

Manual for RC_Link Version 2.0

26 September 2005

RC_Link is a tool for genetic linkage analysis. *Genetic linkage analysis* is a statistical method for ordering genes on a chromosome and determining the distance between them, and is very useful in predicting and detecting diseases and associating functions to genes.

There are many genetic linkage analysis tools, most notably Fastlink, Linkage, GeneHunter, and Vitesse. A list of these along with many more is located at: <http://linkage.rockefeller.edu/soft/list.html>

Many of these tools handle either large populations or many locations on the chromosome, but not both. A more recent tool called Superlink showed that by using Bayesian networks they were able to model this domain and solve many problems which previous state-of-the-art tools were unable to handle, including those with either large populations or many chromosome locations, as well as many problems with both.

RC_Link similarly models the genetic linkage analysis problem as a Bayesian network, and then simplifies the network and computes the results using the *Recursive Conditioning* algorithm. These techniques allow RC_Link to be one of the most efficient tools for solving many genetic linkage analysis problems.

Below are the options used for running RC_Link and a brief description of the input files. Additional information and contact details can be obtained from RC_Link's webpage at http://reasoning.cs.ucla.edu/rc_link.

RC_Link is written entirely in Java and therefore should run on any computer with a Java virtual machine. It has been tested on Windows and Linux under Java version 1.4.

Input Files:

RC_Link uses two input files, one for the Loci information and one for the Pedigree information. They are in exactly the same format used by Superlink, and very similar to those used by Linkage/Fastlink and Tlinkage (with the exception of a few additional options). RC_Link currently supports program codes 4 and 5 in the locus file. For a detailed description of the files, please refer to Superlink's website at <http://bioinfo.cs.technion.ac.il/superlink/>

Command Line Options:

The command-line format for RC_Link 2.0 is:

```
java -cp rc_link.jar edu.ucla.belief.rc2.tools.RC_Link [options] Locus [file]
Pedigree [file]
```

Where [**options**] can contain:

- Help
- Method Generate-Net
- Scalar [integer greater than 6]
- ExtendedSearch

The meanings of the extra options are:

“**Help**”: displays a message with the command line format

“**Method Generate-Net**”: only creates a network without doing any calculations.

This is useful for allowing RC_Link to work with other tools, for example the Ace tool (<http://reasoning.cs.ucla.edu/ace/>). These tools must be careful to check for underflows in the probability calculations, as the probabilities generated are frequently very small. Additionally, these networks have had the evidence incorporated into them and have been simplified; therefore standard BN tools may not allow them to be opened.

This option generates two files: network_[locus file].net and constant_[locus file].txt. The first contains the network and the second contains a constant that any likelihood query must be multiplied by. Since the constants can become very small, they are stored in a scaled format. The actual constant is the value in the file raised to the power of the scalar value (also included in the file).

“**Scalar**” Computations are done using scaled numbers to prevent underflow. If the values do underflow, please increase this value above its default of 6.

“**ExtendedSearch**”: Tells RC_Link to spend more time searching for a good dtree, which can speed up the eventual computations.

Java Command Line:

Additional command line options may be issued to Java to improve performance, for example you may allocate specific amounts of memory, however these options depend on the specific java virtual machine. An example command line using the Sun virtual machine might be:

```
java -Xss8M -Xmx1800M -cp rc_link.jar edu.ucla.belief.rc2.tools.RC_Link
locus datafile.dat pedigree pedfile.dat
```

The option `-Xss8M` sets the stack size to 8MB and `-Xmx1800M` sets the maximum heap to 1.8GB.

The above command lines are designed to have `rc_link.jar` in current directory, however this file can be moved to another location so long as the `CLASSPATH` updated (see the Java manual for details on changing this for your specific Java virtual machine).

Program Output:

RC_Link displays the probability values for each of the requested recombination values on the command line. Since the actual probability of evidence values are usually very small (and frequently underflow), the results are displayed in three forms: $\ln(\text{likelihood})$, $\log(\text{likelihood})$, and $-2*\ln(\text{likelihood})$. Also, the LOD score for each value is computed and displayed.

Download Files:

The downloadable zip file contains 5 files.

rc_link.jar: The program file

pedfile.dat and datafile.dat: Example input files

rc_link.bat (Windows Bat file)

rc_link.sh (Unix-Bash Shell Script)

The last two files are scripts to allow RC_Link to be more easily invoked (when used with Sun's Java virtual machine). Initially the parameter `-Xmx` should be updated to the amount of memory you would like Java to have access to (the scripts default to 1.8GB). Then simply type `rc_link.bat locus datafile.dat pedigree pedfile.dat` (change the name of the script and input files accordingly)(on Unix systems you may need to give the script executable permission).