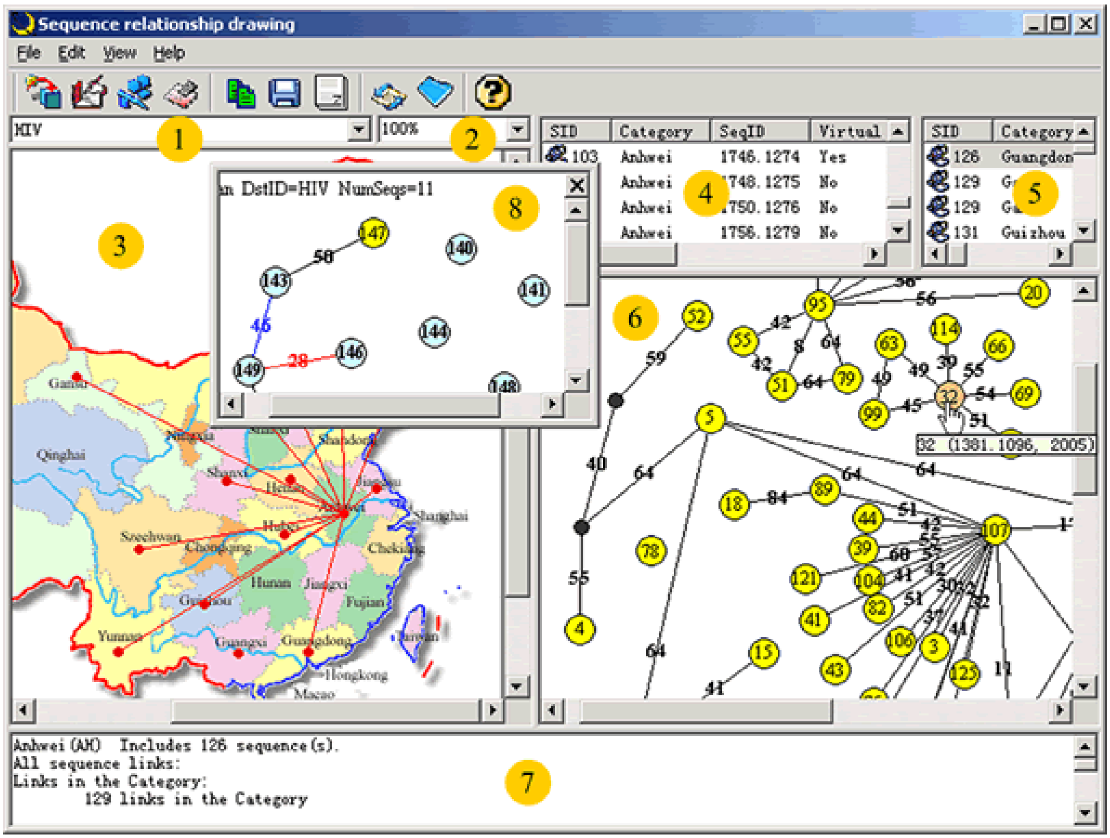
SRD User Manual

Table of contents:

Overview

SRD (Sequence Relationship Drawing program) was developed in order to dynamically draw plane undirected graphs for a wide range of sequence relationship visualization analysis, either for

Figure 1H. Overview of SRD main GUI window.



SID	Category	SeqID	Virtual
103	Anhwei	1746.1274	Yes
	Anhwei	1748.1275	No
	Anhwei	1750.1276	No
	Anhwei	1756.1279	No

SID	Category
126	Guangdon
129	
129	
131	Guizhou

```
Anhwei (AH)  Includes 126 sequence(s).
All sequence links:
Links in the Category:
        129 links in the Category
```

nucleic acid sequences or for peptide sequences. Categories analysis is also supported, which allows visualizing intra-category sequence relationships, extra-category sequence relationships and category-category relationships in several specific windows, respectively.

The following picture is the overview of SRD main GUI window. The sample dataset used here is a dataset of China envelope glycoprotein (env) gene of HIV sequences derived from Chinese center for disease control and prevention.

The main GUI contains several sub-windows:

Number	Window	Description
(1)	Dataset selector	A drop-down menu to select dataset to visualize.
(2)	Zoom selector	A drop-down menu to set or return the zoom factor used for displaying all graphs.
(3)	Inter-category drawing window	Used for drawing category-category relationships, which depicts Chinese province-province relationships in this example since sequences in the dataset come from several different provinces in China.
(4)	Current sequence list window	Lists all sequence within the current category.
(5)	Foreign sequence list window	Gives a list of all related sequences linked to the current category or to the selected sequence.
(6)	Intra-category drawing window	The visualization graph for the current category.
(7)	Information window	Dynamically displays the corresponding text information of all generated graphs at a time.
(8)	Extra-category drawing window	A floating window to visualize another category related to the current category.

First use of SRD

When the program is first started, users will find a Start Wizard to quickly access general functions in the system. Novice users should follow the steps and click the commands in the wizard window to perform their visualization.

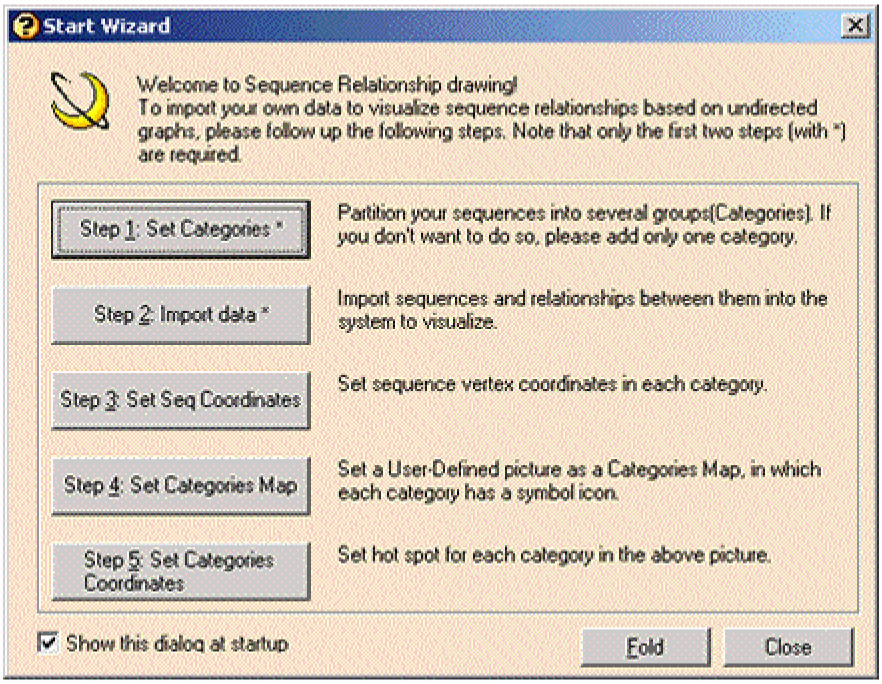
To begin the visualization:

1. In the first step, partition your sequences dataset to several categories. Input the CategoryID and the name of each category into the system. Both the CategoryID and the name are strings no more than 50 characters. However, CategoryID should be unique. If one would like to visualize all sequences as a whole, i.e., not perform the partition into several categories, he must also input one category, i.e., "partition" all sequences into one group.

2. Assign a unique zero-based number to each of the sequence, which is called SID number.

Then create a 'Sequence Information Script File' to contain all sequences information and a 'Sequence Links Script File' to contain the relationship values between each two sequences. Import the two script files into the system using the 'Import New data' dialog.

Figure 2H. Start Wizard in SRD.



Welcome to Sequence Relationship drawing!
To import your own data to visualize sequence relationships based on undirected graphs, please follow up the following steps. Note that only the first two steps (with *) are required.

Partition your sequences into several groups(Categories). If you don't want to do so, please add only one category.

Import sequences and relationships between them into the system to visualize.

Set sequence vertex coordinates in each category.

Set a User-Defined picture as a Categories Map, in which each category has a symbol icon.

Set hot spot for each category in the above picture.

☑ Show this dialog at startup

- To open the 'Import New data' dialog, click 'Import New data' command in the 'File' menu.

3. (This step is optional) Manually rearranged the graph for each category by mouse-dragging the sequence vertices to make the graphs more beautiful, using the 'Update database' dialog.

- To open the 'Update database' dialog, click 'Update data' command in the 'File' menu.

4. (This step is optional) To vividly depict different categories, create a picture (called Categorys Map) using any image creation and editing tools, such as MS Paint, Photoshop, Fireworks, etc., in which each category has an icon on it. Import the picture using the 'Update database' dialog.

