PHYLOViZ Documentation

Release 2.0

PHYLOViZ Team

September 07, 2016

Contents

PHYLOViZ is a platform independent JAVA software that allows the analysis of sequence-based typing methods that generate allelic profiles and their associated epidemiological data.

CHAPTER 1

Download and install

PHYLOViZ core and several plugins are free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see http://www.gnu.org/licenses/.

Certain source files distributed by the PHYLOViZ Team are under the terms of the GNU General Public License Version 3 with the following clarification and special exception, but only where PHYLOViZ Team has expressly included it in the particular source file's header.

```
Linking this library statically or dynamically with other modules is
making a combined work based on this library.  Thus, the terms and
conditions of the GNU General Public License cover the whole combination.

As a special exception, the copyright holders of this library give you
permission to link this library with independent modules to produce an
executable, regardless of the license terms of these independent modules,
and to copy and distribute the resulting executable under terms of your
choice, provided that you also meet, for each linked independent module,
the terms and conditions of the license of that module.  An independent
module is a module which is not derived from or based on this library.
If you modify this library, you may extend this exception to your version
of the library, but you are not obligated to do so.  If you do not wish
to do so, delete this exception statement from your version.
```

Code licensed under this license may be reused in commercial products provided that changes made directly in the sources - bug fixes or enhancements - must be contributed back to PHYLOViZ, but new source files (as in new plugins) which you write that link to PHYLOViZ code do not need to be.

Choose the appropriate version for your operating system or the .jar file. The OS specific versions already contain some memory specific parameters to enhance the software performance when using large datasets.

See details about available plugins and the licenses under which they are covered.

1.1 Binaries

A cross-platform zip distribution package is available.

Just unzip the package, enter the created directory and in the sub-directory `bin/` run `phyloviz.exe` or `phyloviz64.exe` (Windows) or 'phyloviz' (Linux/MacOS) accordingly to your operating system.

NOTE: You may need to adjust some parameters in `etc/phyloviz.conf` with respect to memory usage. These settings have a strong impact on visualization features. For instance, in Windows, you may achieve better results with:

```
default_options="--branding phyloviz -J-Xss8M -J-Xms32m -J-Xmx1024M --laf javax.swing.plaf.metal.Meta
```

IMPORTANTE NOTICE: After installing always go to the "Help" menu and "Check for updates" to install any novel plugins or latest updates to PHYLOViZ software. The SNP analysis plugin is installed in this way to demonstrate the plugin capability.

1.2 Source

All the Source code is available in the new code repository for in bitbucket.org. Check it out at https://bitbucket.org/phyloviz/phyloviz-main.

PHYLOViZ is built on top of the NetBeans Platform, thus we recommend NetBeans for the development of new plugins.

CHAPTER 2

Loading data

2.1 File formats

To be able to analize and visualize your data, PHYLOViZ needs two separate files: One file contains the allelic profile data of the method you are using (Typing Data), while the other will contain accessory data (Isolate Data). In the example image below they are *sampleAPfile.txt* and sampleADfile.txt_respectively.

sampleAPfile.txt

ST	gki	gtr	murI	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	13	9
14	10	4	7	6	12	13	8
15	3	3	2	2	9	8	2
16	4	4	1	2	17	1	2
17	4	4	1	2	17	6	2
18	4	2	4	1	8	7	2
19	3	8	4	1	8	7	2
20	3	3	2	8	9	6	6
21	3	8	2	2	9	8	2
22	3	3	2	8	1	11	6
23	3	3	4	1	3	1	1
24	3	2	1	5	15	4	3
25	3	2	1	5	7	4	3
26	3	2	1	1	7	10	2
27	2	2	4	1	13	12	7
28	3	3	4	2	16	14	2
29	3	2	4	2	7	1	3
30	3	2	4	1	7	10	2
31	3	2	4	1	8	7	2
32	3	2	4	1	4	10	5
33	3	8	2	8	9	6	6
34	3	7	4	1	14	15	10
35	3	7	4	1	14	15	2
36	4	4	1	2	17	6	3
37	3	2	1	1	4	1	3
38	1	1	1	1	1	1	1
39	1	1	1	1	1	21	4
40	1	1	1	4	1	1	4
41	1	1	1	9	1	1	1
42	1	1	4	1	1	1	4
43	2	2	4	1	1	1	2
44	2	2	4	2	3	7	1
45	2	2	4	10	8	19	2
46	2	4	4	1	19	17	6
47	3	2	1	1	20	1	3
48	3	2	1	2	10	4	2
49	3	2	3	1	5	5	2
50	3	2	3	1	5	18	2
51	3	2	4	1	11	1	2
52	3	2	4	1	11	3	5

sampleADfile.txt

Strain	emm type	Group	carbohydrate	ST	Location	Collection
168554	stG485 G	47	Portugal	UL		
171712	stG480 G	38	Portugal	UL		
220269	stG2078 G	15	Portugal	UL		
223754	stC839 C	3	Portugal	UL		
230631	stG480 G	8	Portugal	UL		
231995	stC74a G	29	Portugal	UL		
241940	stC36 C	50	Portugal	UL		
273600	stG166b G	65	Portugal	UL		
299298	stG643 G	8	Portugal	UL		
313247	stG6 G	25	Portugal	UL		
363962	stG2078 G	17	Portugal	UL		
378119	stC839 G	15	Portugal	UL		
380870	stG480 G	41	Portugal	UL		
386041	stC839 C	3	Portugal	UL		
394314	stG2078 G	72	Portugal	UL		
423738	stG62647 C	20	Portugal	UL		
450784	stG10 G	15	Portugal	UL		
460880	stG10 G	15	Portugal	UL		
493188	stG485 C	69	Portugal	UL		
542567	stG6 G	62	Portugal	UL		
618280	emm57 G	57	Portugal	UL		
SH0004	stG6792 G	4	Portugal	UL		
SH0015	stG6 G	25	Portugal	UL		
SH0032	stG166b G	15	Portugal	UL		
SH0102	stG2078 G	17	Portugal	UL		
SH0107	stG643 G	52	Portugal	UL		
SH0110	stG6 G	25	Portugal	UL		
SH0113	stG6792 G	4	Portugal	UL		
SH0124	stG6792 G	4	Portugal	UL		
SH0218	stG245 G	15	Portugal	UL		
SH0254	stG485 C	69	Portugal	UL		
SH0257	stC6979 C	80	Portugal	UL		
SH0259	stG652 G	71	Portugal	UL		
SH0275	stG485 G	55	Portugal	UL		
SH0330	stC36 C	49	Portugal	UL		
SH0336	stG5420 G	25	Portugal	UL		
G121	stC74a G	29	Australia	QIMR		
G122	stC74a G	29	Australia	QIMR		
GCS10128	stC1400 C	46	Australia	QIMR		
GCS2816	stG62647 C	20	Australia	QIMR		
GCS6894	stG62647 C	20	Australia	QIMR		
GCS6929	stG62647 C	20	Australia	QIMR		
GGS075	stG166b G	56	Australia	QIMR		
GGS101	stG643 G	12	Australia	QIMR		
GGS10b	stG6 G	44	Australia	QIMR		
GGS11172	stC74a G	29	Australia	QIMR		
GGS11543	stG643 G	12	Australia	QIMR		
GGS120	stG4831 G	74	Australia	QIMR		
GGS19	stC1400 G	64	Australia	QIMR		
GGS2	stG10 G	15	Australia	QIMR		
GGS24	stG6 G	44	Australia	QIMR		
GGS430	stG643 G	12	Australia	QIMR		

The Typing data should be a tab separated file containing the allelic profiles, formatted as follows: the first line should contain the column headers (usually locus identifiers be it either SNP, MLST or cg/wgMLST locus). The first column should be the allelic profile identifier (for MLST this would be the Sequence Type number, for any other method could be an unique strain ID. however if two strains have the same profile they should be given the same ID). The following columns are the loci used in the analysis.

If the Isolate data file is not used, the Typing data file should also represent the number of repeated profiles in a

collection, that is to say that if a given profile appears in a collection n times it should be repeated in the Typing data file n times.

In case of an Isolate data file is used the frequency of each type will be represented by the number of entries with a given Sequence type, in the Isolate file only and the frequency represented by repeated profiles in the Typing data file will not be used.

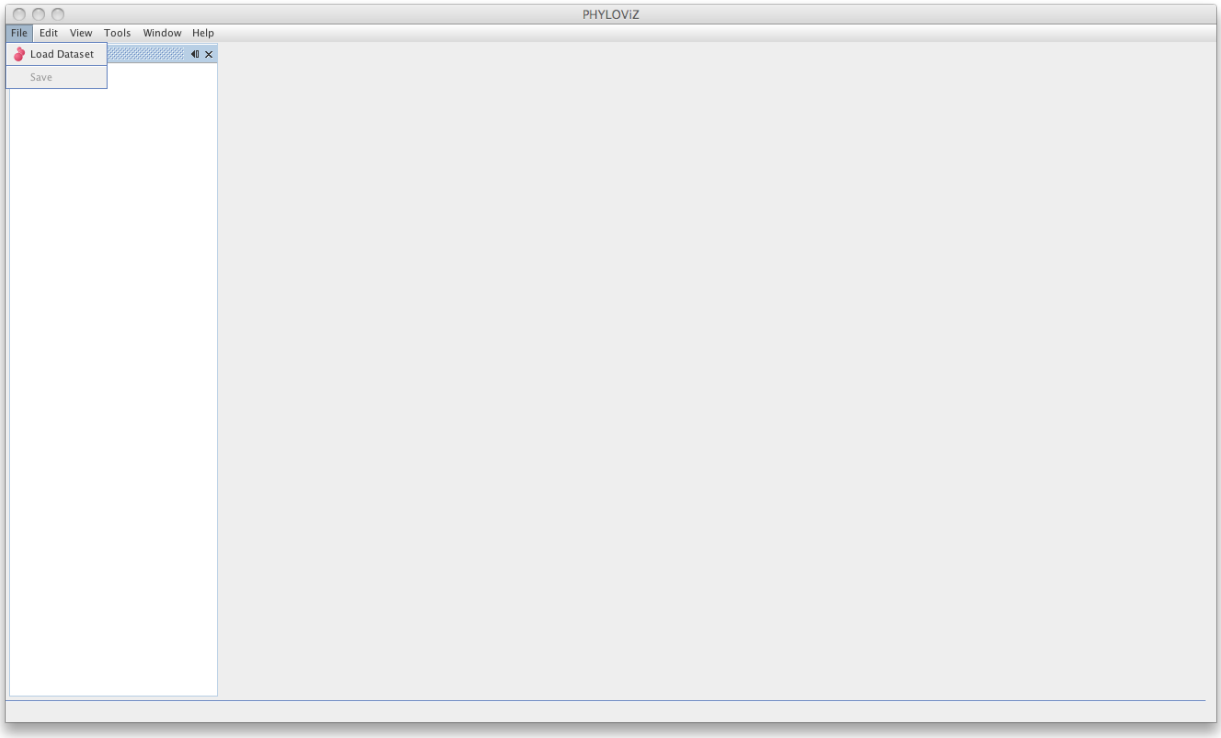
You can find an example of MLST data correctly formatted here. Note that in this file several STs are represented by more than one isolate (e.g. ST3 was found in 6 isolates).

The Isolate data file can contain epidemiological and/or demographic data or any other data you want to visualize overlaid onto the results of the analysis algorithms. The link between the data in the two files is made by the Sequence Type identifier. You can find an example file correctly formatted here.

2.2 Loading a Dataset

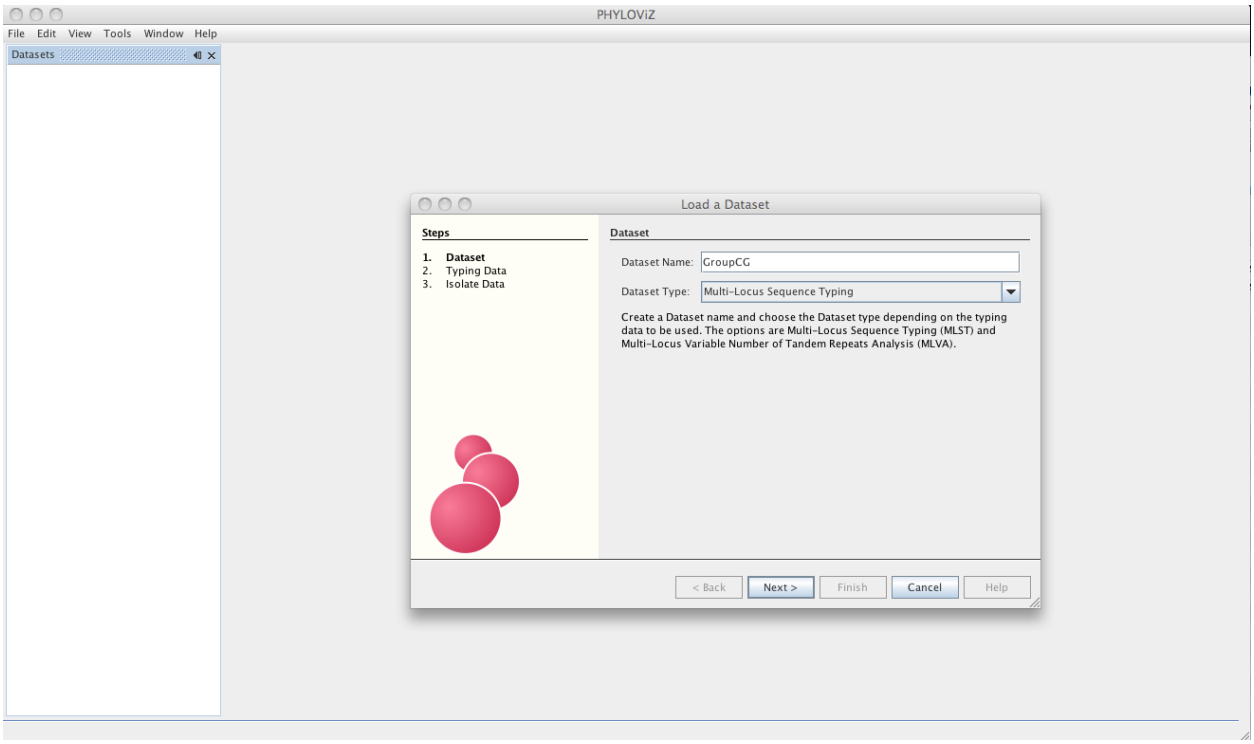
Go to *File* menu and choose *Load Dataset*.

