Behavior Genetics, Vol. 36, No. 1, January 2006 (© 2006)
DOI: 10.1007/s10519-006-9045-z

ViewPoint and ViewDist: Utilities for Rapid Graphing of Linkage Distributions and Identification of Outliers

Harry N. Beeby,[1,3] Sarah E. Medland,[1,2] and Nicholas G. Martin[1]

Received 3 Apr. 2005—Final 20 July 2005

ViewPoint and ViewDist are standalone Java utilities designed to provide an interactive visual representation of data from files which may be created by the user or produced directly from analyses packages such as Mx. ViewPoint is designed for representation of linkage results and can be used for summarizing multiple univariate linkage results or providing detailed comparison of univariate and multivariate results, with optional dynamic sub-plots designed to display the QTL path coefficients at each marker. ViewDist is an easy to use distribution plotter designed for use with raw data or case-wise likelihood statistics. The utility can produce Q–Q plots (assuming either normal or χ^2-distributions) and by allowing data to be read simultaneously from two files readily produces difference plots allowing comparison of the case-wise likelihood under two competing models.

KEY WORDS: Java utility; Mx; graphical representation; linkage.

Utilities for accurate, customizable and portable graphical representation of linkage results, summary statistics and raw data are important both for the presentation and dissemination of results and to aid in the understanding and interpretation of data. While there are many programs currently available, the purpose of the present note is to introduce two utilities, ViewPoint and ViewDist, which are specialized for graphical representation of linkage results and raw data/summary statistics for easy detection of outliers, respectively. Both utilities are written in Java (requiring Java v1.3 or higher), making them readily portable between those platforms which support graphical displays (Windows, Linux, Solaris, etc.). Input is in the form of simple tab delimited data files which may be created by the user or read directly from analyses packages such as Mx (Neale *et al.*, 2002).

VIEWPOINT

ViewPoint, is designed to provide a graphical representation of linkage results and produces two styles of graph. The first Univariate style provides simultaneous display of up to ten sets of results. The second Multivariate style is designed specifically for comparison of multivariate and univariate results. In both cases the display has two tabbed panels, the first for displaying one chromosome at a time (with easy navigation provided through drop-down menus and arrow buttons), the second for displaying all chromosomes together in individual graphs scaled for chromosome length. All displays can be printed, or saved to a JPEG file.

Univariate style graphs provide an easy way to compare multiple sets of results. If the first line of the input file contains column headings they will be used for labeling. The first three columns are assumed to be Chromosome, Map Distance (assumed to be in cM), and Marker ID. These three header columns are followed by an unlimited number of columns containing linkage results. After the data file is read in, a drop-down menu allows the user to select which columns they wish to display, with up to ten sets of

---

[1] Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Australia.

[2] Cognitive Psychophysiology Laboratory, University of Queensland, Toowoomba, Australia.

[3] To whom correspondence should be addressed at Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, 4029, Australia. Tel.: +61-7-3362-0242; Fax: +61-7-3362-0101; e-mail: harry.beeby@qimr.edu.au

results displayed on the graph at one time. The choice of results displayed can be altered through the menu options without needing to re-read the file.

Marker position is indicated as a series of active points along the *X*-axis, and will initiate a pop-up showing the marker name and position in cM when the mouse pointer passes over them. These pop-ups can be pinned on the display by double clicking the active point. A vertical guide line appears at the same time as the pointer moves over the active point in order to assist in identifying the marker(s) beneath any peaks of interest. For reference, significance and suggestive thresholds are also displayed (by default these are set to asymptotic values of 3.6 and 2.2 (Lander and Kruglyak, 1995) although these maybe changed to empirical values or removed by the user). The resulting graphs are highly customizable allowing users to choose line colors and styles, font sizes, background colors and specify axis labels and horizontal grid lines. The user is also able to either fix the *X*-axis, scaling each chromosome by the length of the longest chromosome, or allow the axis to vary with the graphs scaling to window size. An example graph is given in Figure 1 below.