5. (This step is optional) Using the 'Update database' dialog again, set hot spots for each category on the Categorys Map imported in the previous step.

Sample datasets are available on the website.

The ‘Sequence Information Script File’ format

Users must create a “Sequence Information” script file that contains all sequences in the dataset. “Sequence Information” script file adopts a format as a FASTA file that begins with a single-line description, followed by lines of sequence data.

The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. There should be no space between the ">" and the first letter of the identifier. The description line must exist for each sequence while the rest sequence data lines are optional. The description line must adopt a specific format with several fields separated by a “|” character. The format of the description line is as follows:

>SID | CategoryID | SeqID | [IsVirtual] | [Caption] | [Memo]

, where fields in brackets are optional and can be omitted. Each sequence must be assigned a unique zero-based identifier number called SID. And each category must also be given a string identifier (a string not more than 50 characters), which is called CategoryID. The field SeqID can be the name or the database ID of the sequence (e.g. a gi number of NCBI, a PDBID of PDB database, etc). IsVirtual indicates it is a virtual sequence (=1) or not (=0). Caption is a text string that describe the sequence (could also be none). Other corresponding information can be included in the field Memo.

Figure 3H. A sample of ‘Sequence Information Script File’.

```
>154|JL|5691.7614|0|1998
TAGTACAgTTGAAtGAAtCTGTAGAAATTAAtTGTACaAGA-CCCAACAACAATACAAGA AAAAGTATACaTCTAGGAC
CAGGGCAA-GCATGGTATACAACAGGACAAATAATAGGAGA TATAAGACAAGCACATTGTAAcATTAATAgA-ACAAAA
TGGAATAAcACTTTA-aGA-cA GATAAcTGaAAAATTAAgA------GAaCAaTTTggg-------aacAAAACAATAA
tCT TT---AAtCaATCT---TCAGGAGGGGACCCAGAGATTGTAATGCACAGTTTTAATTGTG GAGGGGAATTTTTC-
-TACTGTAATACATCAcAaCTGTTTaatagtacttggaataatac tagt---------------------acttggaat
gatactacaggagatagcAAtATcAC G---CTCCCTTGCAGAATAAAACAA---ATTATAAAcATG-TGGCAG-GAAGT
A--AGG- AAAGCA-ATG-TATGCCCCTCCCATT-aGAGGA-cAaATTAgA-TGTTcATCA-AATATT ACA--GGGTTG
cTATTAACAAGAGATGGTGGT---------------------------- -----TAtAAt-GAgAAcAaG--ACcACcG
AN--ACC-TCANA-CCT-GNAGGA-GG
>155|JL|5692.7615|0|1998
TAGTACAgCTGAAtGAAtCTGTAGAAATTAAtTGTACaAGA-CCCAACAACAATACAAGA AAAAGTATACcTCTAGGAC
CAGGGCAA-GCATGGTATACAACAGGACAAATAATAGGAGA TATAAGACAAGCACATTGTAAcCTTAGTAgA-ACAAAA
TGGGATAAcACTTTA-aAA-cA GATAAcTGaAAAATTAAgA------GAaCAaTTTggg-------aacAAAACAATAA
tCT TT---AAtCaATCT---TCAGGAGGGGACCCAGAGATTGTAATGCACAGTTTTAATTGTG GAGGGGAATTTTTC-
-TACTGTAATACATCAcAaCTGTTTaatagtacttggaatggtac tagt---------------------acttggaat
gatactacaggaaatagcACtATcAC G---CTCCCTTGCAGAATAAAACAC---ATTGTAAAcATG-TGGCAG-GAAGT
A--GGG- AAAGCA-ATG-TATGCCCCTCCCATT-aGAGGA-cAaATTAcA-TGTTcATCA-AATATT ACA--GGGTTG
cTATTAACAAGAGATGGTGGT---------------------------- -----AAtAAt-GAgAGcAtT--ACcACcG
AG--ACCTTCAAA-CCTgGGAGGA-GG
>156|JL|5693.7616|0|1998
TAGTACAgTTGAAtGAAtCTGTAGTAATTAAtTGTACaAGA-CCCAACAACAATACAAGA AAAAGTATACcTCTAGGAC
CAGGGCAA-GCATGGTATACAACAGGACAAATAATAGGAGA TATAAGACAAGCACATTGTAAcCTTAGTAgA-ACAAAA
TGGGATAAcACTTTA-aAA-cA GATAAcTGaAAAATTAAgA------GAaCAaTTTagg-------aacAAAACAATAA
tCT TT---AAtCaATCT---TCAGGAGGGGACCCAGAGATTGTAATGCACAGTGTTAATTGTG GAGGGGAATTTTTC-
-TACTGTAATACATCAcAaCTGTTTaatagtacttggaatgataa tagt---------------------acttggaat
aatactacaggagatagcAAtATcAC G---CTCCCTTGCAGAATAAAACAA---ATTGTAAAcATG-TGGCAG-GAAGT
A--GGG- AAAGCA-ATG-TATGCCCCTCCCATT-aGAGGA-cAaATTAgA-TGTTcATCA-AATATT ACA--GGGTTG
cTATTAACAAGAGATGGTGGT---------------------------- --AaAGa-ATg-AGaACgCcG--ACcACcG
AG--ACCTTCAAA-CCTgGGAGGA-GG
>157|JL|5694.7617|0|1998
TAGTACAgTTGAAtGAAtCTGTAGAAATTAAtTGTACaAGA-CCCAACAACAATACAAGA AAAAGTATACaTCTAGGAC
```

The sequence ends if another line starting with a ">" or end of file (EOF) appears.

The script file is text-based and can be created and edited using any text-processing tools or by simply pasting from some other software tools, such as Excel, Orgin Pro, or Matlab. It can be imported directly into the system by an interactive ‘‘Import New data’ dialog provided by SRD, which is activated by mouse clicks throughout the entire importing process. An input module of the software will automatically detect the file formats and an internal syntax analyzer will give back any error information. When the input process completes, the parser recognizes the text and stores all the information into the computerized database along with the application.

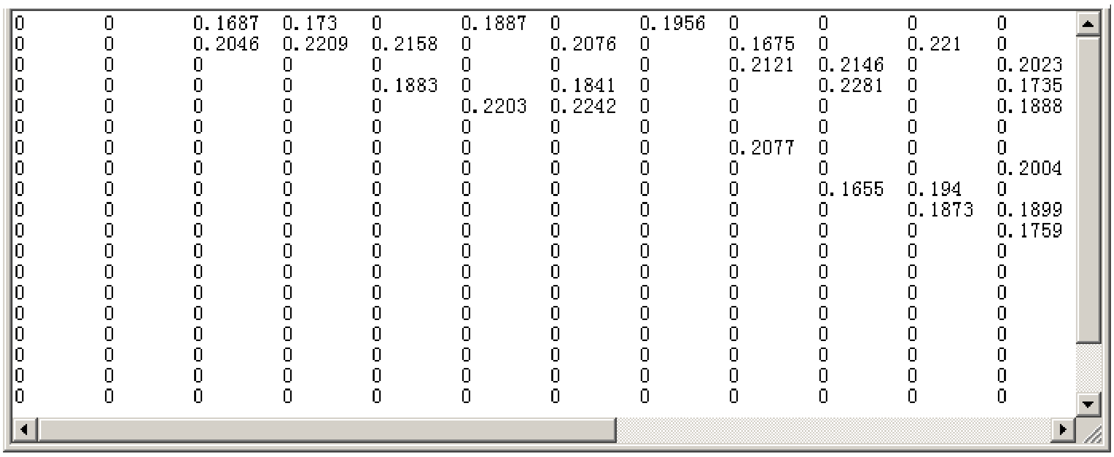
The ‘Sequence Links Script File’ format

Users must create a “Sequence Links” script file that contains relationships between each two sequences. Both the script files are text-based and can be generated and edited using any text-processing tools.

'Sequence Links script file' can adopt one of the following format:

- Matrix format: a $n \times n$ matrix separated by spaces in each line. The separator can be one space or several spaces or even Tab, which is useful to load a table style text file.