If any errors in the data loading process are found they will be displayed in the session *Tab*. In the following screenshot you can see an example where allelic profiles were repeated with different identifiers. In the example data,we created ST81 as copy of ST1 profile and PHYLOViZ detects it and eliminates it from the analysis.

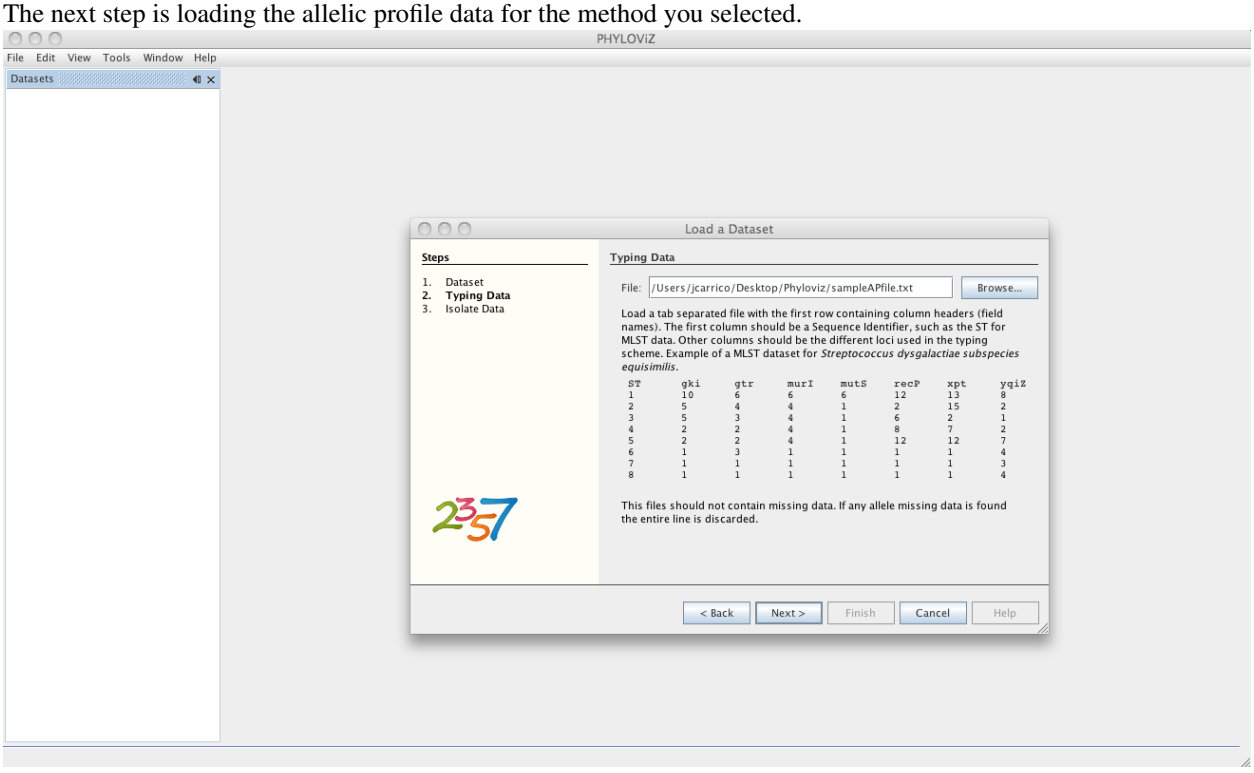


The dialog will now guide the user in the loading of the data. The first step is choosing a name for your Dataset since now PHYLOViZ supports multiple datasets open simultaneously. You must also choose the Dataset Type from the dropbox menu.

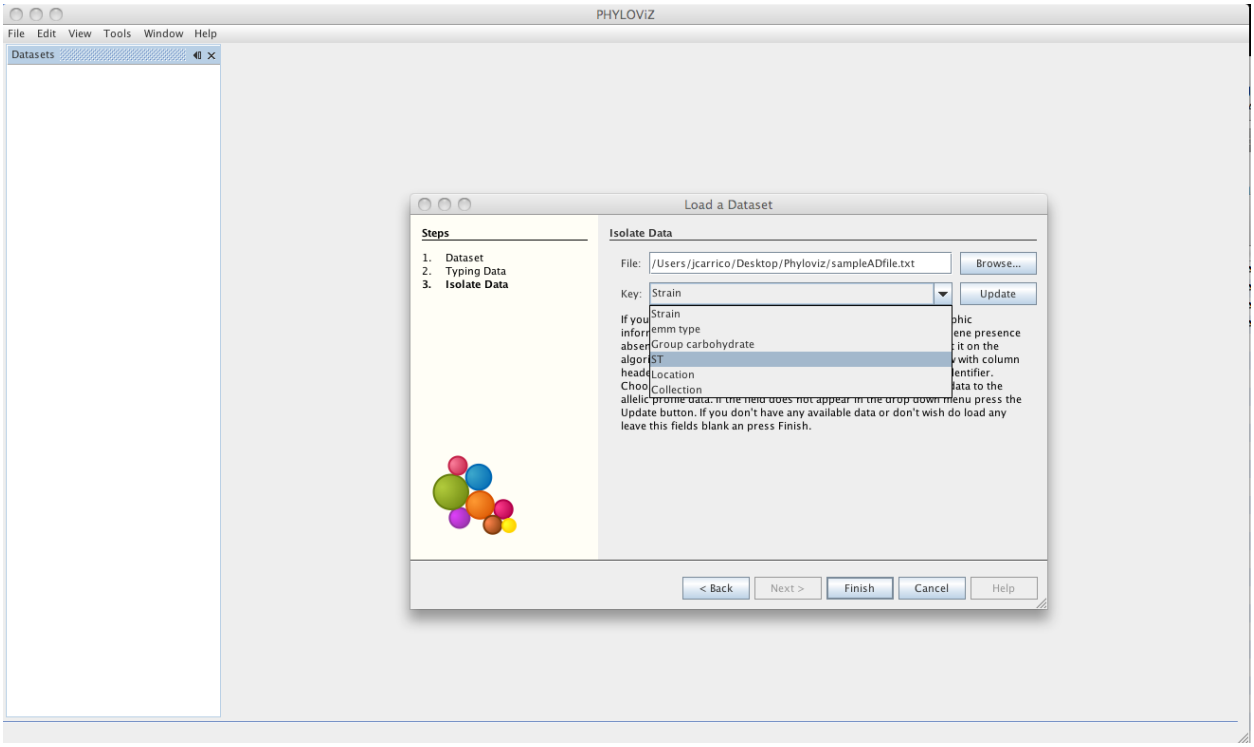
The Dataset type can be MLST or MLVA datasets with any number of loci, without any missing data. Lines with missing data will be excluded on load. If you have installed the Single Nucleotide Polymorphism (SNP) plugin, you can also access it on the Dataset type. See the Sample Datasets page to access some test data for the sequence-based typing methods available.



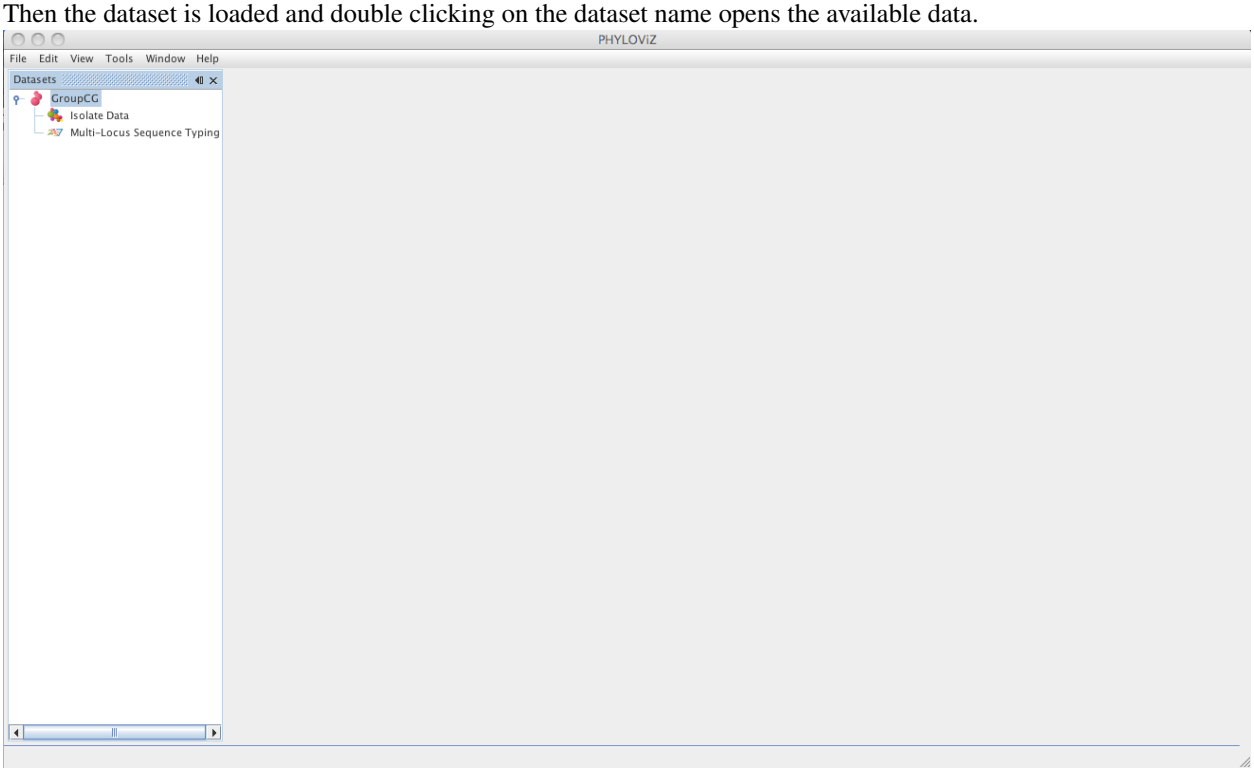
The next step is loading the allelic profile data for the method you selected.



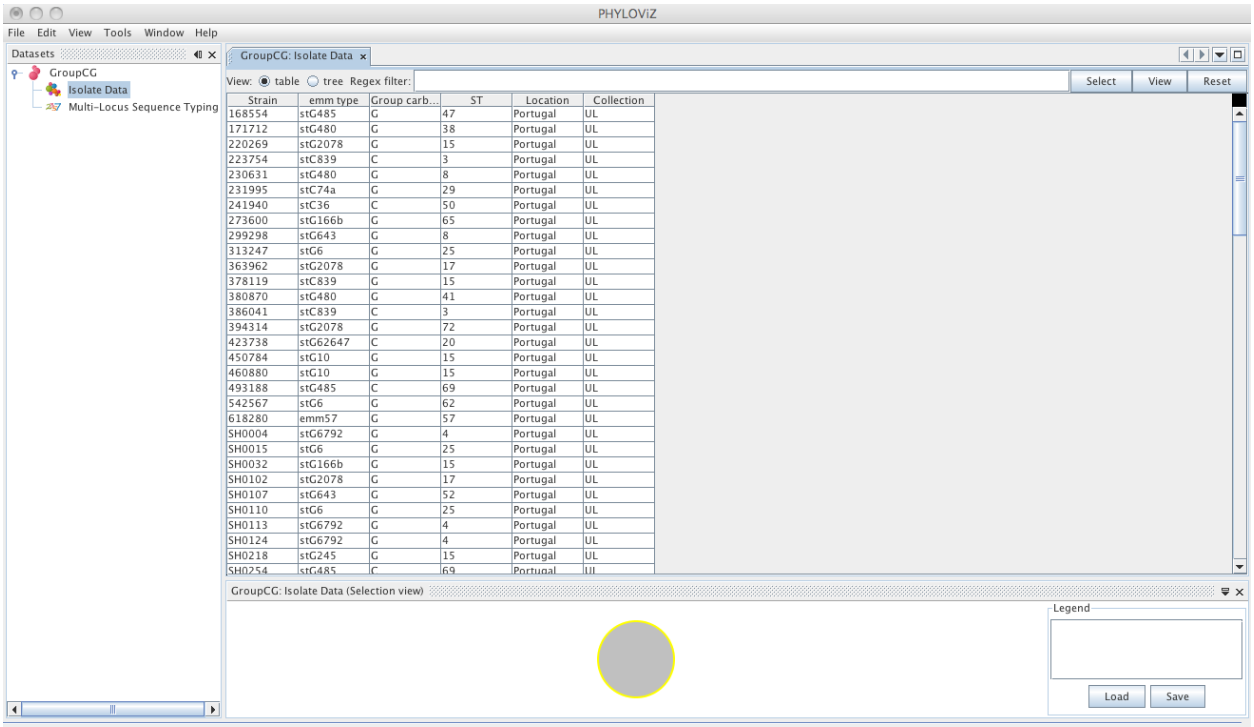
After loading the allelic profile data, you can choose a file with information on your isolates for which the allelic profile was loaded. The linking field, as explained before, should be the Sequence Identifier and should be selected in the Key dropdown menu.