While multivariate and univariate results can easily be compared using the Univariate style graph, the Multivariate style graph provides a dynamic way to display additional information such as standardized QTL path coefficients within the graph, thereby summarizing both the evidence for linkage and the QTL factor loadings for each variable at a glance. As in the univariate case, the three header columns in the input file (Chromosome, Map Distance (assumed to be in cM), and Marker ID) are followed by an optional univariate LOD score, such as linkage to a mean or other summary measure (users should provide a column of zeros if they do not intend to plot a univariate linkage result), the multivariate LOD score, and up to ten additional values (standardized path coefficients or similar). These additional values will be plotted on a histogram in the pop-up window.

The pop-up can display both positive and negative values. Marker name, score, and position are shown in the top left hand corner of the display. Active points at marker locations appear as points on the multivariate plot line. By double clicking on an active point, the pop-up window can be pinned to the display. The top left corner of the information sub-window will be connected to the point to which it pertains by a red connecting line. Up to ten information sub-windows can be displayed at any one time. When comparing a multivariate linkage to

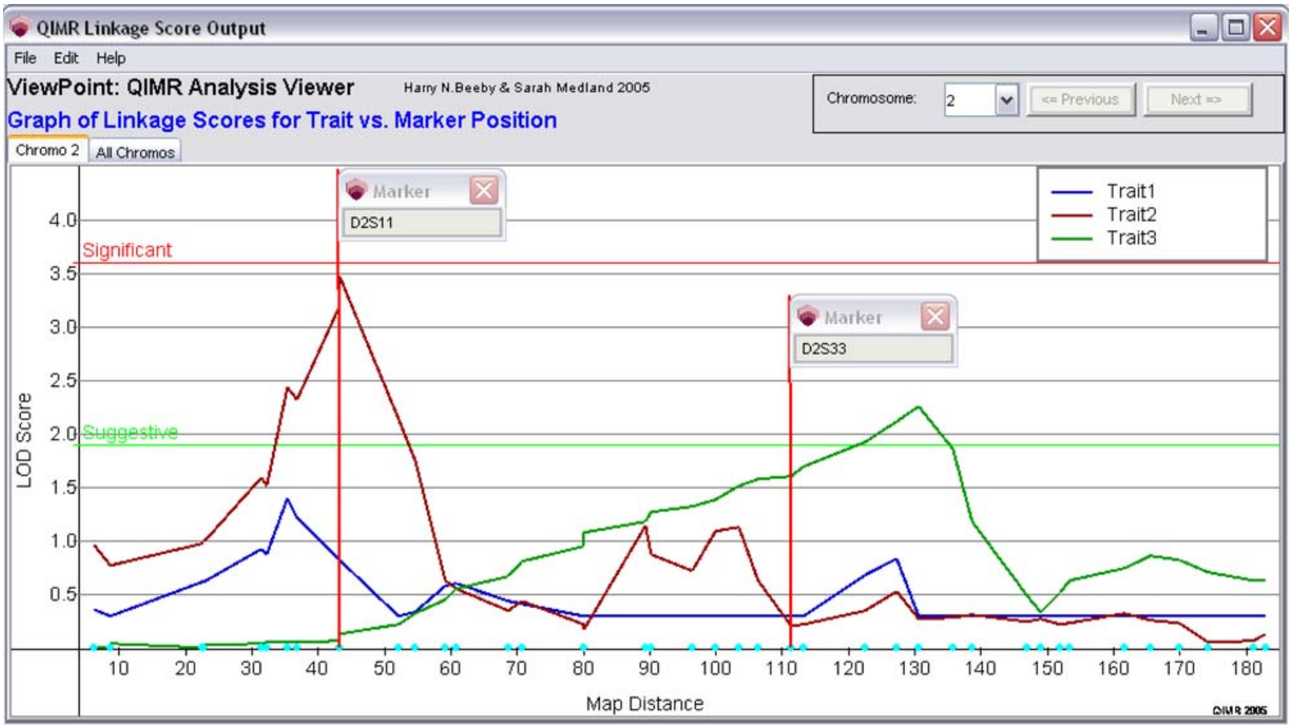


Fig. 1. An example data set plotted using ViewPoint showing The Univariate style graph.