Figure 4H. A sample of 'Sequence Links script file' (Matrix format)



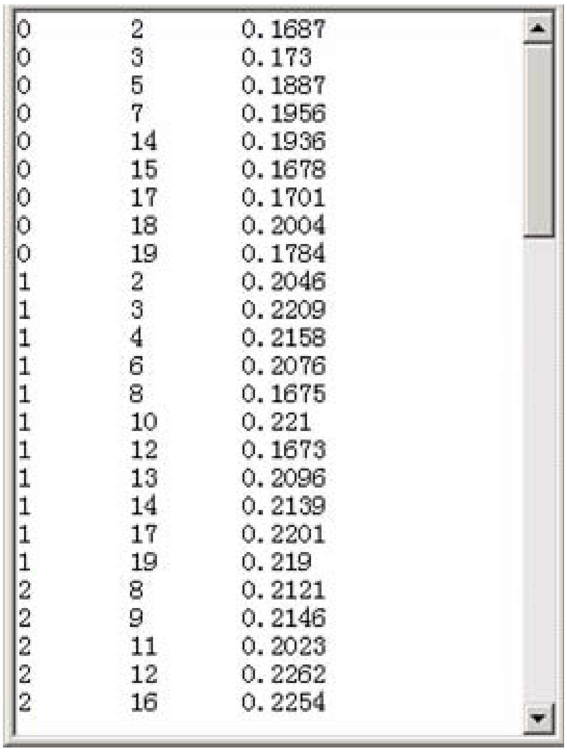
```
0       0       0.1687  0.173   0       0.1887  0       0.1956  0       0       0       0
0       0       0.2046  0.2209  0.2158  0       0.2076  0       0.1675  0       0.221   0
0       0       0       0       0       0       0       0       0.2121  0.2146  0       0.2023
0       0       0       0       0.1883  0       0.1841  0       0       0.2281  0       0.1735
0       0       0       0       0       0.2203  0.2242  0       0       0       0       0.1888
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0.2077  0       0       0
0       0       0       0       0       0       0       0       0       0       0       0.2004
0       0       0       0       0       0       0       0       0       0.1655  0.194   0
0       0       0       0       0       0       0       0       0       0       0.1873  0.1899
0       0       0       0       0       0       0       0       0       0       0       0.1759
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0       0       0
```

- Relationship lines format: each line indicates one relationship and its value between one pair of sequences. For instance, if sequence *SID1* and sequence *SID2* has a relationship value *V*, the corresponding line should be:

SID1   *SID2*   *V*

The separator can either be a "Tab" or any number of spaces. SID1 and SID2 must be consistent with the ones in the 'Sequence Information script file'.

Figure 5H. A sample of 'Sequence Links script file' (Relationship lines format)



```
0       2       0.1687
0       3       0.173
0       5       0.1887
0       7       0.1956
0       14      0.1936
0       15      0.1678
0       17      0.1701
0       18      0.2004
0       19      0.1784
1       2       0.2046
1       3       0.2209
1       4       0.2158
1       6       0.2076
1       8       0.1675
1       10      0.221
1       12      0.1673
1       13      0.2096
1       14      0.2139
1       17      0.2201
1       19      0.219
2       8       0.2121
2       9       0.2146
2       11      0.2023
2       12      0.2262
2       16      0.2254
```

The script file is text-based and can be created and edited using any text-processing tools or by simply pasting from some other software tools, such as Excel, Orgin Pro, or Matlab. It can be imported directly into the system by an interactive ''Import New data' dialog provided by SRD, which is activated by mouse clicks throughout the entire importing process. An input module of the software will automatically detect the file formats and an internal syntax analyzer will give back any error information. When the input process completes, the parser recognizes the text and stores all the information into the computerized database along with the application.

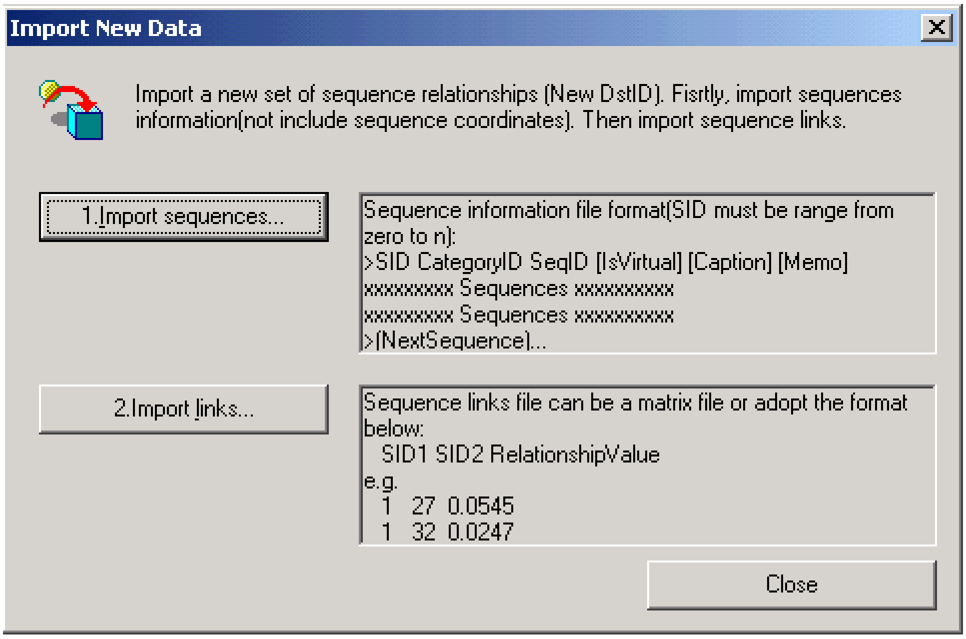
Import new data to visualize

To start a sequence relationship visualization, sequences and the relationship values between them must be imported into the system. Users must create a 'Sequence Information script file' that contains all sequences in the dataset and a 'Sequence Links script file' that contains relationships between each two sequences.

Import both the script files using the ''Import New data' dialog.

- To open the 'Import New data' dialog, click ''Import New data' command in the 'File' menu.

Figure 6H. The "Import New Data" dialog



Import a new set of sequence relationships (New DstID). Fisrtly, import sequences information(not include sequence coordinates). Then import sequence links.

```
Sequence information file format(SID must be range from
zero to n):
>SID CategoryID SeqID [IsVirtual] [Caption] [Memo]
xxxxxxxxx Sequences xxxxxxxxxx
xxxxxxxxx Sequences xxxxxxxxxx
>[NextSequence]...
```

```
Sequence links file can be a matrix file or adopt the format
below:
  SID1 SID2 RelationshipValue
e.g.
  1   27  0.0545
  1   32  0.0247
```

Click 'Import sequences' command to import a 'Sequence Information script file'. Then click 'Import Links' to import a 'Sequence Links script file'. Data set ID (DstID) will be asked when click the commands. Please Input a string no more than 50 characters as your dataset ID. The dataset ID you input must be consistent (the same) with the two askings when you click the two commands separately.

When the 'Sequence Links script file' importing is complete, graphs layout for all categories will be performed automatically (automatically set vertices coordinates of all sequences).

Note:

1). Before importing, one must define all categories (CategoryID and CategoryName). And the CategoryID in a ‘Sequence Information script file’ must be consistent with that defined before.

2). When importing new data, if a Data set ID (DstID) a user inputs already exists in the system, the previous dataset will be overwritten. And the system will ask the user if he want to do so.

Intra-category graph drawing

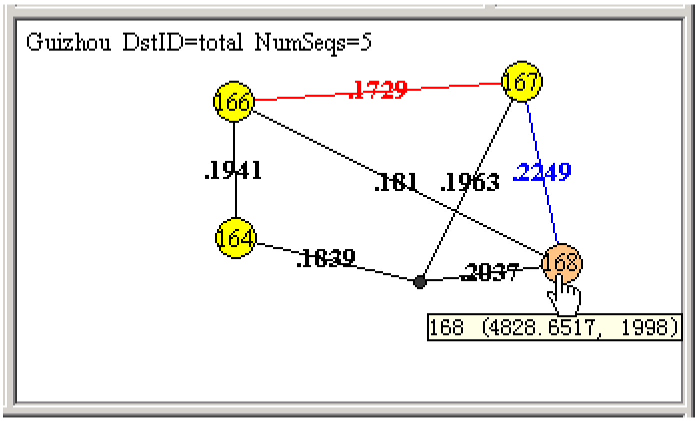
An intra-category graph is used to depict all sequences and their connections within a category (the current category). Unlike some regular graph drawing software, in which the location of an object is fixed, SRD provides an interactive dynamic drawing by using a hot spot technique jointly with a user-steered object location.

Figure 7H. A Categories Map and the categoriy hot spots on it.