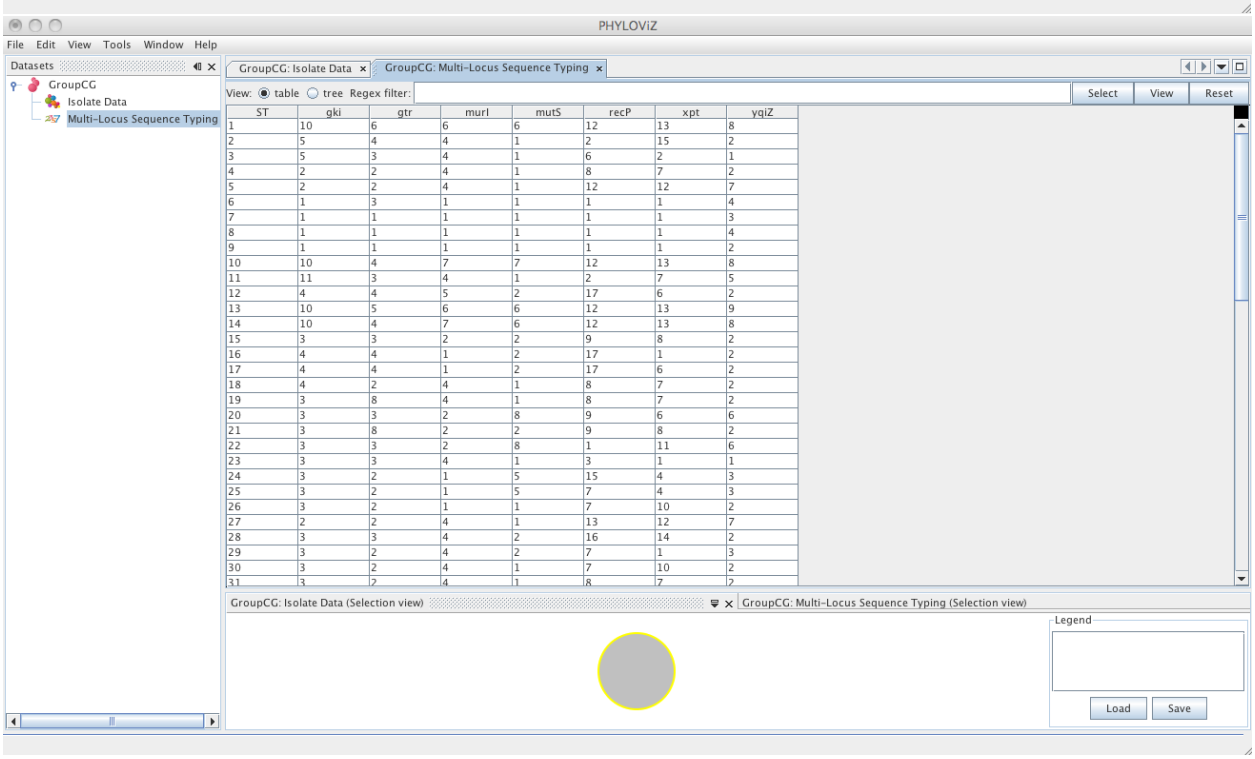


Then the dataset is loaded and double clicking on the dataset name opens the available data.

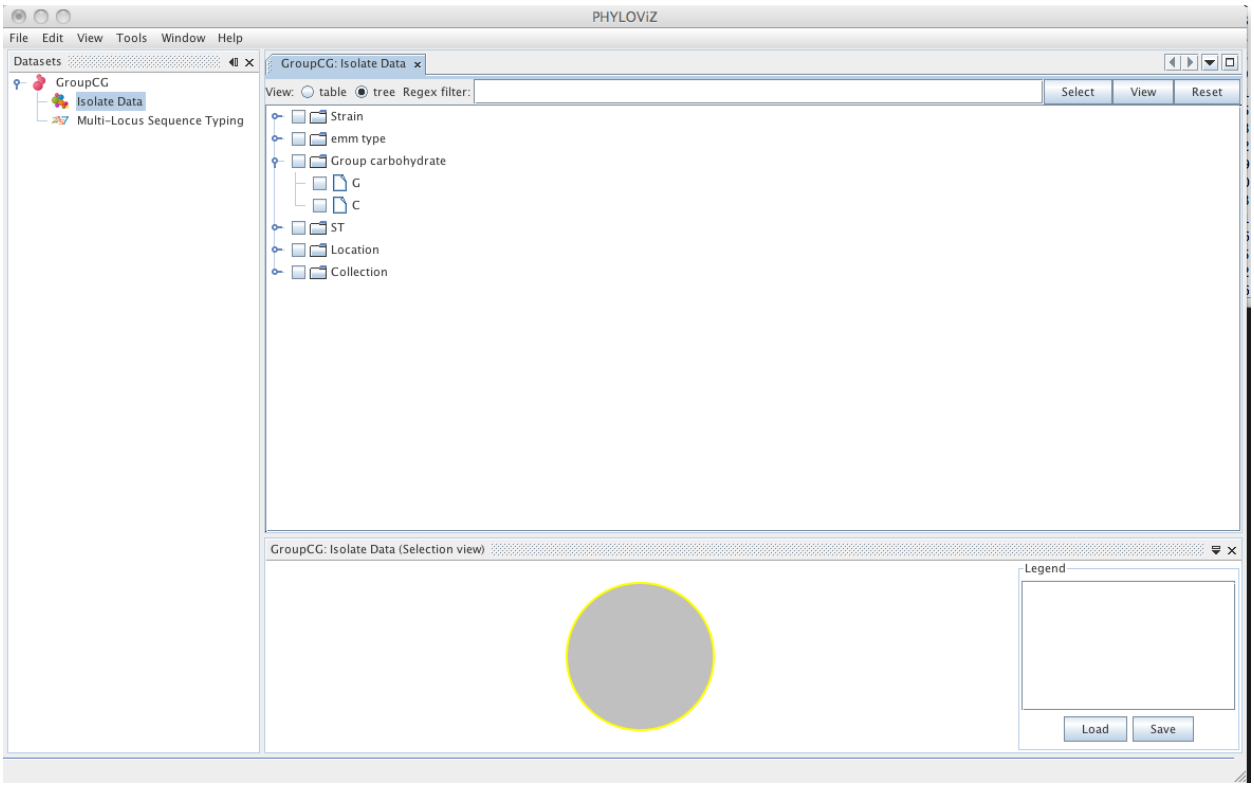


Double clicking on *Isolate Data* and *Typing Data* in the tree menu under the dataset name opens the respective tabs.



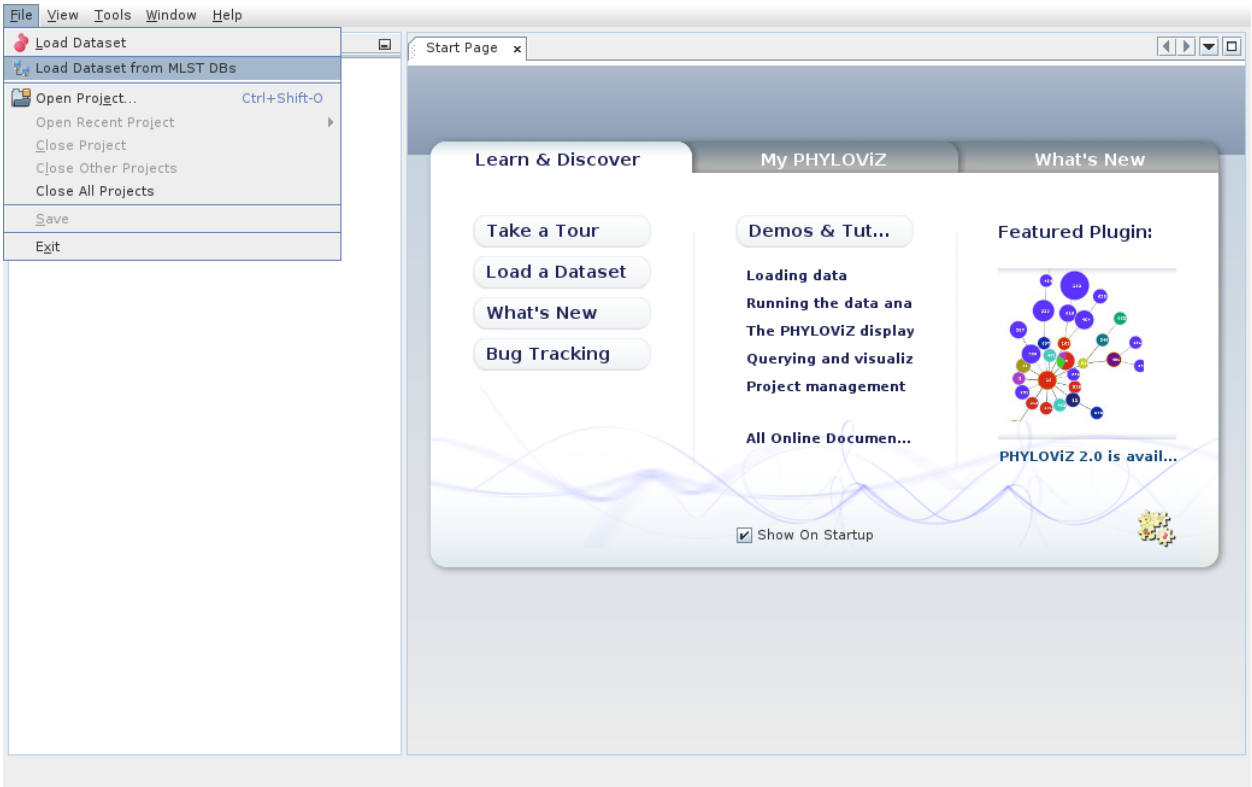


The default view is the *table* view. Also available is the *tree* view, where it is easier to visualize what information is available in the different fields and to select combinations of fields with specific values.