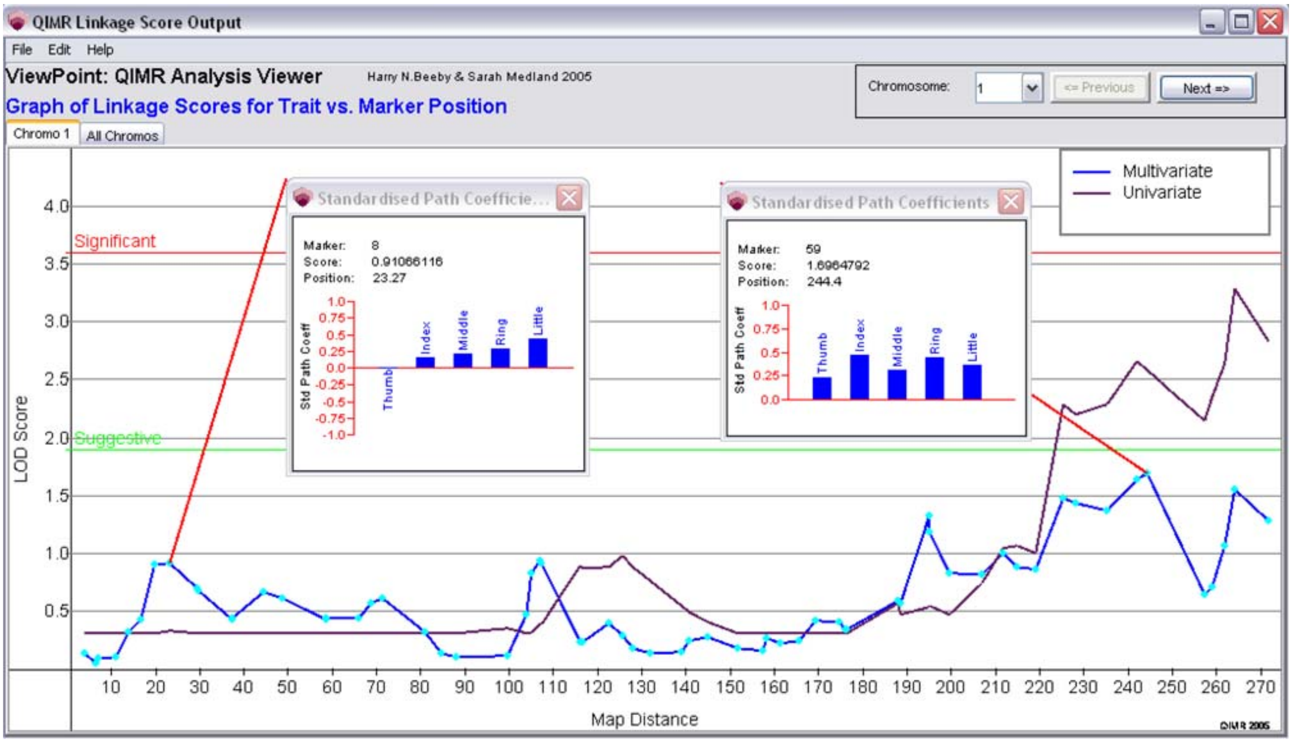
univariate linkage to a summary variable (e.g. IQ sub-test results vs. full scale IQ), this feature allows an easy comparison of factor loadings at points where the multivariate LOD score outperforms the summary measure, and those where the summary measure performs better. As in the Univariate style graph the Multivariate graph is customizable and may be saved as a JPEG file or printed. An example graph is given in Figure 2 below.