Categories are displayed graphically as a user-defined beautiful picture (Categories Map) in the ‘inter-category drawing window’. The picture has a hot spot for each category, which is also user-defined. When a user clicks one of these hot spots, the corresponding category becomes the “current category”, and SRD displays all the sequences of the category and their connections simultaneously in the ‘intra-category drawing window’.

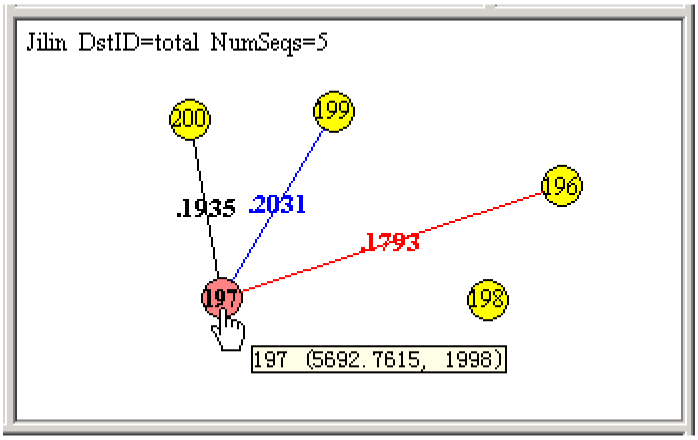
In intra-category graphs, each sequence is represented by an open circle except for the virtual ones, which are displayed as black solid circles. The number marked in the middle of the circle denotes the SID of each sequence. For clarity, only SID is printed in graphs, other properties such as SeqID, Caption, sequence strings, etc. are not printed. However, full details are available in the 'current sequence list window'. On the other hand, SRD provides a hot spot for each sequence vertex. When a user's mouse moves onto a vertex circle, the mouse pointer changes to the pointing finger shape and the cursor returns the SID, SeqID and caption of the corresponding sequence in a ToolTip text box.

Figure 8H. An intra-category graph generated by SRD.



The reticulate relationships between each pair of sequences are displayed as connection lines labeled with the relationship values at the middle of the line. The minimum and the maximum relationship values are calculated and the corresponding connection line is drawn as red color and blue color, respectively. This is useful for the user to find out the minimum and the maximum relationship and acquire the value range at a glance.

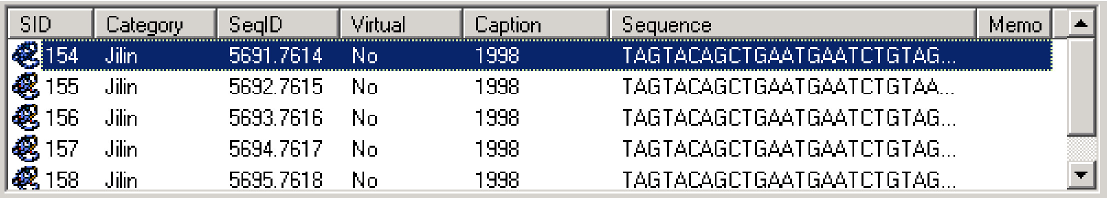
Figure 9H. Select a sequence and display edges separately in an intra-category graph.
(The sequence with SID=197 is selected in this example.)



When there are a large number of sequences and a large number of relationship links, connection lines will mix up together. To overcome this problem, separate edges displaying is supported in the system. Any sequence can be selected by clicking on the corresponding vertex. When a sequence is selected, the corresponding vertex is highlighted with a particular color, and edges only linked to the selected sequence are drawn, other edges are omitted. The minimum and the maximum relationship values are re-calculated and colored lines are also re-marked according to the current lines drawn. When users click anywhere else of the graph, all connection lines will display together again. Which part of the edges to be displayed depends on the user's selection.

Furthermore, sequence selection displays synchronously with the ‘current sequence list window’. When a sequence is selected in the graph, the corresponding item in the list window is highlighted and brought into view automatically, which is useful for a user to check the detail information about the selected sequence. Conversely, sequence items in the list window are also clickable and selectable, when an item is selected, the corresponding vertex in the graph will be highlighted automatically and the drawing window will be scrolled if necessary to make a view of putting the vertex circle in the center. This is very useful for a user to find out the vertex of a specific sequence in the graph from the detail list.

Figure 10H. Details of sequences within the current category are list in the 'current sequence list window'.



SID	Category	SeqID	Virtual	Caption	Sequence	Memo
154	Jilin	5691.7614	No	1998	TAGTACAGCTGAATGAATCTGTAG...	
155	Jilin	5692.7615	No	1998	TAGTACAGCTGAATGAATCTGTAA...	
156	Jilin	5693.7616	No	1998	TAGTACAGCTGAATGAATCTGTAG...	
157	Jilin	5694.7617	No	1998	TAGTACAGCTGAATGAATCTGTAG...	
158	Jilin	5695.7618	No	1998	TAGTACAGCTGAATGAATCTGTAG...	

Category-Category graph drawing

The intra-category graph shows sequences and their relationships within one category. Another undirected graph is provided in SRD in the ‘inter-category drawing window’, with each category as a vertex, to depict category-category relationships.

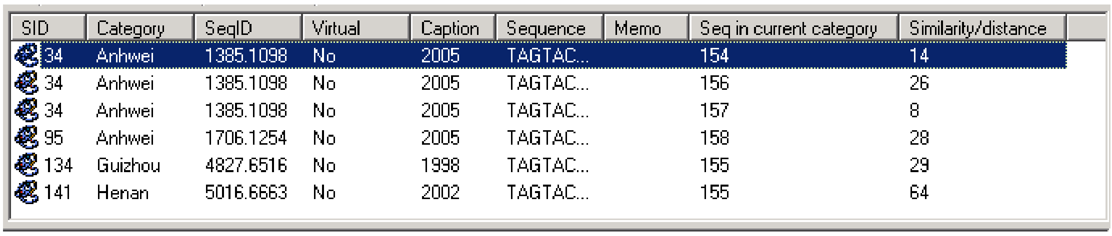
The category-category graph is somewhat different from the intra-category one, although both of them are undirected graphs. Firstly, the fact that a category linked to another only denotes that there exists some relationships between the sequences within the two categories, so we needn’t define a category-category relationship value and therefore no value will be labeled on the connection lines in the category-category graph. Moreover, when drawing an intra-category graph to illustrate sequence relationships within the current category, at a larger scale, the user will want

Figure 11H. A category-category graph.

to see whether and how her current category are related to someone else. At this scale, in the category-category graph, it should be straightforward to display edges that only linked to the current category, while other edges should be omitted at the moment for clarity.

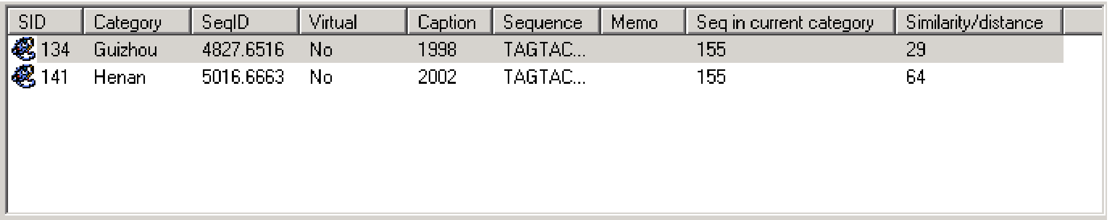
In SRD, when the current category is selected by clicking the corresponding hot spot in the Categories Map, an intra-category graph is generated and simultaneously, edges linked to the current category are displayed as several red lines in the ‘inter-category drawing window’. At the same time, the ‘foreign sequence list window’ is also filled with some sequences, which are called “Foreign sequences” since these sequences listed here denote that each of them is related at least to one of the sequences within the “Current Category”. It is useful to get a detail list of related sequences in other categories while visualizing an intra-category relationship.

Figure 12H. Foreign sequences are list in the ‘foreign sequence list window’.



SID	Category	SeqID	Virtual	Caption	Sequence	Memo	Seq in current category	Similarity/distance
34	Anhwei	1385.1098	No	2005	TAGTAC...		154	14
34	Anhwei	1385.1098	No	2005	TAGTAC...		156	26
34	Anhwei	1385.1098	No	2005	TAGTAC...		157	8
95	Anhwei	1706.1254	No	2005	TAGTAC...		158	28
134	Guizhou	4827.6516	No	1998	TAGTAC...		155	29
141	Henan	5016.6663	No	2002	TAGTAC...		155	64

Figure 13H. The foreign sequence list is altered automatically when a sequence with SID=155 is selected by clicking the corresponding vertex in the ‘intra-category drawing window’ or clicking the corresponding item in the ‘current sequence list window’.