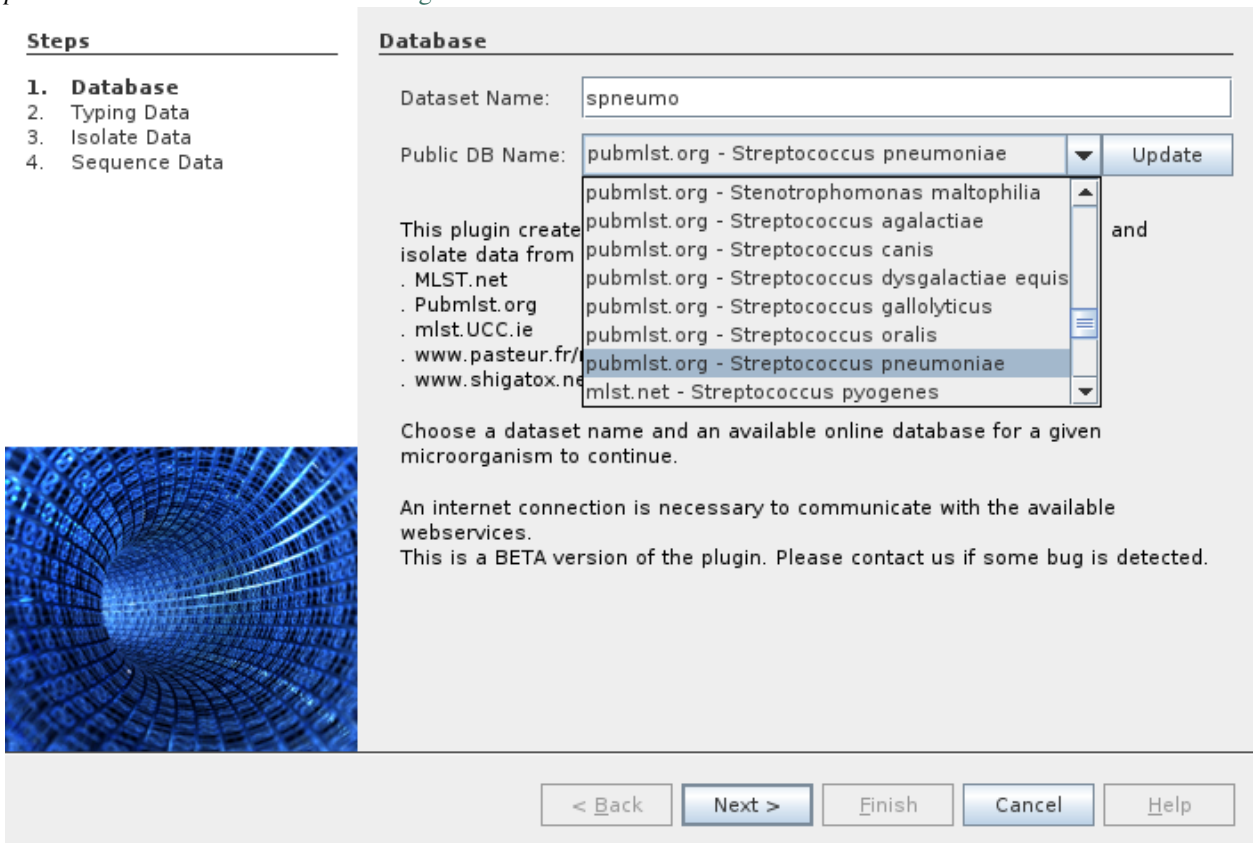


2.3 Loading a remote Dataset

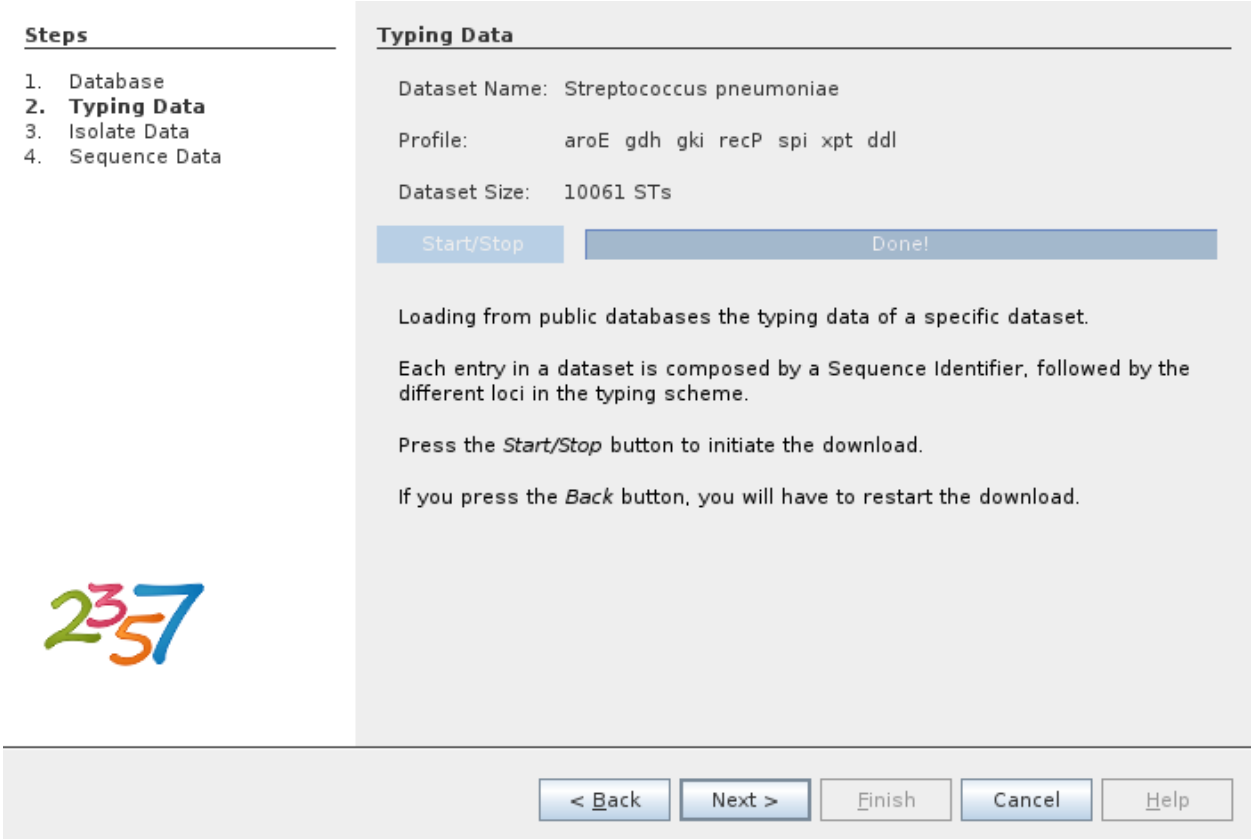
We can also load datasets from remote databases and services. PHYLOViZ contains already a list of available databases. We can choose *Load Dataset from MLST DBs*.



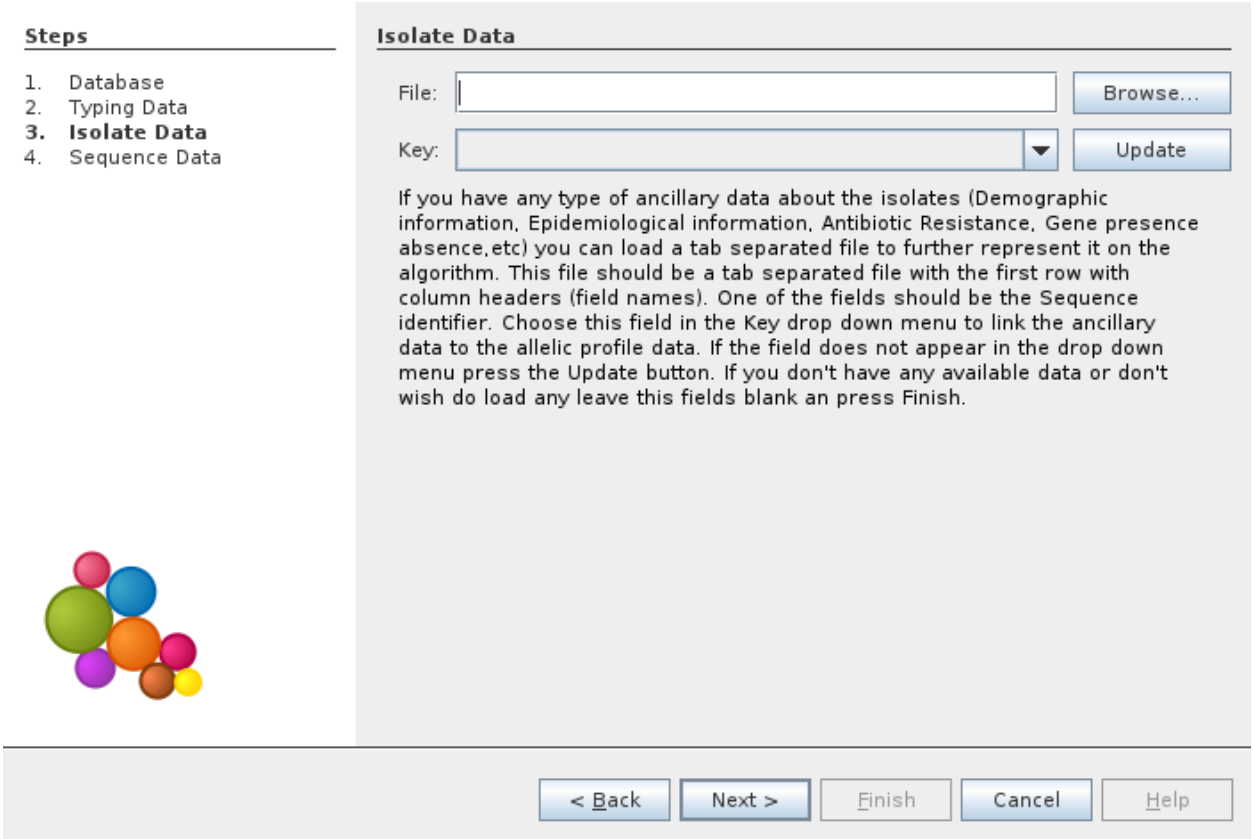
There are several datasets available from several providers. In the following example we select the *Streptococcus pneumoniae* dataset from PubMLST.org.



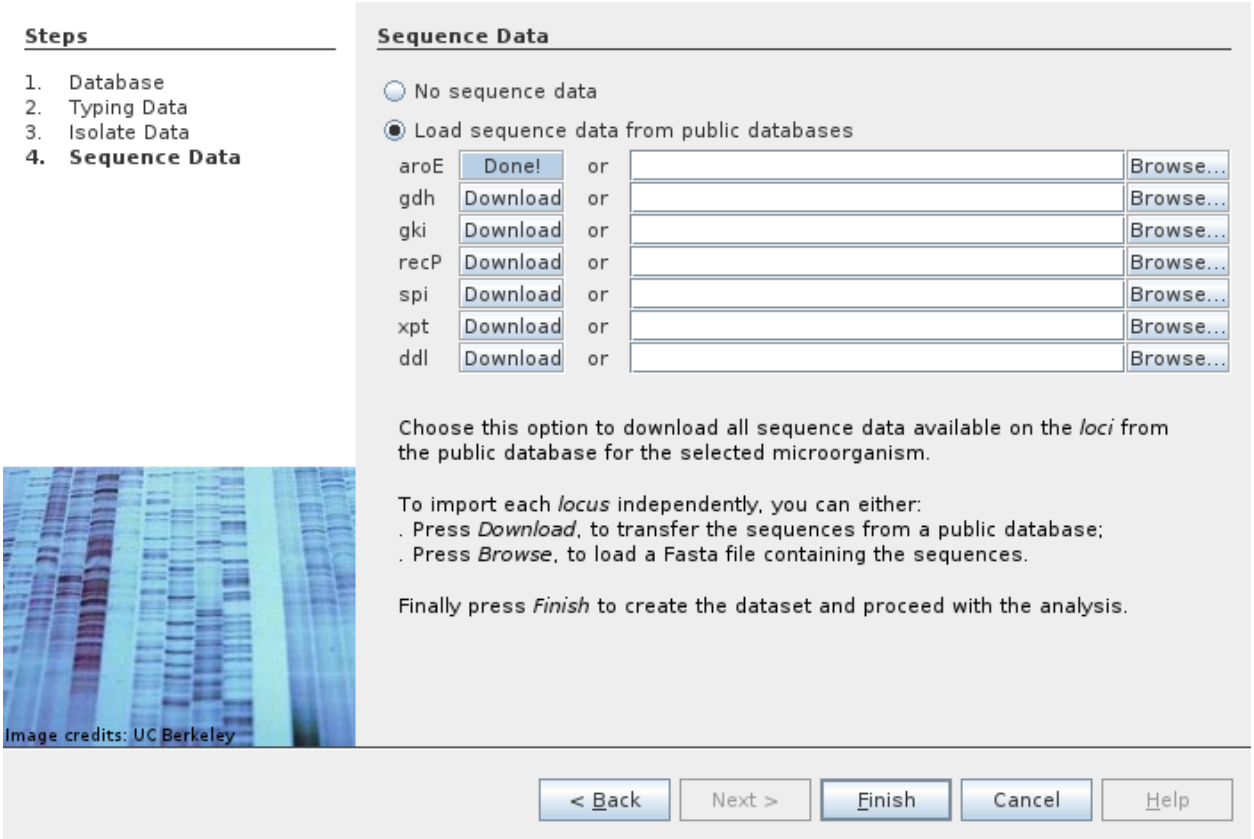
The next step is to download the dataset.



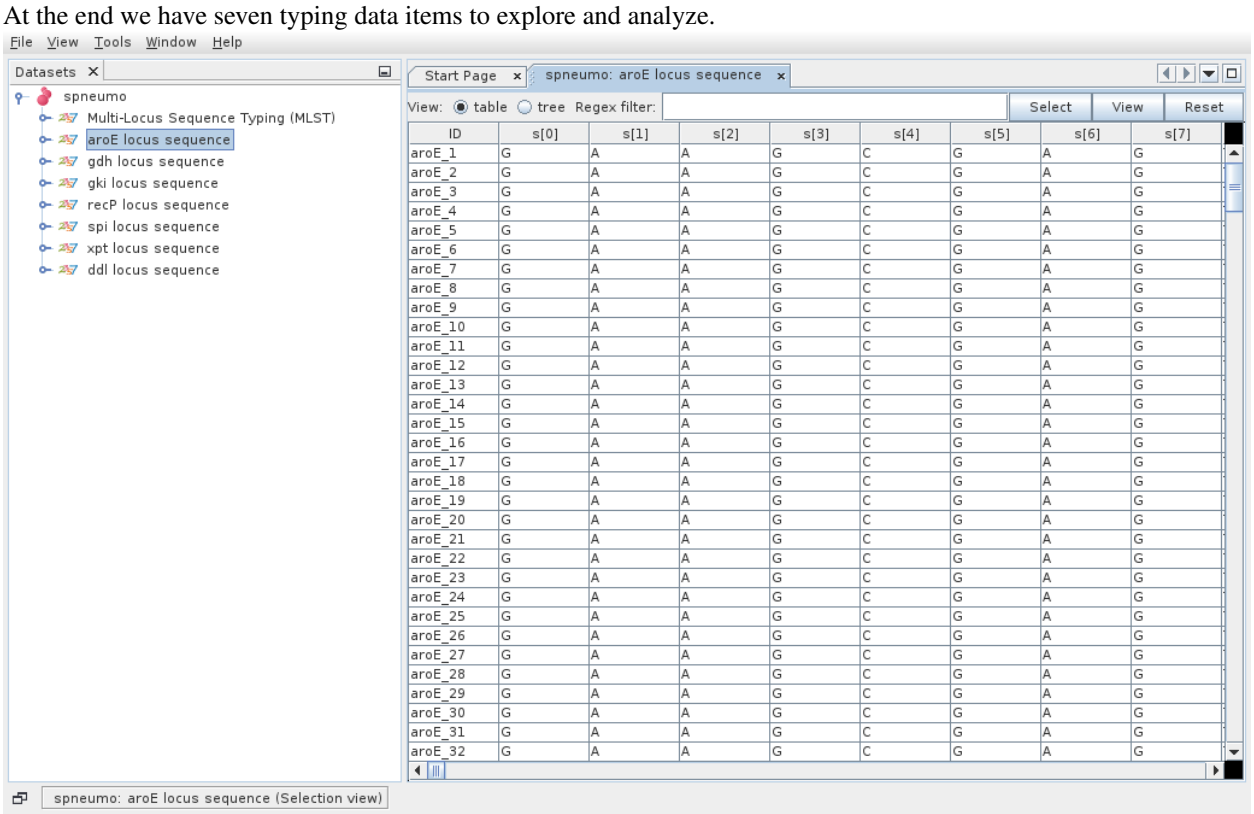
In the next window we can load ancillary data on isolates. In this example we choose to not load any data.



We can also load sequence data for each allele. They are downloaded individually and loaded as typing data.



At the end we have seven typing data items to explore and analyze.