VIEWDIST

ViewDist aids in the detection and exclusion of outliers, and can be used to produce graphs of raw data or summary statistics. The utility produces Q–Q plots against either a normal or χ^2-distribution or can be used to plot the data against itself (an internal plot) which can be useful in detecting outliers in other distributions. In all cases the application takes the input data and sorts it into numerical order. The normal and chi-squared distributions are based on the SourceForge JSci Open Source library (SourceForge, 2005). The graphs produced are customizable allowing user specification of Background Color, Axis Color, Title and Font Size. In addition the display may be printed, or saved to a graphic JPEG file.

Although the utility will accept any tab delimited file, it is specifically designed to complement the Mx matrix algebra program and the default input file is an Mx %p file (from Mx version 1.57 or above, which prints the first definition variable i.e. Family number, in the first column of the output file). This file is produced using the command *Option Mx%p = filename.p* and contains the following columns for each case: identifier, −2 log likelihood (−2LL), the square root of the Mahalanobis distance, estimated *z*-score, number of the observation, number of data points, number of times the −2LL was found to be incalculable, incalculability code and model number. If the %p file contains data from multiple models, that is if the file was generated from an Mx script using the *NModels* command such as a mixture-distribution linkage analysis, ViewDist will simultaneously plot the data from each model using color to differentiate them. By default the first column will be used as the data labels, and the data in the fourth column will be plotted and the ninth column will be assumed to contain model number. Use of alternate file formats simply requires specification of the column numbers that contain these data.

As with other Q–Q plots, if the observed values follow the specified distribution, the plotted points



Marker: 8
Score: 0.91066116
Position: 23.27

Marker: 59
Score: 1.6964792
Position: 244.4

Fig. 2. An example data set plotted using ViewPoint showing the Multivariate style graph, with two pop-up histograms.

will align with the diagonal ($x = y$) which is drawn on the display for easy reference. An example graph is given in Figure 3 below. One of the key features of this application is to be able to select a point on the graph and display the identifier to which it belongs. Placing the mouse pointer over a point displays the case identifier. Points on the graph may be selected either by clicking the left mouse button whilst the pointer is over the point, or by pulling out a selection box. The selection box will appear when the left mouse button is held down and the mouse is moved, and shows as a rectangle bounded by a dark boundary line. All points that are encompassed by the box when the mouse button is released will become selected. The Control key may be held down to select multiple points or groups, and there is a menu option to select all. A list of the selected data points can be output in the form of a series of Mx select statements which allows for easy exclusion of the selected data in subsequent analyses. Following the selection of this option from the menu, the user will be asked for a variable name to be used in the output file. For example, if the default value of "fam" is taken, the output script will look something like the following:

```
select if fam ^= 1234;
select if fam ^= 5678;
```

In addition to plotting the data contained in a single file, ViewDist has the ability to simultaneously access two data files and calculate a difference score which may be graphed as either a Q–Q plot or an internal plot. This feature is particularly useful if %p files are used as input, as it can provide a graph of the difference in −2 log likelihood between two models (i.e. a linkage model and a null model) providing an easy way to determine which families are contributing the most to the linkage signal. Both files are required to be in the same format and contain data for exactly the same set of identifiers (though not necessarily in the same order).

RUNNING THE UTILITIES

ViewPoint and ViewDist are freely available as beta test versions and can be downloaded from http://genepi.qimr.edu.au/viewapps/. Both the ViewPoint and ViewDist utilities contain help files that provide information regarding file formats and menu options. In the Windows environment, both utilities may be opened by simply double clicking the .jar files. Alternatively, they can be run from a command prompt using the syntax provided in the Appendix A.