SID	Category	SeqID	Virtual	Caption	Sequence	Memo	Seq in current category	Similarity/distance
134	Guizhou	4827.6516	No	1998	TAGTAC...		155	29
141	Henan	5016.6663	No	2002	TAGTAC...		155	64

There is also a mechanism to get extra-connections separately, i.e. to get extra-connections only linked to the selected one sequence in the current category rather than all of them. This can be simply done by selecting one sequence by clicking on the corresponding vertex in the ‘intra-category drawing window’ or clicking the corresponding item in the ‘current sequence list window’. As mentioned in previous section, this will lead to the selected sequence vertex to be highlighted and connection lines displayed separately in the intra-category graph. At the same time, this will also affect extra-connections. Category-category edges and foreign sequence list will also be altered automatically according to this selection. Once again, how extra-connections are displayed depends on a user’s selection. Combining this feature with intra-category graph provides a macroscopic understanding of sequence relationships.

Extra-category graph drawing

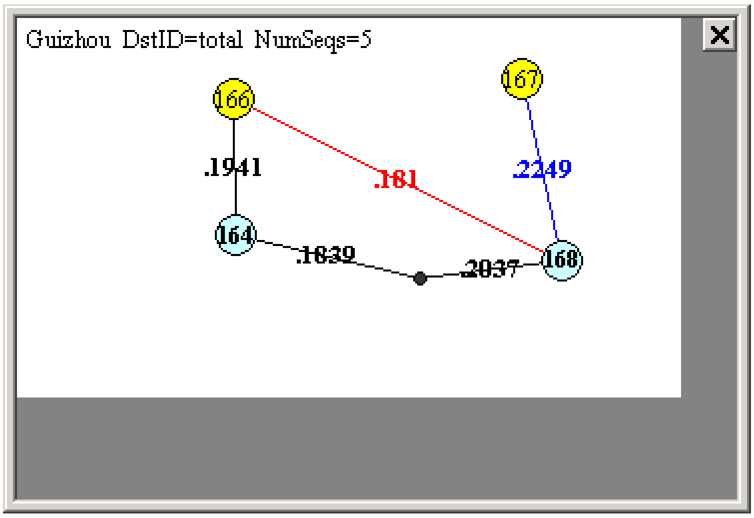
Continue with the previous section, when a user sees several red lines in the ‘inter-category drawing window’, hot spots of all linked categories are assigned a new functional property. With mouse moving onto the hot spots, users will find that the ToolTip has been changed by being added a new line of text: “Right click to depict sequences in related category”, which can not be found on normal category hot spots. If a user right-clicks one of them, the extra-category drawing window will be animated from small to big to become visible in the main GUI. The animation indicates the opening of the right-clicked category. An extra-category graph has been generated in this window, in which sequences and the relationships between them in the right-clicked category are illustrated, just similar as that in an intra-category one. This is a particularly important point for interactive graphic displays, which is useful for a user to visualize the detail foreign sequences and relationships in another related category.

Figure 14H. Hot spots of all linked categories are assigned a new functional property.

Guizhou (5 sequences)
Right-Click to display sequences in related cluster.

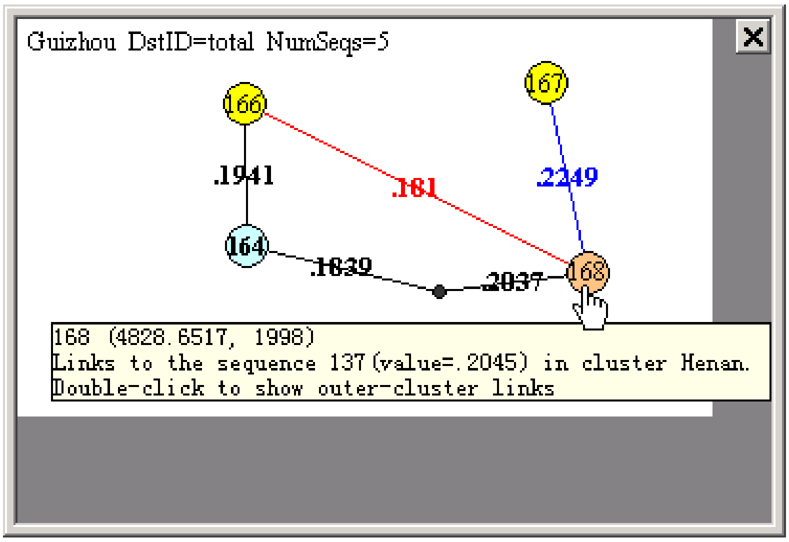
An extra-category graph has most the same characteristics as an intra-category one, such as circular vertices, connection lines, hot spots, etc. Sequence vertices are also clickable to draw connection lines separately. A special characteristic of an extra-category graph is that vertices of all foreign sequences in this graph are filled with a particular color, which denotes that there are some relationships between them and the current selected one. Furthermore, SIDs of related sequences and relationship values can be displayed in the prompt ToolTip text when mouse moving onto the foreign sequences vertices, which allows a user to quickly get the information about which a colored sequence is linked to and how about the relationship values.

Figure 15H. An extra-category graph.



Alternatively, another way to open the ‘extra-category drawing window’ is to click one of the items list in the ‘foreign sequence list window’. In this way, the drawing window will scroll automatically to make the selected sequence into view.

Figure 16H. Cursor gives the linking information when mouse moves onto a foreign sequence vertex.



168 (4828.6517, 1998)
Links to the sequence 137(value=.2045) in cluster Henan.
Double-click to show outer-cluster links

Moreover, this window is a floating window, which is always placed above and can be dragged to anywhere in the main GUI of SRD. It can be closed at any time by clicking the “close” button at the top-right of the window. Animation is used again when the window disappears as it returns to its corresponding category.

Recursion extra-intra categories

As mentioned above, it is easy to view category-category relationships and to generate an extra-category graph. However, although clicking a sequence in an extra-category graph can control edges displaying separately, there is no effect on category-category relationships and foreign sequences, which is different from selecting a sequence in an intra-category graph. This is because there can be only one “current category” at a time, and category-category relationships and foreign sequences are relative to the “current category”. But sometimes, one may concern about the category-category relationships and foreign sequences relative to the extra-category, rather than “current category”, when he is looking at an extra-category graph.

The proposed software adopts an “extra-intra category transferring” technique, which have combined these three perspectives to create a highly integrated multi-window visualization in the SRD environment. This can be simply done by another mouse action—double click. A double click on one vertex in the extra-category graph leads to a “current category transferring”. The ‘extra-category drawing window’ is animated from big to small again but to disappear at the middle of the ‘intra-category drawing window’, which indicates that the current category now has become the one previously in the extra-category window. The double-clicked sequence is then automatically selected in the now ‘intra-category graph’. All other windows are changed according to the new current category and the new selected sequence.

One may right-click a related category again to open the ‘extra-category drawing window’ and double-click a sequence in this window to make a second extra-intra transferring. Then right-click another category, and transfer again … This recursion is helpful to trace a sequence relationship path between different categories, especially in investigating a revolutionary path between different organisms or investigating a disease transmission path between different places.

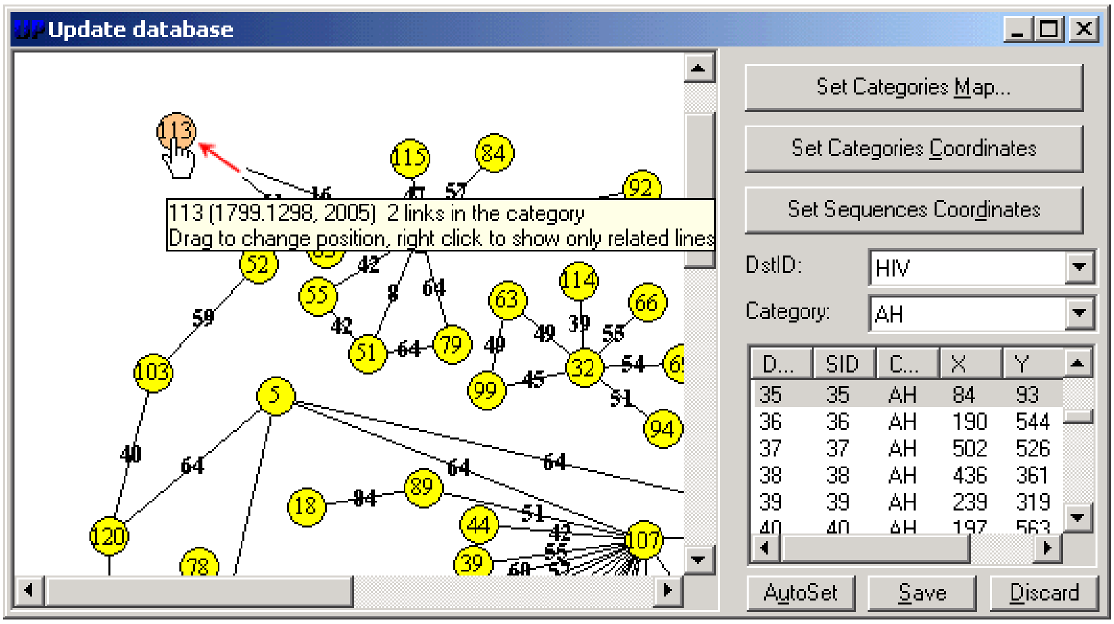
Fully customizable interface

To cater to the broadest array of biological scientists users, the program is fully customizable for categories definition, category-category map picture, all vertices locations and even help document. Users can alter all these things using the ‘update database’ dialog.