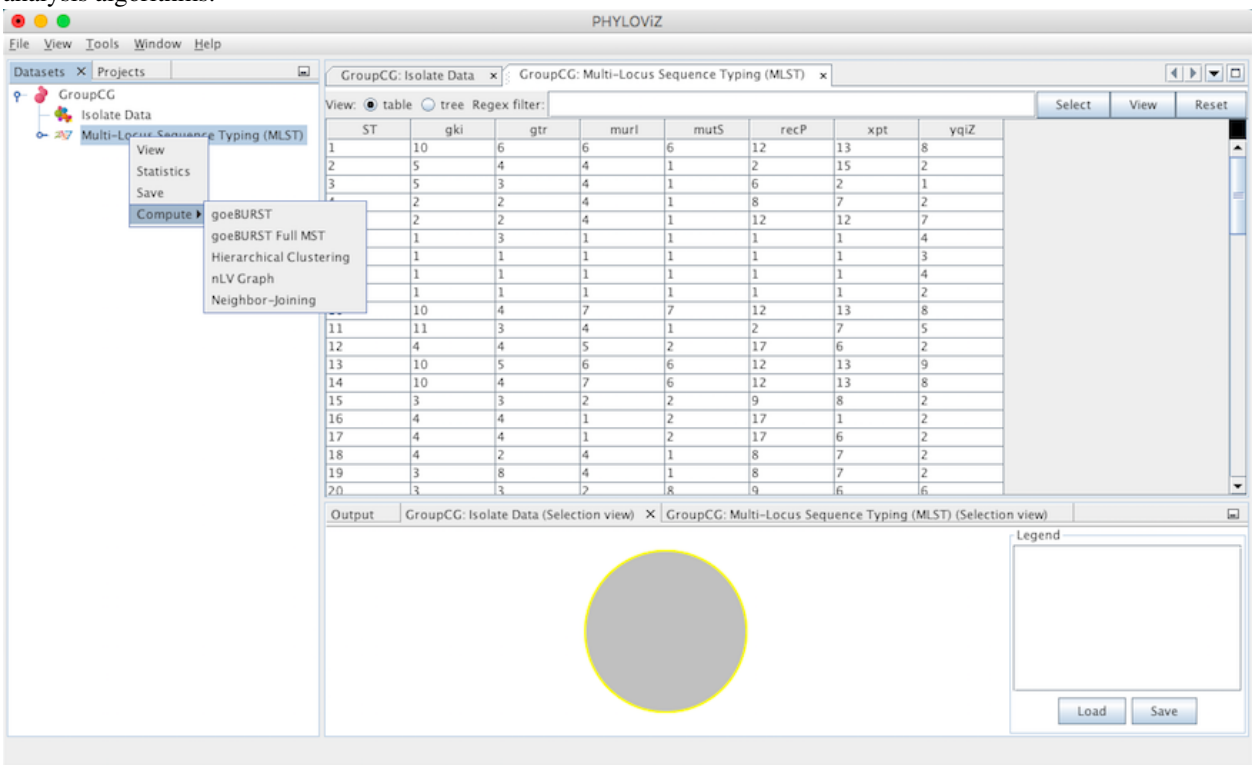


ID	s[0]	s[1]	s[2]	s[3]	s[4]	s[5]	s[6]	s[7]
aroE_1	G	A	A	G	C	G	A	G
aroE_2	G	A	A	G	C	G	A	G
aroE_3	G	A	A	G	C	G	A	G
aroE_4	G	A	A	G	C	G	A	G
aroE_5	G	A	A	G	C	G	A	G
aroE_6	G	A	A	G	C	G	A	G
aroE_7	G	A	A	G	C	G	A	G
aroE_8	G	A	A	G	C	G	A	G
aroE_9	G	A	A	G	C	G	A	G
aroE_10	G	A	A	G	C	G	A	G
aroE_11	G	A	A	G	C	G	A	G
aroE_12	G	A	A	G	C	G	A	G
aroE_13	G	A	A	G	C	G	A	G
aroE_14	G	A	A	G	C	G	A	G
aroE_15	G	A	A	G	C	G	A	G
aroE_16	G	A	A	G	C	G	A	G
aroE_17	G	A	A	G	C	G	A	G
aroE_18	G	A	A	G	C	G	A	G
aroE_19	G	A	A	G	C	G	A	G
aroE_20	G	A	A	G	C	G	A	G
aroE_21	G	A	A	G	C	G	A	G
aroE_22	G	A	A	G	C	G	A	G
aroE_23	G	A	A	G	C	G	A	G
aroE_24	G	A	A	G	C	G	A	G
aroE_25	G	A	A	G	C	G	A	G
aroE_26	G	A	A	G	C	G	A	G
aroE_27	G	A	A	G	C	G	A	G
aroE_28	G	A	A	G	C	G	A	G
aroE_29	G	A	A	G	C	G	A	G
aroE_30	G	A	A	G	C	G	A	G
aroE_31	G	A	A	G	C	G	A	G
aroE_32	G	A	A	G	C	G	A	G

CHAPTER 3

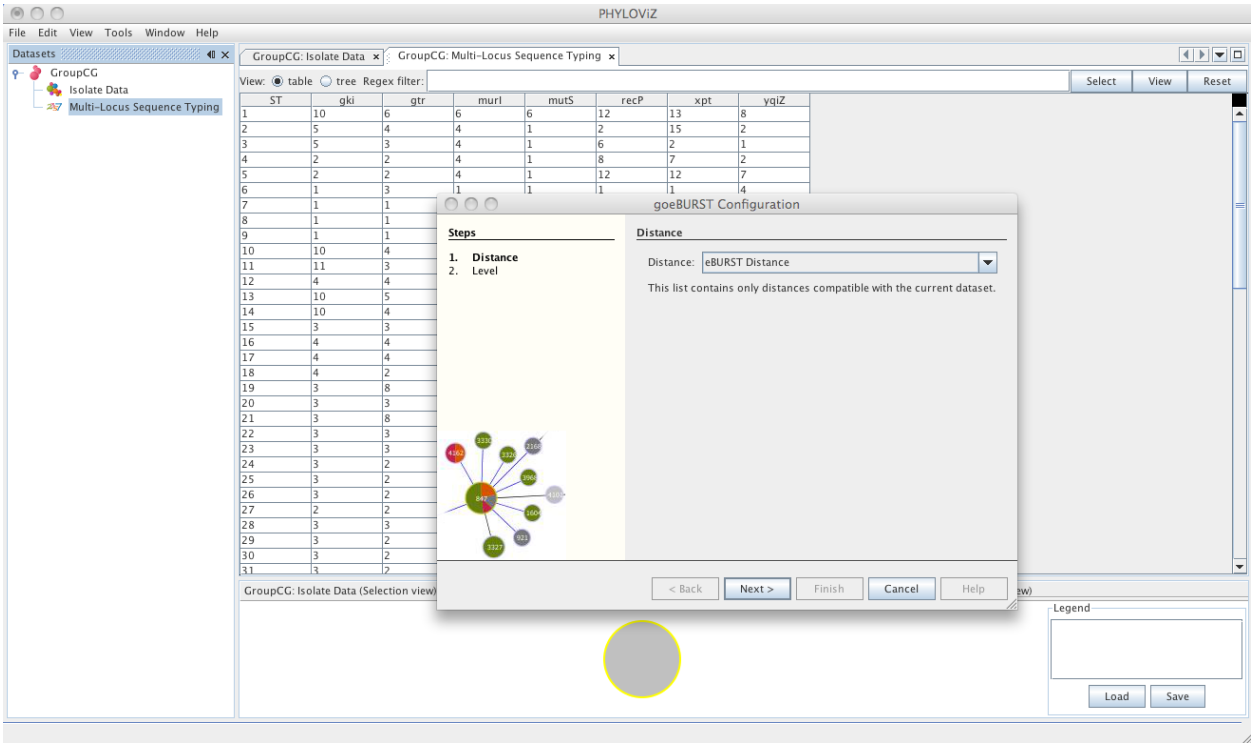
Data analysis

In the current version of PHYLOViZ, you can analyze your data using the several algorithms described below. Press the *Right Mouse Button* on the *Typing Data* (now named with the method) and choose compute to access the available analysis algorithms.

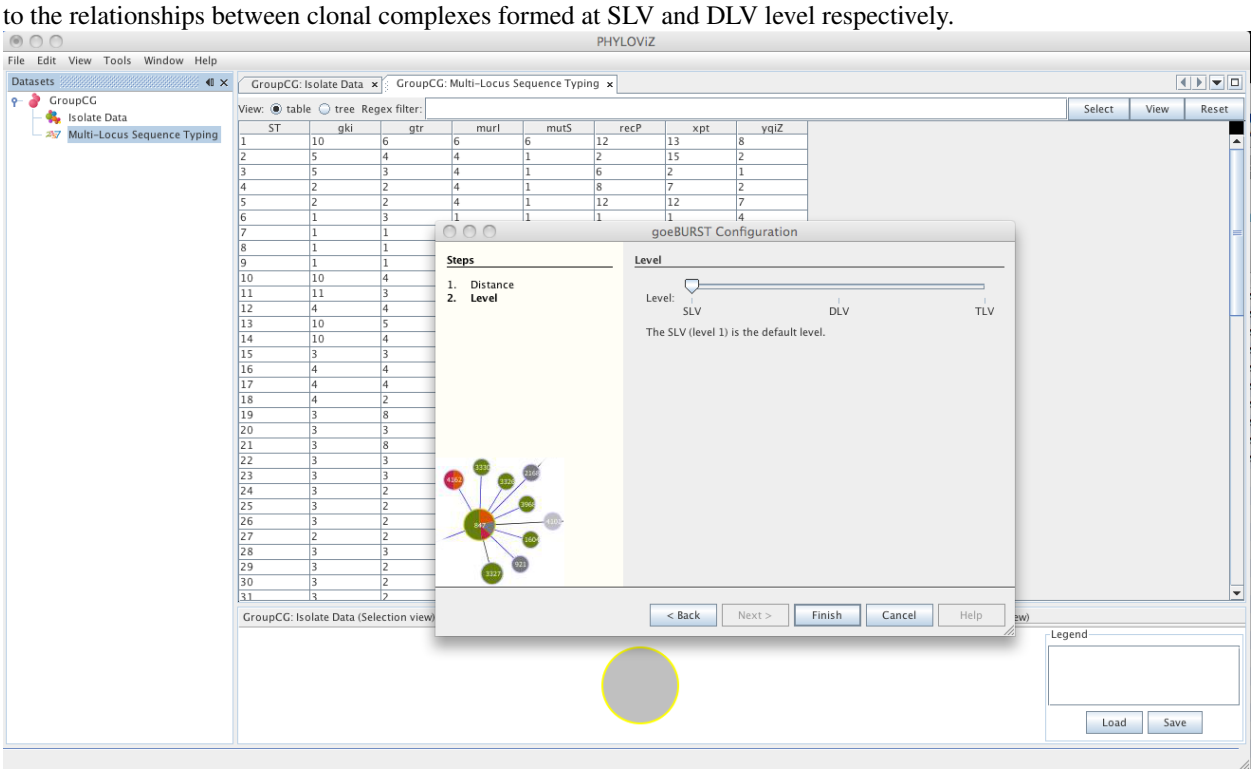


3.1 goeBURST algorithm

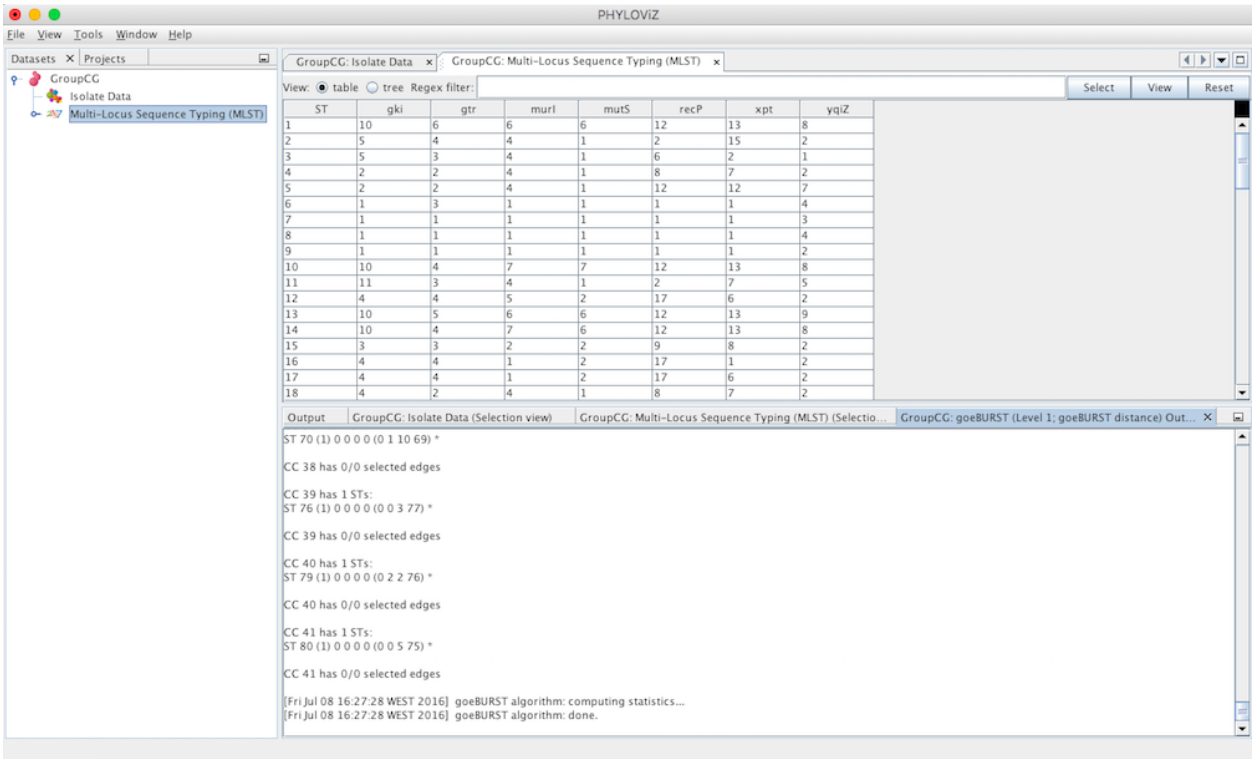
Selecting the goeBURST algorithms opens the dialog for the goeBURST algorithm. This algorithm was typically used for MLST data analysis and was originally described in the article Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. The first step is choosing the *Distance* to be used. Currently eBURST Distance is the only one available, but others could be implemented. The eBURST distances follows the tiebreak rules discussed in the article.



