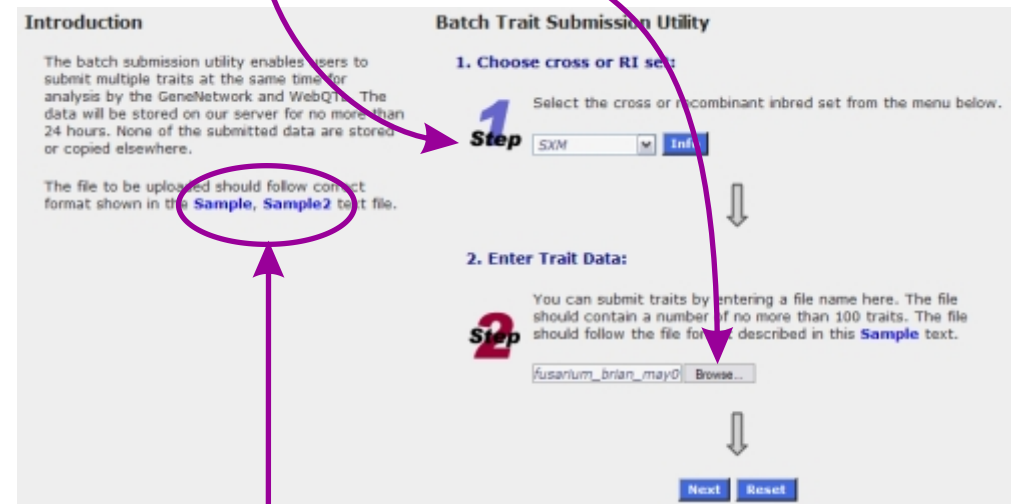
missing value



## How to enter your own data set into the GeneNetwork?



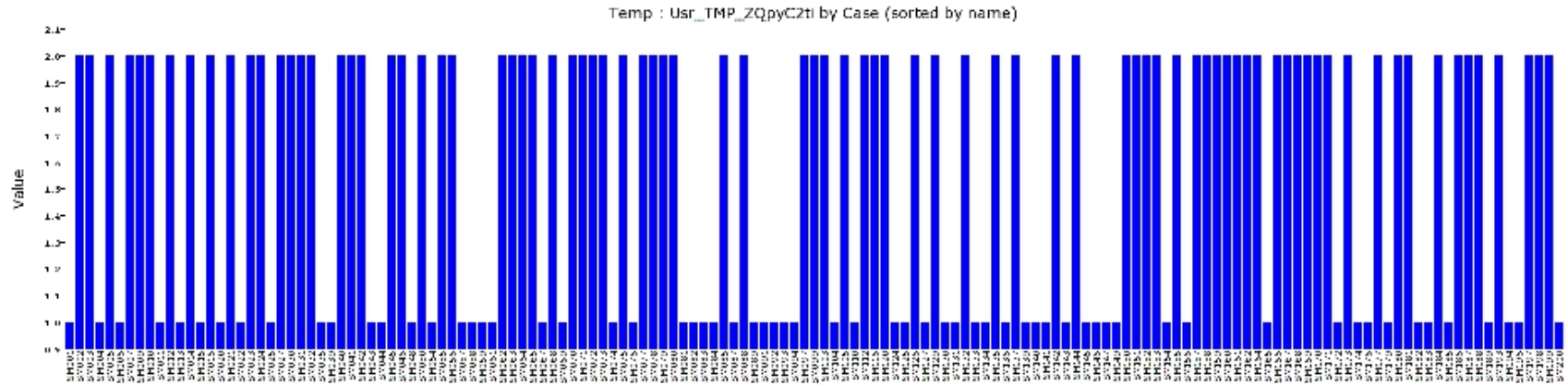
Batch Submission  
Select SXM  
Type in or browse to your data file



Describes the format of the data file

# Mapping genes or Mendelian (binary) traits using GeneNetwork

Output generated by using 'Basic Statistics' function with genotype scores (Steptoe '1' and Morex '2') of one of the markers.



1) convert genotype scores or trait values to binary numerical values, eg 1,2; -1,1 or 0,1;

2) format as shown in the previous page;

3) enter the data into the GeneNetwork;

4) submit to the 'Marker Regression' to identify map position.



**Marker Regression Report**

The following loci in the SXM data set have associations with the above trait data.

LRS	Chromosome	Locus	Additive Effect
162.371*	2	snp_1061	-0.410
192.512*	2	snp_1054	-0.429
214.643*	2	MWG557	-0.442
235.056*	2	snp_1272	-0.447
235.856*	2	CD0537	-0.450
302.873*	2	ABG316C	-0.469
334.595*	2	snp_0474	-0.475
376.050*	2	olad_259	-0.480
376.050*	2	P40FA	-0.480
334.152*	2	Hvex1	-0.473
302.409*	2	snp_1121	-0.465
237.836*	2	snp_0545	-0.444
168.552*	2	snp_0265	-0.411
158.332*	2	snp_0407	-0.404
148.992*	2	bBE54D	-0.398
132.498*	2	snp_0650	-0.386
111.777*	2	snp_1449	-0.365
105.611*	2	snp_1435	-0.358

High LRS clearly indicates co-segregating markers.