- To open the ‘Update database’ dialog, click ‘Update data’ command in the ‘File’ menu.

1. To edit categories definition, click the 'Set Categories Coordinates' button in the dialog. A category list will be shown at the right bottom. Click the 'Add' or 'Delete' button to add a new category or delete the selected category in the list.
2. To customize a category-category map, click 'Set Categories Map' button in the dialog. Choose a picture as the map. Any picture with bmp, jpg, gif format is supported. In the example, we choose a China map as the category-category map; because we partition all the sequences to several Chinese provinces.
3. To re-arrange objects in each intra-category graph or extra-category graph, click the 'Set Sequences Coordinates' button in the dialog. Then choose the dataset (DstID) and the category in the two drop down menus bellow. The dialog presents the sequence list in the selected category and illustrates the undirected graph to users graphically. All objects

Figure 18H. Re-arrange objects in each intra-category graph by mouse-dragging.



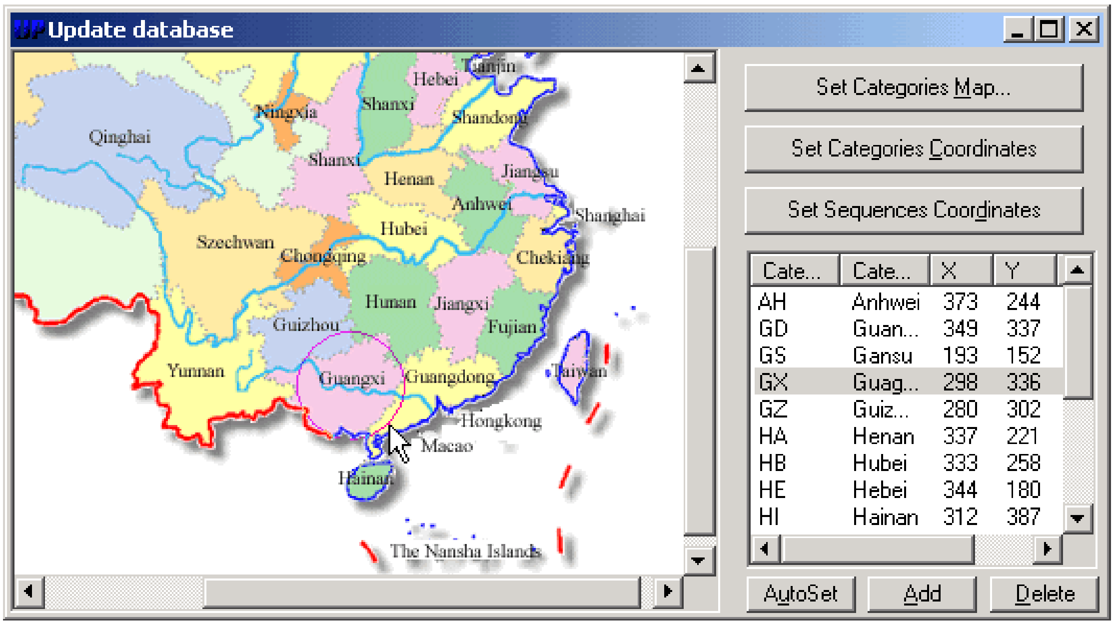
113 (1799.1298, 2005) 2 links in the category
Drag to change position, right click to show only related lines

DstID: HIV

Category: AH

D...	SID	C...	X	Y
35	35	AH	84	93
36	36	AH	190	544
37	37	AH	502	526
38	38	AH	436	361
39	39	AH	239	319

Figure 19H. Set hot spots for categories.
It is setting a hot spot for category 'Guangxi' in this figure.



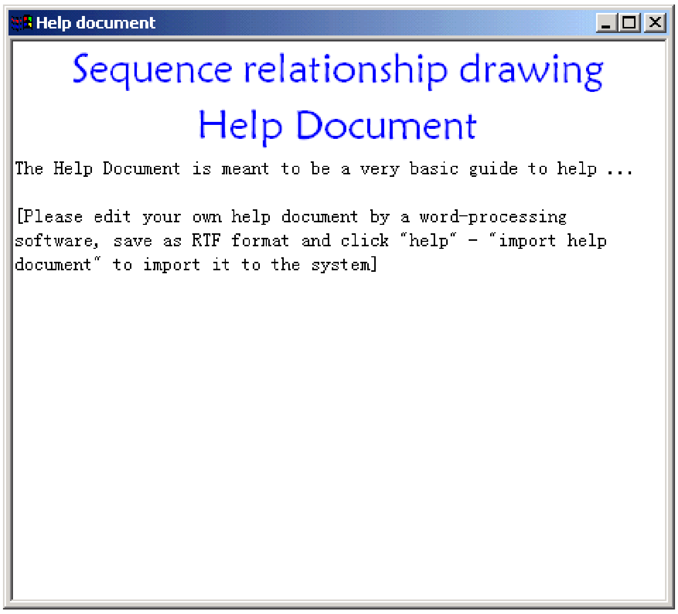
Cate...	Cate...	X	Y
AH	Anhwei	373	244
GD	Guan...	349	337
GS	Gansu	193	152
GX	Guag...	298	336
GZ	Guiz...	280	302
HA	Henan	337	221
HB	Hubei	333	258
HE	Hebei	344	180
HI	Hainan	312	387

coordinates can then be adjusted by users mouse-dragging in this graph. Click ‘save’ button to save all the changes to the database.

4. To re-arrange categories, i.e., set categories hot spots, click the ‘Set Categories Coordinates’ button in the dialog. Select a category in the list. Then click mouse at somewhere in the picture map to set the center of the hot spot, drag mouse to draw a proper circle to set the hot spot range for the category.

5. SRD allows importing a user-defined available RTF (Rich Text Format) file into the system as a help document, in which information appears as color-coded text. The RTF file can be generated by most word-processing software, such as Microsoft Word, Windows WordPad, Open Office, etc, and saved as a RTF file. The RTF format help file can then be further imported using the ‘Import help document’ function in the help menu. The help document can be displayed in another window by clicking the ‘help’ function in the help menu or press Ctrl+F1.

Figure 20H. A Help document displaying as RTF format.



Sequence relationship drawing

Help Document

The Help Document is meant to be a very basic guide to help ...