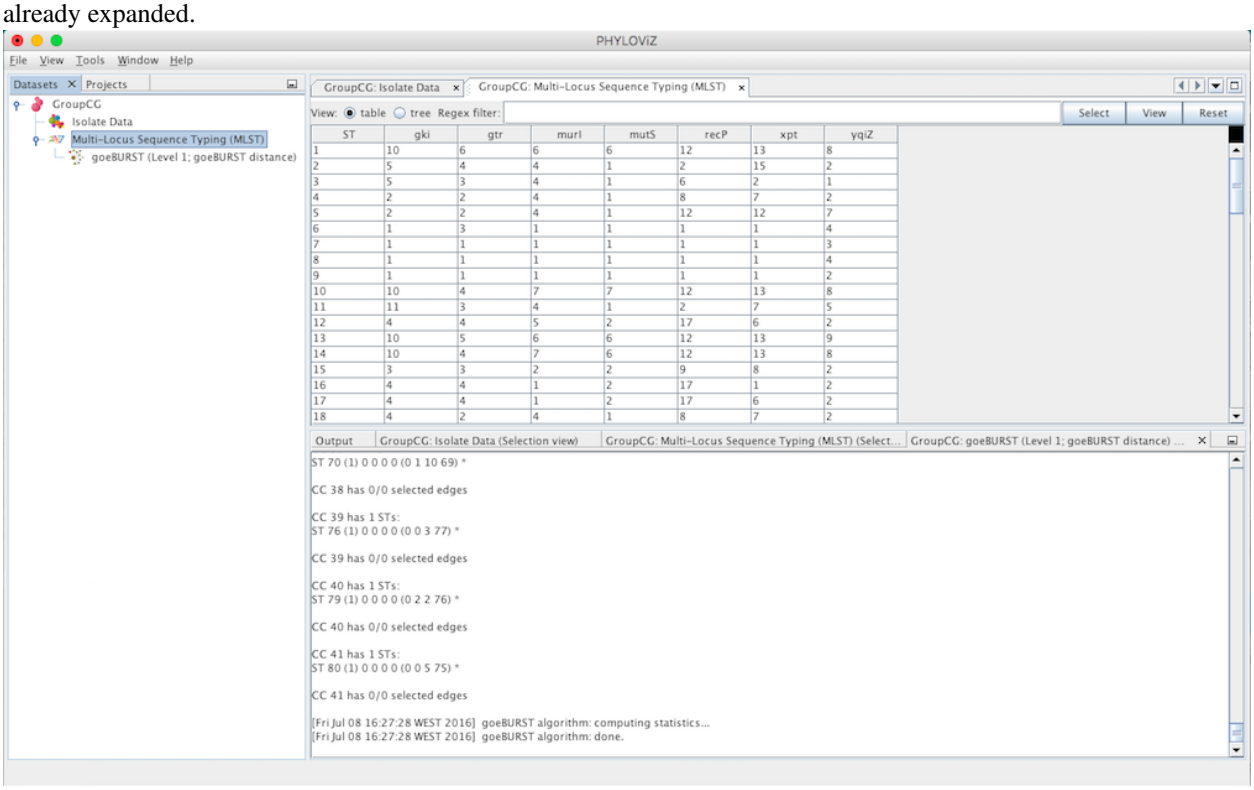
The second step is the choice of the level to which clonal complexes will be formed. The usual default for MLST analysis is SLV Level. Choosing DLV or TLV level will take longer calculation times, but could provide some insight to the relationships between clonal complexes formed at SLV and DLV level respectively.



A goeBURST *Output* tab will appear and display the goeBURST algorithm results. It will contain information about the Clonal Complexes (CCs), namely the Sequence Types that compose them and what edges (the links between STs) were drawn in each CC.

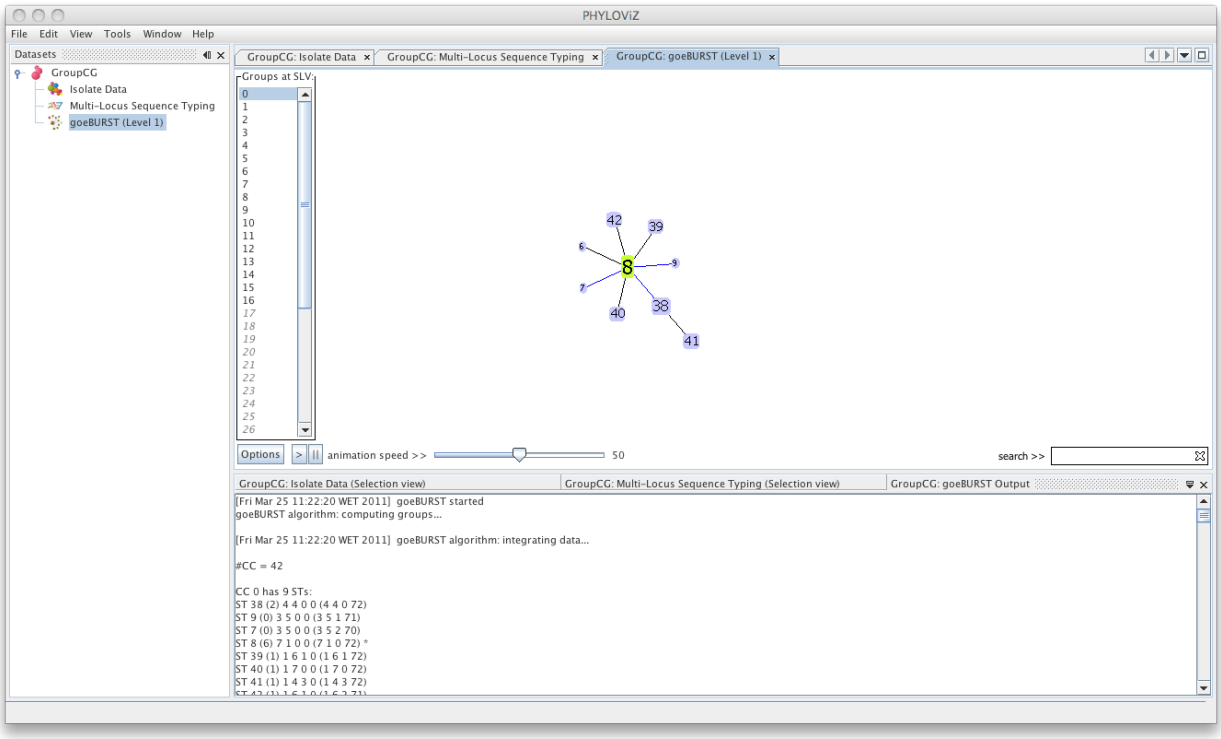


In order to display the goEburst tree view, it is necessary to expand the typing data on the DataSets' tab, if it is not already expanded.

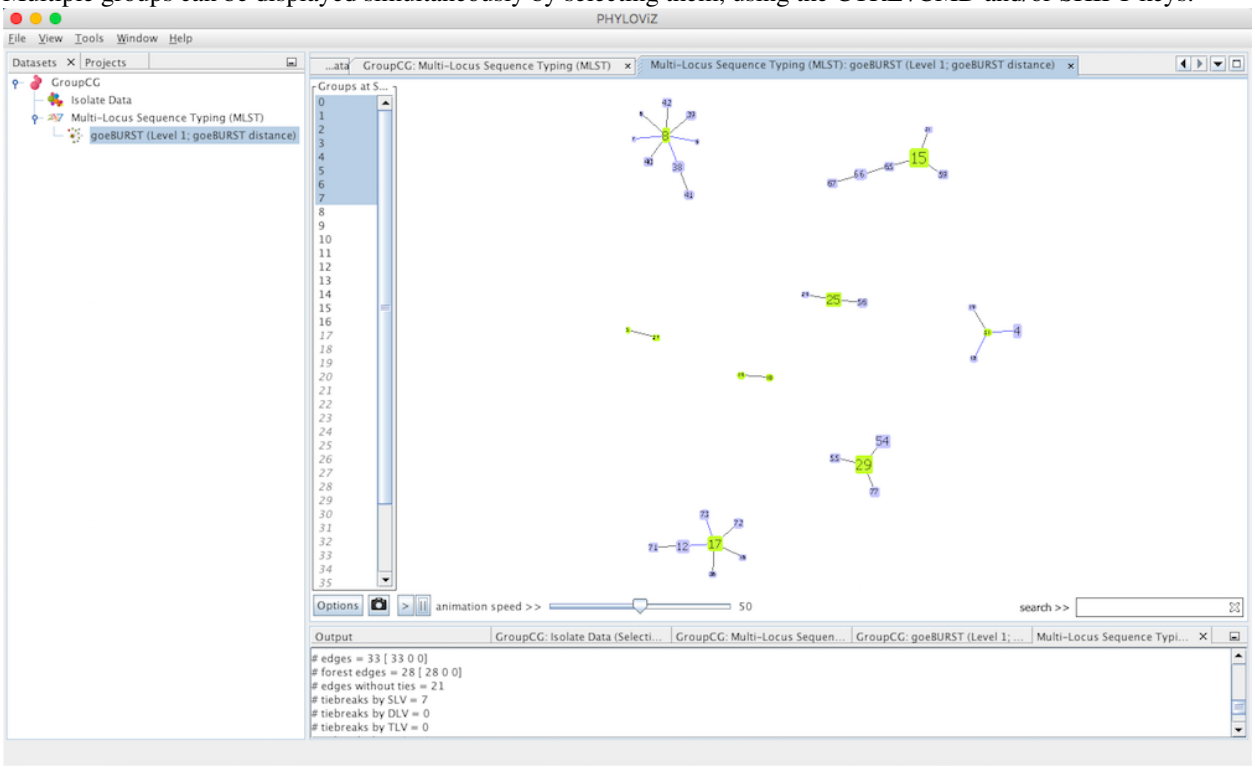


Double clicking on the goeBURST item that is now on the Dataset tree menu will show the display. The clonal complexes will be arbitrarily numbered starting from 0 (for the CC with most STs) and contains all the data relevant to the goeBURST analysis (STs in each group and the drawn SLVs edges). The following screenshot summarizes the

output for a single clonal complex with the test dataset used.



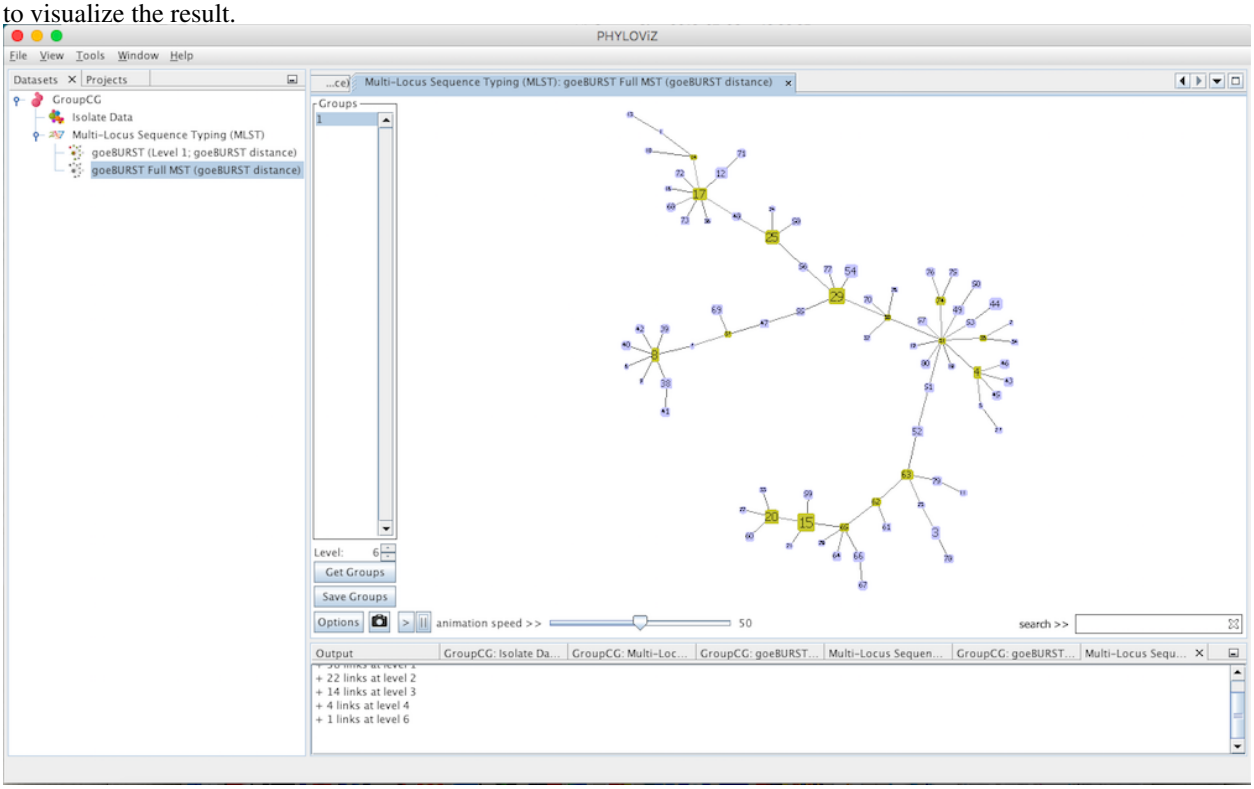
Multiple groups can be displayed simultaneously by selecting them, using the CTRL /CMD and/or SHIFT keys.