APPENDIX A. SYNTAX FOR RUNNING VIEWPOINT AND VIEWDIST FROM THE COMMAND PROMPT

ViewPoint

```
java -jar ViewPoint.jar
```

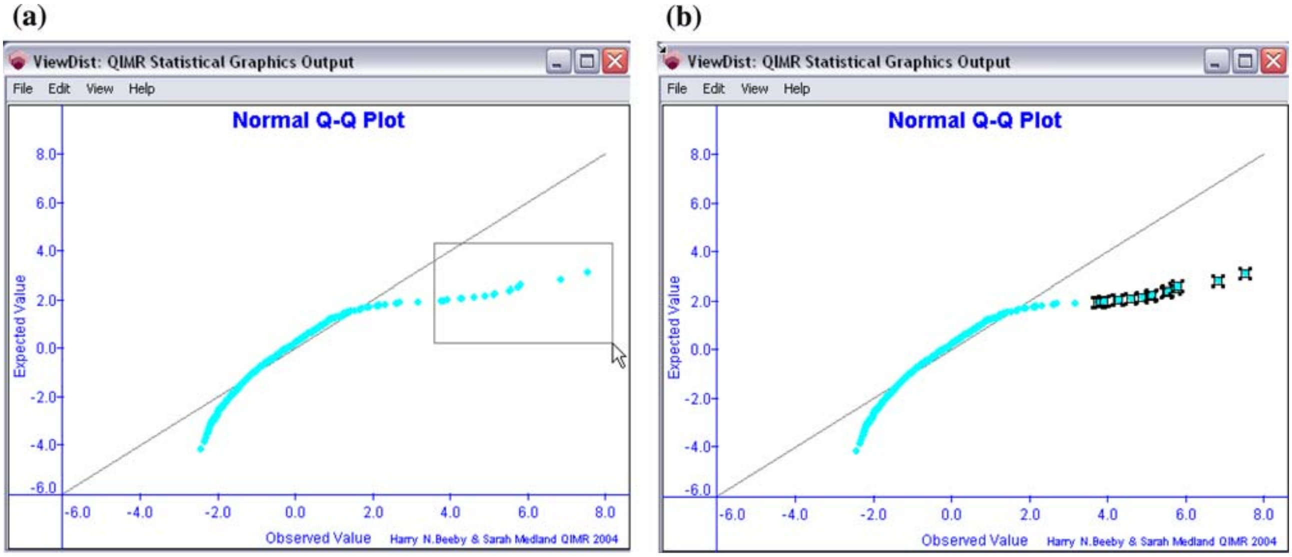


Fig. 3. An example data set plotted using ViewDist showing the two methods of selection (a) multiple selections with a selection box, (b) multiple selection using the Crl key.

```
  [-infile <input data filename>]
  [-suggestive <suggestive score threshold (float)>]
  [-significant <significant score threshold (float)>]
  [-fixXaxis (X axis units set to max chromo length)]
  [-plot <plot type U or M>]
  [-help (usage)]
```

```
ViewDist
java -jar ViewDist.jar
  [-infile <input data filename>]
  [-diffile1 <first difference filename>]
  [-diffile2 <second difference filename>]
  [-infoCol <info column number>]
  [-scorecol <score column number>]
  [-intObsCol < observation count column number>]
  [-intModelCol <model column number>]
  [-plottype <plot type: internal, normal, chi>]
  [-help (usage)]
```

ACKNOWLEDGMENTS

The authors wish to acknowledge the support and helpful suggestions of David Duffy and Michael Neale. In addition, the authors would like to thank the staff and students of the Genetic Epidemiology Laboratory of the Queensland Institute of Medical Research, and in particular Narelle Hansell and Manuel Ferriera, for their comments, suggestions and testing throughout the development of these utilities.

REFERENCES

Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**:241–247.

Neale, M. C., Boker, S. M., Xie, G., and Maes, H. H. (2002). *Mx: Statistical Modeling* (6th ed.), VCU Box 900126, Richmond, VA 23298 http://www.vcu.edu/mx/: Department of Psychiatry

SourceForge JSci Science API for Java 2005; an Open Source library of mathematical and scientific functions – website: http://www.sourceforge.net/projects/jsci

Edited by Danielle Posthuma