[Please edit your own help document by a word-processing software, save as RTF format and click "help" - "import help document" to import it to the system]

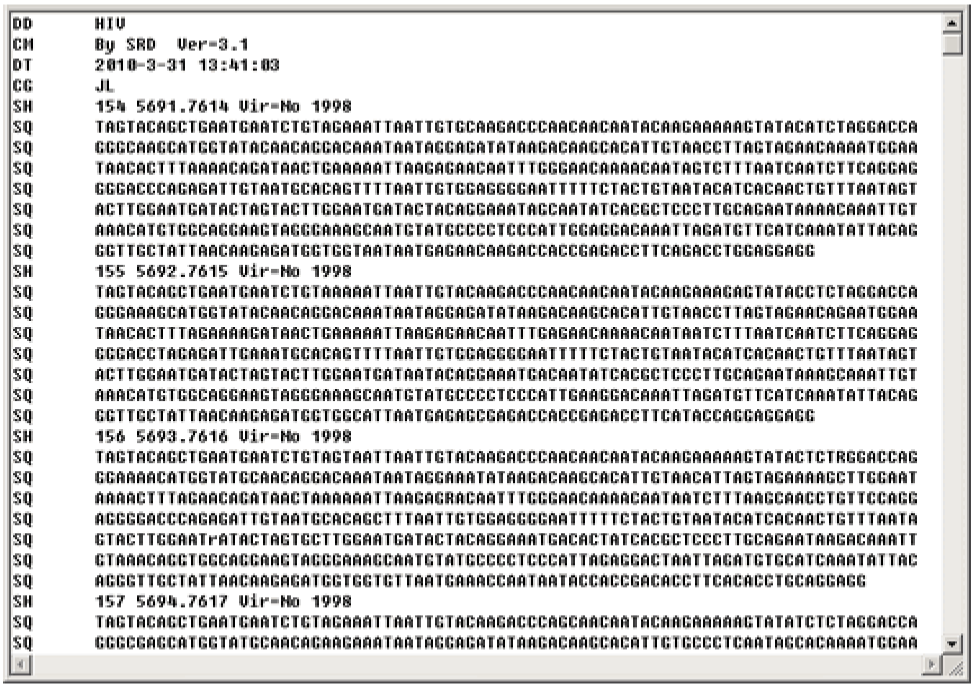
Results Export

Finally, for the purpose of input to other applications or the purpose of further analysis, the capacity to output results is essential. The output layer is able to directly write contents in all list windows to files, copy any graphs into the clipboard and save pictures as BMP (bitmap image format) files. These can be done by a simple mouse right-click in most windows of the main GUI, and choosing the corresponding functions in the pop-up menu.

To efficiently view the results generated by SRD, text output is also supported as complementary to the graphical drawings. A text file can be generated to describe current graphs by clicking the ‘Output Text Information’ function in the ‘edit’ menu. The text file is structured so as to be usable by human readers as well as by computer programs. It is composed of different types of lines, each with its own format, which are used to record the various types of data. Some files will not contain all of the line types, and some line types may occur many times in another file, this depends on the user’s selection in graph drawing window.

A terminator line (//) designates the end of the file.

Figure 21H. An output text file generated by SRD.



```
DD      HIV
CM      By SRD  Ver=3.1
DT      2010-3-31 13:41:03
CG      JL
SH      154 5691.7614 Vir=No 1998
SQ      TAGTACAGCTGAATGAATCTGTAGAAATTAATTGTGCAAGACCCAACAACAATACAAGAAAAAGTATACATCTAGGACCA
SQ      GGGCAAGCATGGTATACAACAGGACAAATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAAAATGGAA
SQ      TAACACTTTAAAACAGATAACTGAAAAATTAAGAGAACAATTTGGGAACAAAACAATAGTCTTTAATCAATCTTCAGGAG
SQ      GGGACCCAGAGATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATACATCACAACTGTTTAATAGT
SQ      ACTTGGAATGATACTAGTACTTGGAATGATACTACAGGAAATAGCAATATCACGCTCCCTTGCAGAATAAAACAAATTGT
SQ      AAACATGTGGCAGGAAGTAGGGAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGATGTTCATCAAATATTACAG
SQ      GGTTGCTATTAACAAGAGATGGTGGTAATAATGAGAACAAGACCACCGAGACCTTCAGACCTGGAGGAGG
SH      155 5692.7615 Vir=No 1998
SQ      TAGTACAGCTGAATGAATCTGTAAAAATTAATTGTACAAGACCCAACAACAATACAAGAAAGAGTATACCTCTAGGACCA
SQ      GGGAAAGCATGGTATACAACAGGACAAATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAGAATGGAA
SQ      TAACACTTTAGAAAAGATAACTGAAAAATTAAGAGAACAATTTGAGAACAAAACAATAATCTTTAATCAATCTTCAGGAG
SQ      GGGACCTAGAGATTGAAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATACATCACAACTGTTTAATAGT
SQ      ACTTGGAATGATACTAGTACTTGGAATGATAATACAGGAAATGACAATATCACGCTCCCTTGCAGAATAAAGCAAATTGT
SQ      AAACATGTGGCAGGAAGTAGGGAAAGCAATGTATGCCCCTCCCATTGAAGGACAAATTAGATGTTCATCAAATATTACAG
SQ      GGTTGCTATTAACAAGAGATGGTGGCATTAATGAGAGCGAGACCACCGAGACCTTCATACCAGGAGGAGG
SH      156 5693.7616 Vir=No 1998
SQ      TAGTACAGCTGAATGAATCTGTAGTAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACTCTRGGACCAG
SQ      GGAAAACATGGTATGCAACAGGACAAATAATAGGAAATATAAGACAAGCACATTGTAACATTAGTAGAAAAGCTTGGAAT
SQ      AAAACTTTAGAACAGATAACTAAAAAATTAAGAGRACAATTTGGGAACAAAACAATAATCTTTAAGCAACCTGTTCCAGG
SQ      AGGGGACCCAGAGATTGTAATGCACAGCTTTAATTGTGGAGGGGAATTTTTCTACTGTAATACATCACAACTGTTTAATA
SQ      GTACTTGGAATrATACTAGTGCTTGGAATGATACTACAGGAAATGACACTATCACGCTCCCTTGCAGAATAAGACAAATT
SQ      GTAAACAGGTGGCAGGAAGTAGGGAAAGCAATGTATGCCCCTCCCATTAGAGGACTAATTAGATGTGCATCAAATATTAC
SQ      AGGGTTGCTATTAACAAGAGATGGTGGTGTTAATGAAACCAATAATACCACCGACACCTTCACACCTGCAGGAGG
SH      157 5694.7617 Vir=No 1998
SQ      TAGTACAGCTGAATGAATCTGTAGAAATTAATTGTACAAGACCCAGCAACAATACAAGAAAAAGTATATCTCTAGGACCA
SQ      GGGCGAGCATGGTATGCAACAGAAGAAATAATAGGAGATATAAGACAAGCACATTGTGCCCTCAATAGCACAAAATGGAA
```

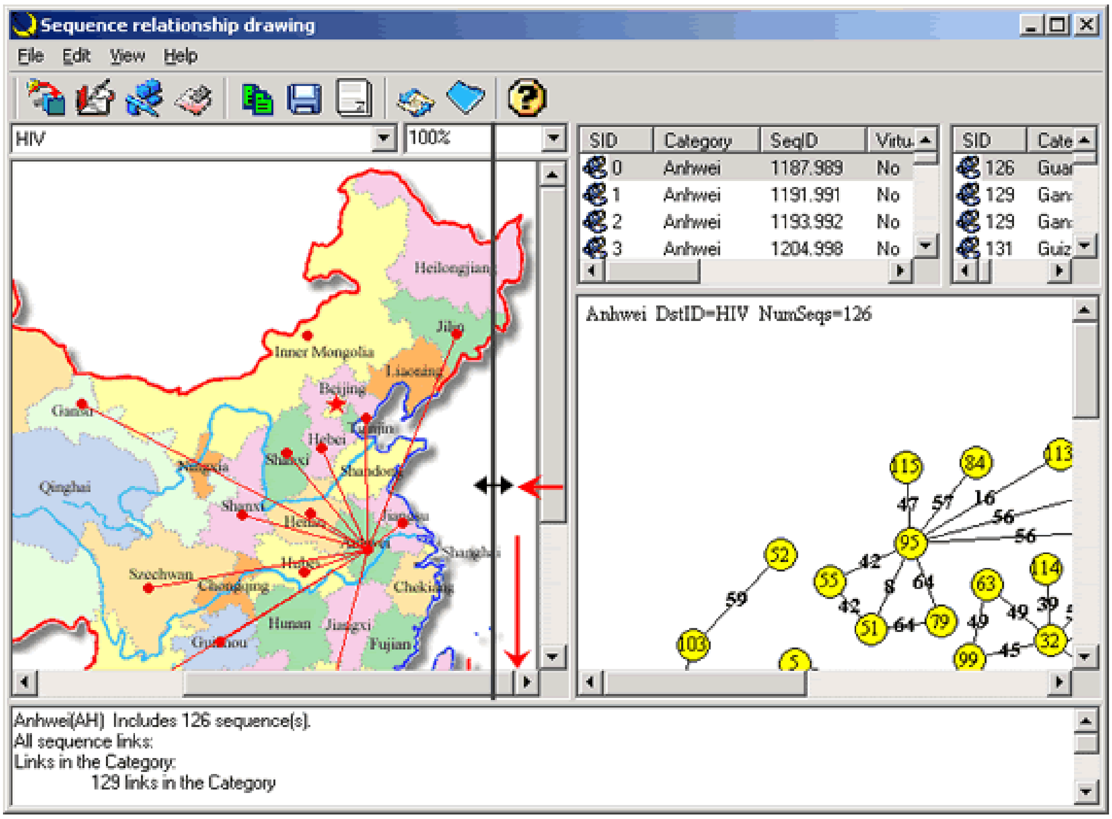
A text file generated by SRD is composed of different types of lines.