3.2 goeBURST Full MST algorithm

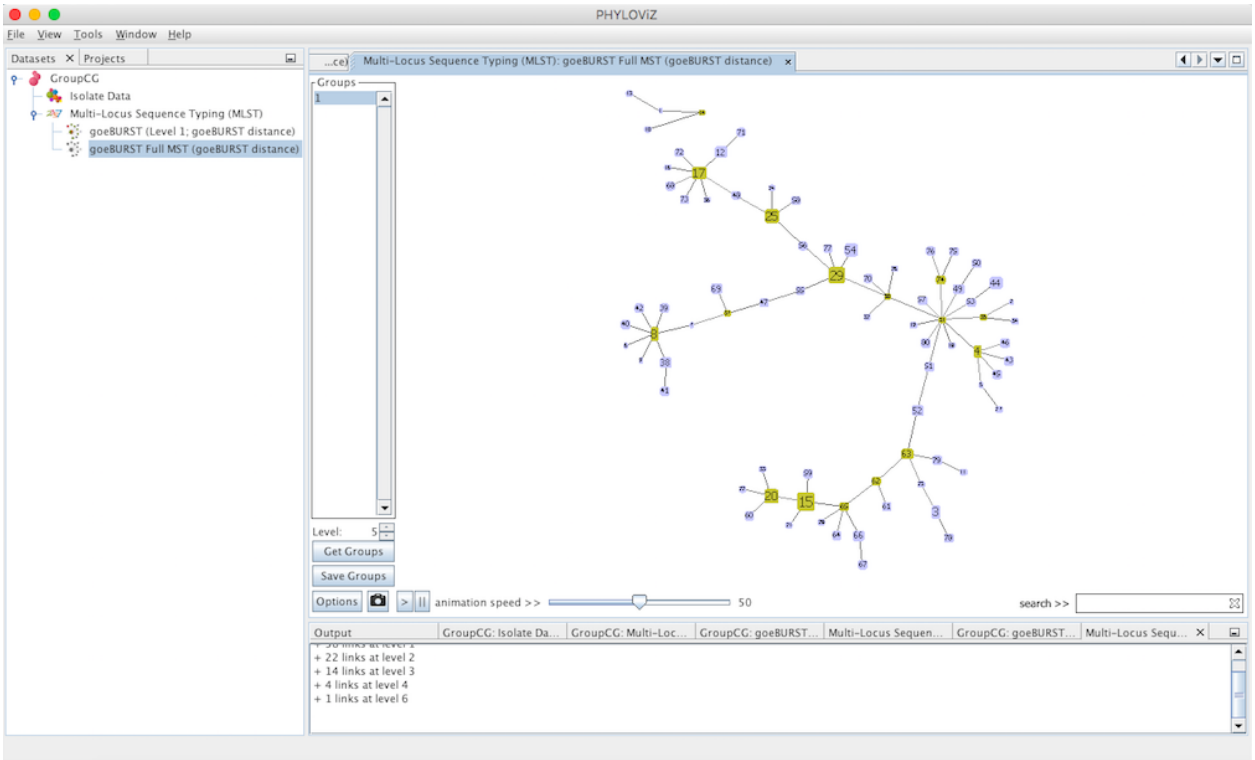
Using an extension of the goeBURST rules up to \(n\)LV level (where \(n\) equals to the number of loci your dataset uses), a Minimum Spanning Tree-like structure can be drawn. This is typicially used for SNP or cg/wgMLST datasets with dozens to thousand of loci.

Select *goeBURST Full MST* in the *Compute* options to draw it. Contrary to the standard goeBURST, the link statistics are not presented. After computation, double click on the *goeBURST Full MST* that appears under the dataset heading to visualize the result.

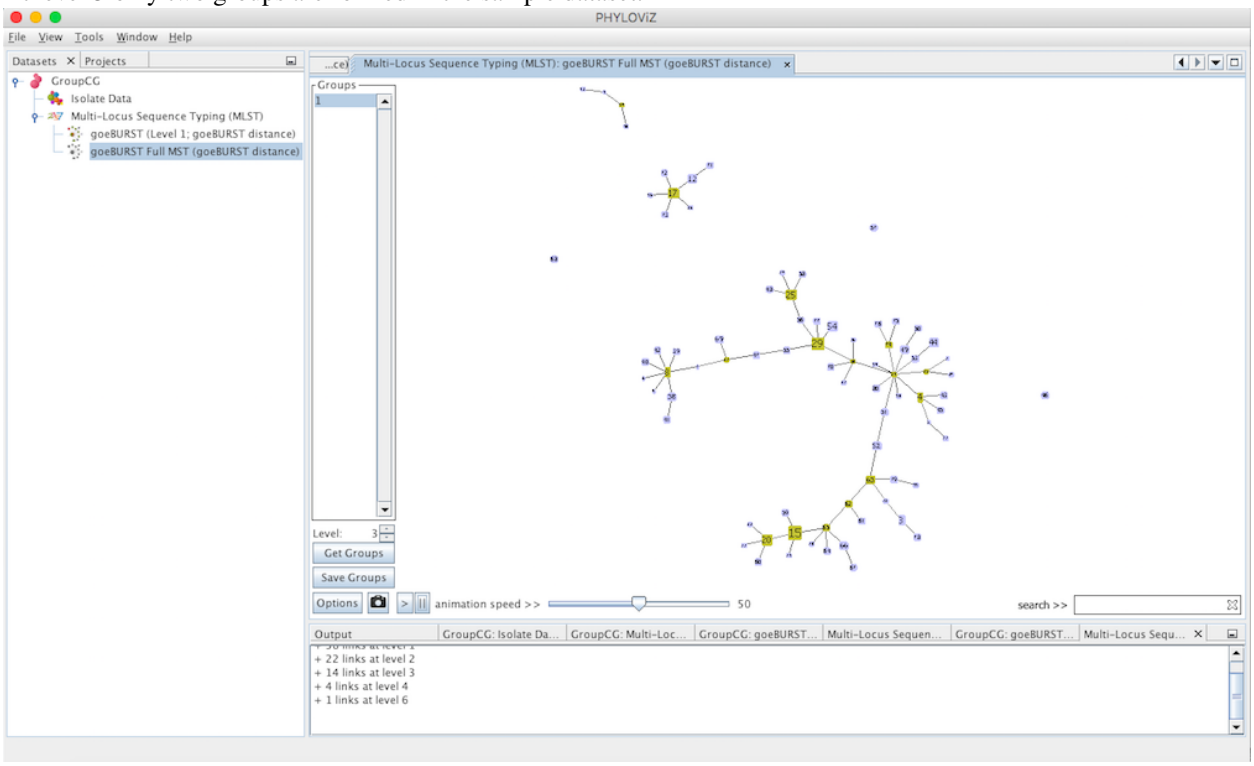


New options appear on the display: The *Level* selector and two new buttons *Get Groups* and *Save Groups*. The *Level* represents the *Locus Variant* level and allows the removal of all the links greater than the number represented. The user can use the up and down arrows or directly edit the number by clicking on it. The *Get Groups* button allows separate the display of groups that are not connected at the level chosen in order to simplify the analysis of larger datasets. This will generate a display very similar to that of goeBURST, but at a higher link level. The Save Groups creates an extra column in the isolate data with the title label *goeBURST MST[\(x\)]* with \(x\) being equal to the level used to create the groups.

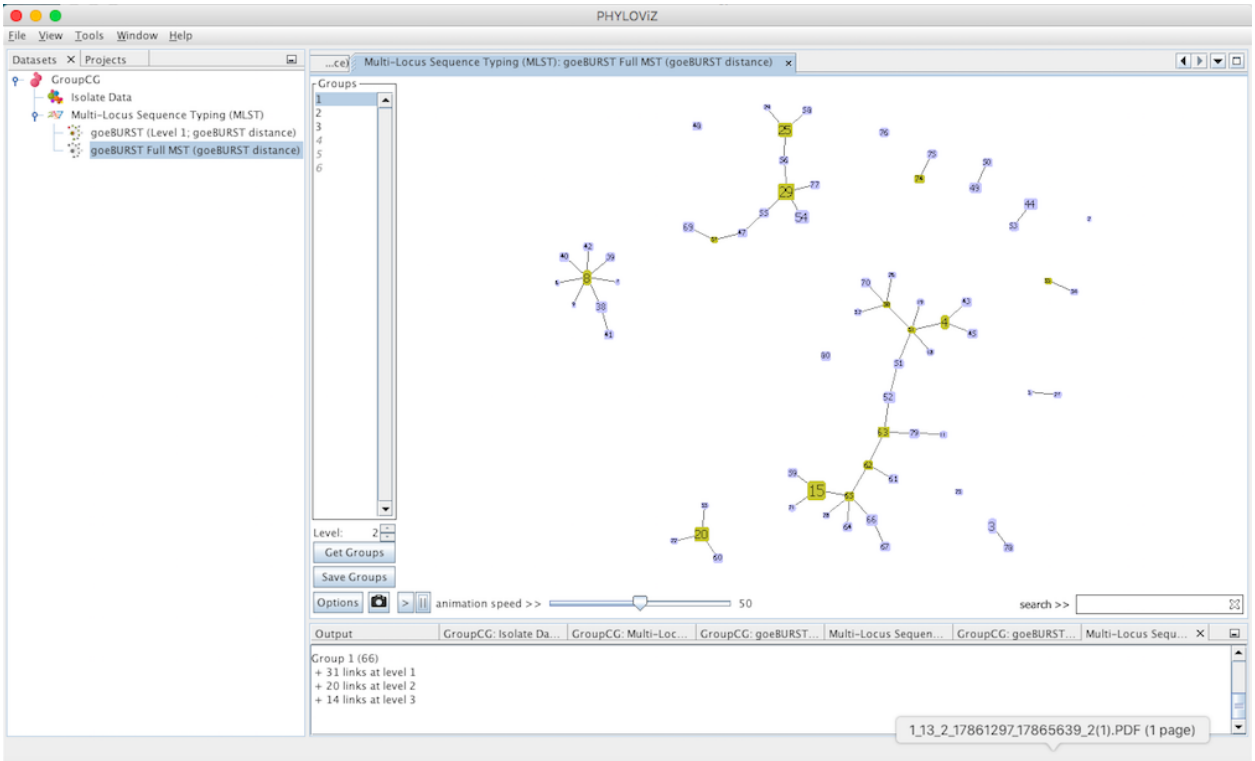
Decreasing the *Level* selector, allows the user to see how clonal complexes would relate to each other at a certain level. Level 1, 2 and 3 are equivalent to calculating goeBURST at those levels (SLV,DLV and TLV respectively). The following images shows what happens to the dataset when you decrease the level. Level 4 is not displayed since no new groups are formed at that level.



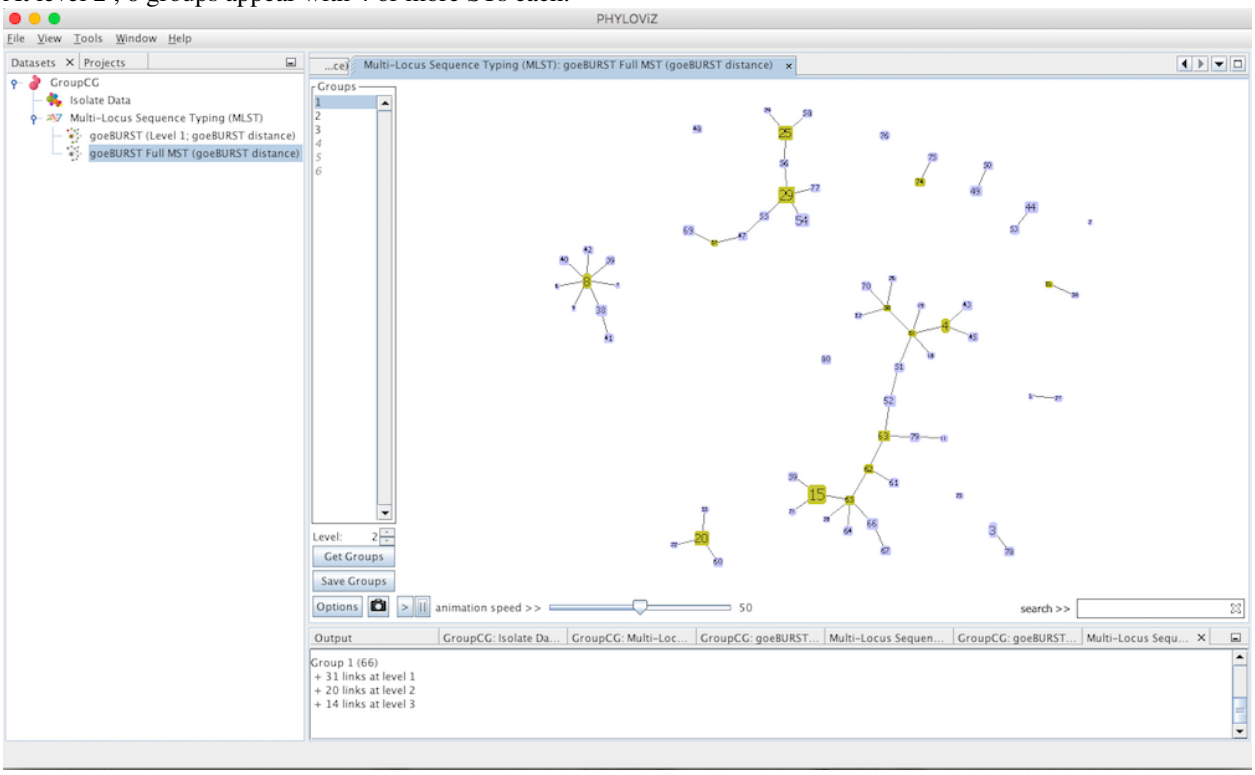
At level 5 only two groups are formed in the sample dataset.



At level 3 (TLV level) some singletons appear. Level 4 is not shown since no changes were observed in the graph. This means that there are no two STs in the dataset that differ in 4 of the loci of their profiles.