Types of lines	Description
DD	Current dataset ID (DstID).
CM	Comments.
DT	Date and time when generate this file.
CG	Current category.
SH	Sequence header (a sequence description followed by lines of sequence data).
SQ	Lines of sequence data.
LS	The links is related to all sequence or to the selected sequence.
LH	Link header (a link description followed link information)
LK	Link Information including the pair of two sequences with their relationship value.
//	End of file.

Splitting sub-windows in different views

The main GUI of SRD composed of several sub-windows. Users are enabled to drag the common border of each two adjacent windows to resize each other. If the dragging horizontal border up and down, or dragging vertical border left and right, the border's orientation can be

Figure 21H. Drag border to resize sub-windows and re-arrange windows layout.



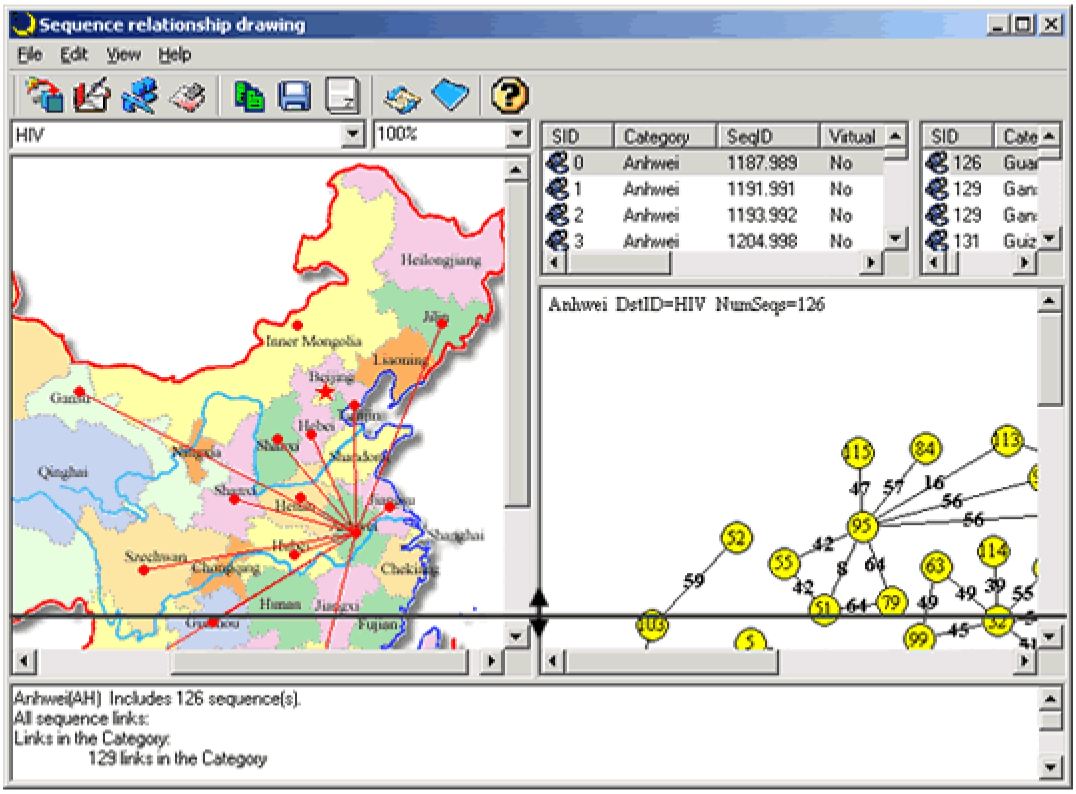
SID	Category	SeqID	Virtu
0	Anhwei	1187.989	No
1	Anhwei	1191.991	No
2	Anhwei	1193.992	No
3	Anhwei	1204.998	No

SID	Cate
126	Guar
129	Gan
129	Gan
131	Guiz

Anhwei DstID=HIV NumSeqs=126

Anhwei(AH) Includes 126 sequence(s).
All sequence links:
Links in the Category:
129 links in the Category

Figure 22H. The border can be draged from vertical to horizontal to re-arrange windows layout.



SID	Category	SeqID	Virtual
0	Anhwei	1187.989	No
1	Anhwei	1191.991	No
2	Anhwei	1193.992	No
3	Anhwei	1204.998	No

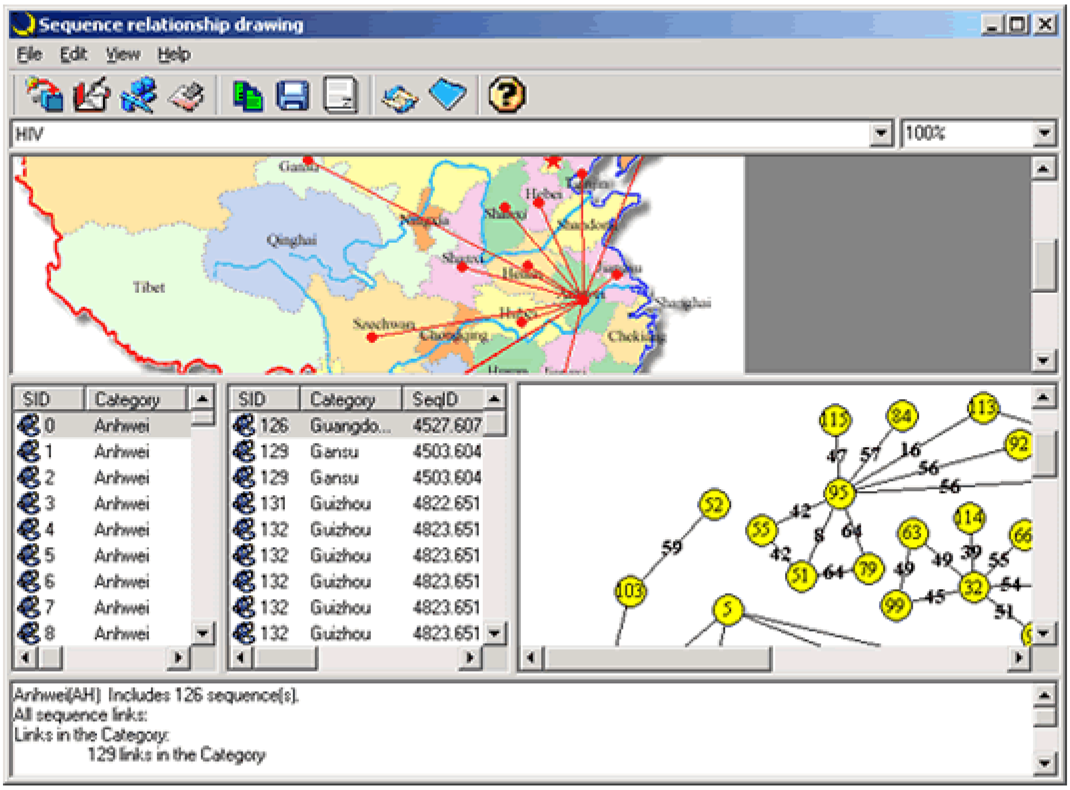
SID	Cate
126	Guar
129	Gan
129	Gan
131	Guiz

Anhwei DstID=HIV NumSeqs=126

Anhwei(AH) Includes 126 sequence(s).
All sequence links:
Links in the Category:
129 links in the Category

changed from horizontal to be vertical, or vice versa. This can be used to re-arrange the sub-windows layout of SRD.

Figure 23H. A new re-arranged sub-windows layout.



SID	Category
0	Anhwei
1	Anhwei
2	Anhwei
3	Anhwei
4	Anhwei
5	Anhwei
6	Anhwei
7	Anhwei
8	Anhwei

SID	Category	SeqID
126	Guangdo...	4527.607
129	Gansu	4503.604
129	Gansu	4503.604
131	Guizhou	4822.651
132	Guizhou	4823.651
132	Guizhou	4823.651
132	Guizhou	4823.651
132	Guizhou	4823.651
132	Guizhou	4823.651

Anhwei(AH) Includes 126 sequence(s).
All sequence links:
Links in the Category:
129 links in the Category