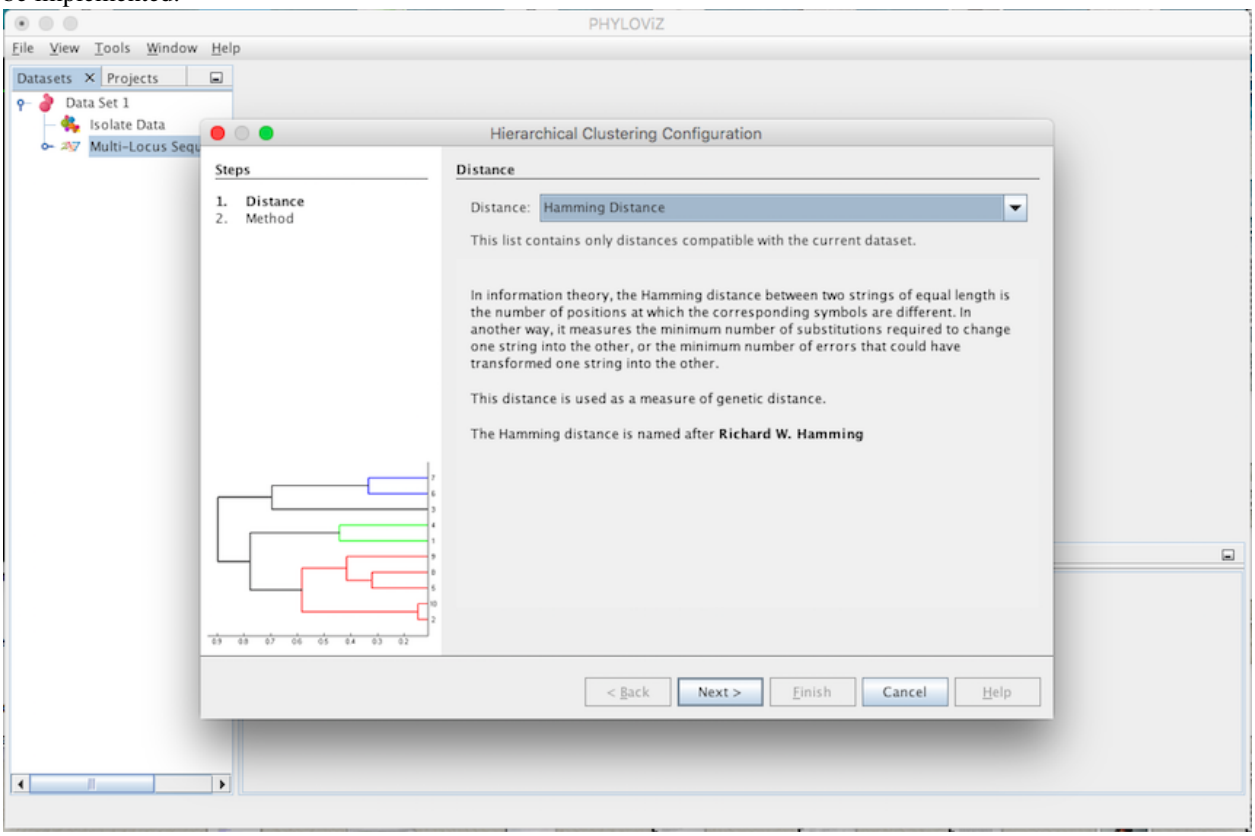
At level 2 , 6 groups appear with 4 or more STs each.



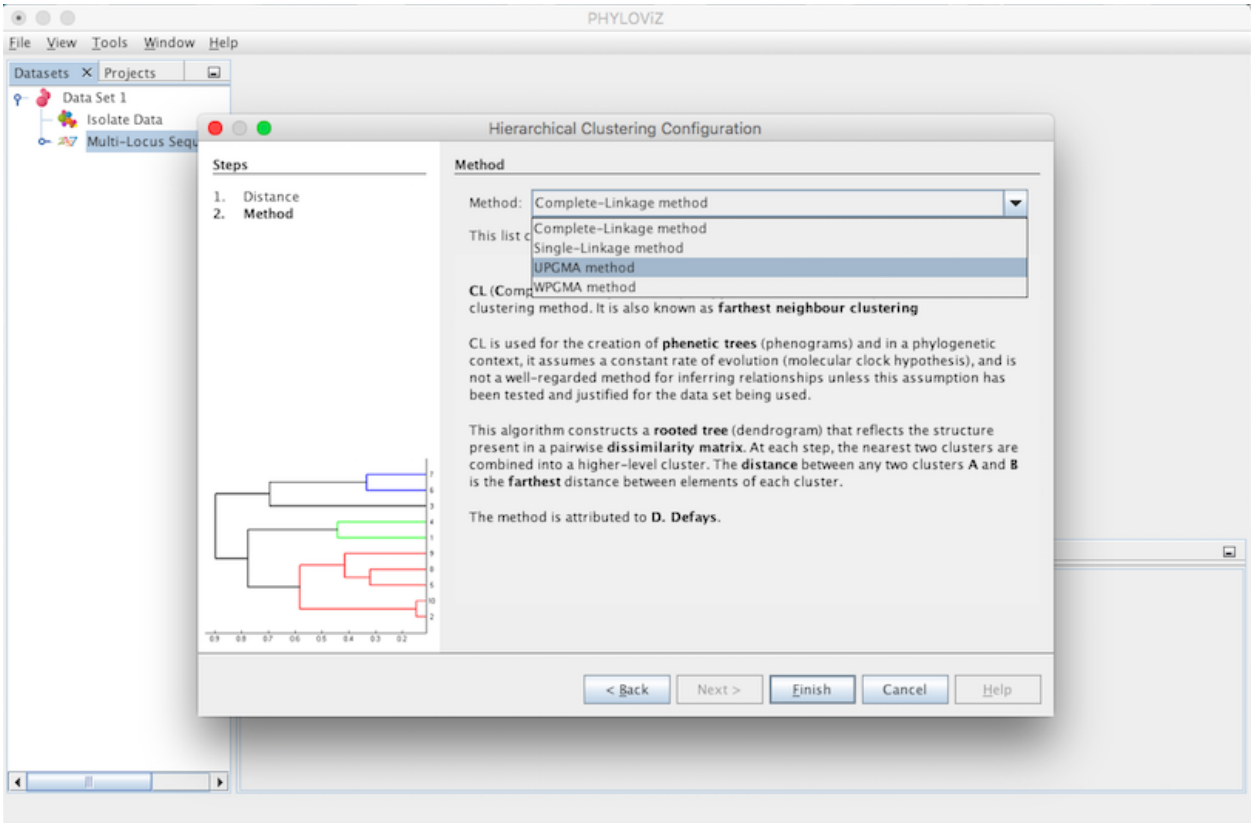
And finally at level 1, the equivalent of the most commonly used Clonal Complex definition by goeBURST, 17 groups with 2 or more STs are formed and there are 25 singletons on the dataset.

3.3 Hierarchical Clustering

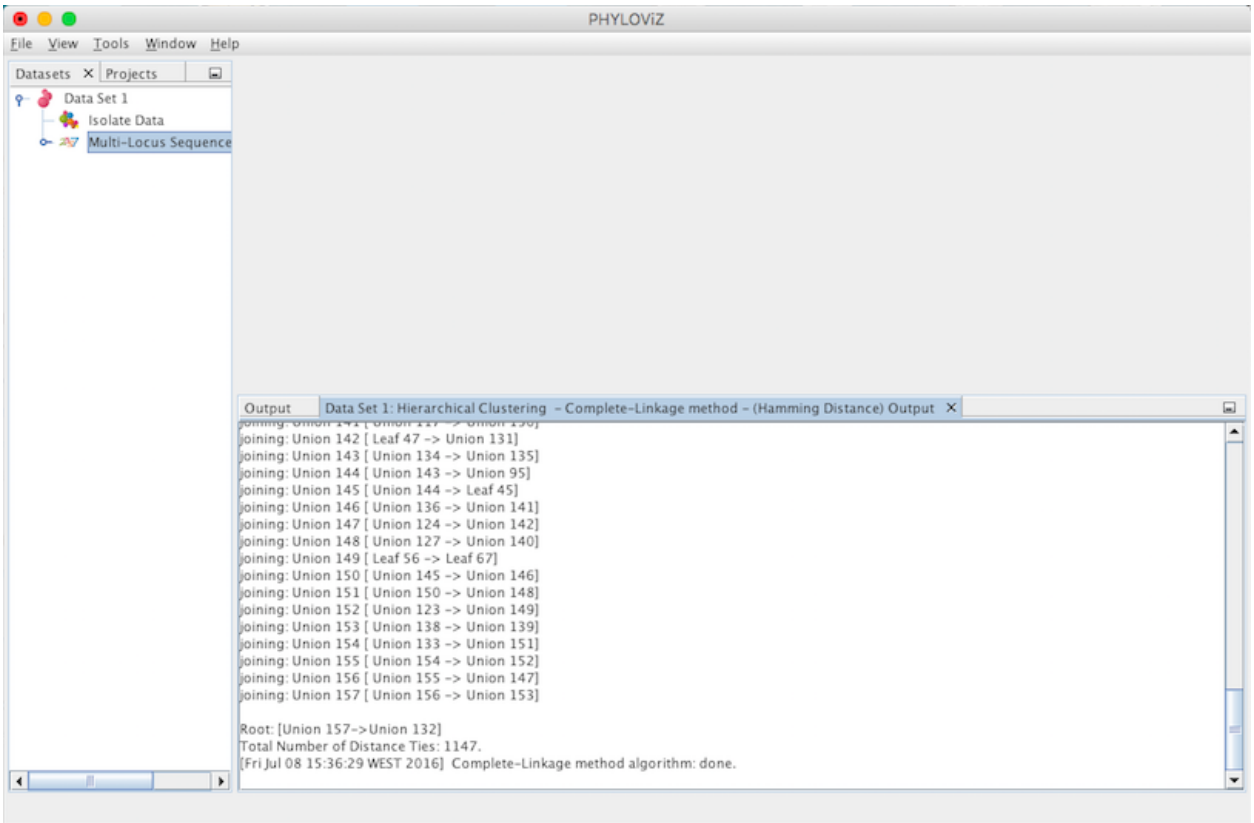
Selecting the Hierarchical Clustering opens the dialog where you can select what method you want to apply. The first step is choosing the *Distance* to be used. Currently the hamming distance is the only one available, but others could be implemented.



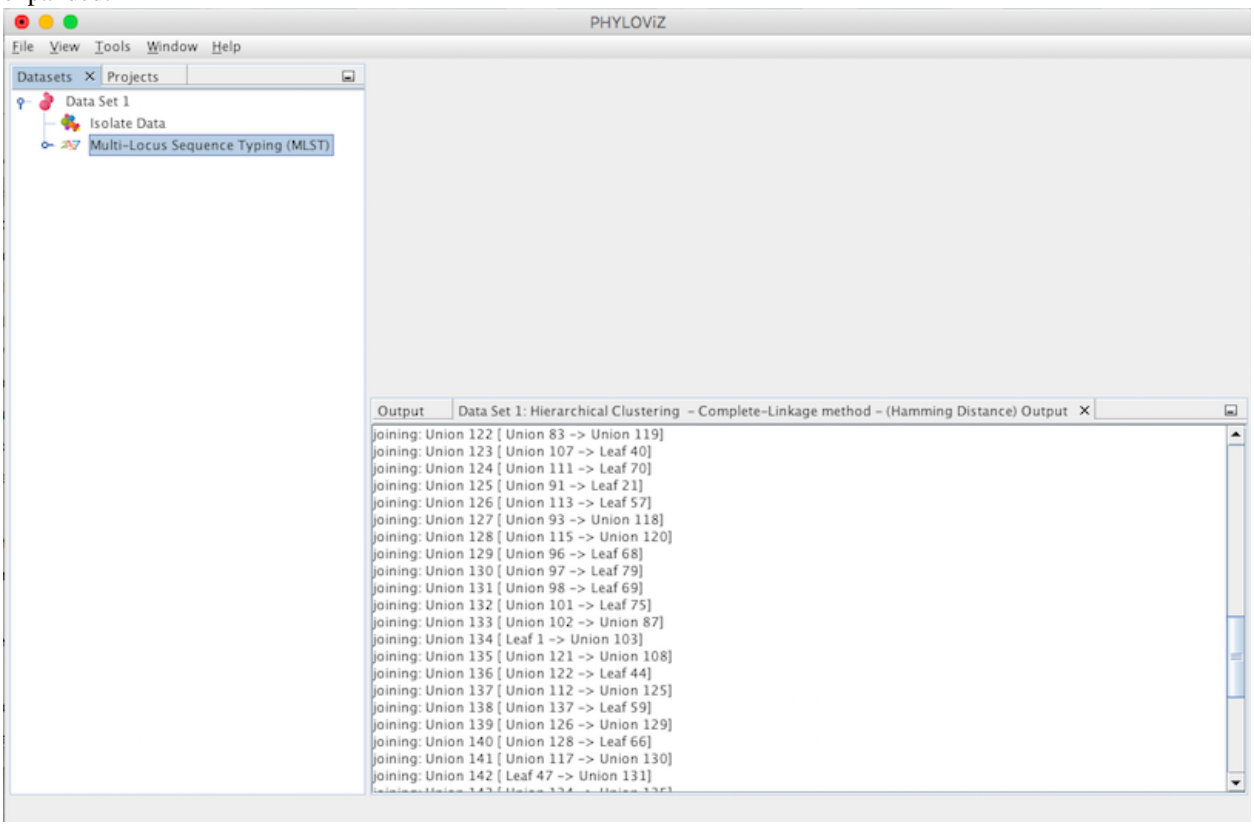
The second step is to select the *Method*. You can choose between complete-linkage, single-linkage, UPGMA (Unweighted Pair Group Method with Arithmetic mean) and WPGMA (Weighted Pair Group Method with Arithmetic mean). Selecting the method corresponds to selecting the criterion of minimal dissimilarity.



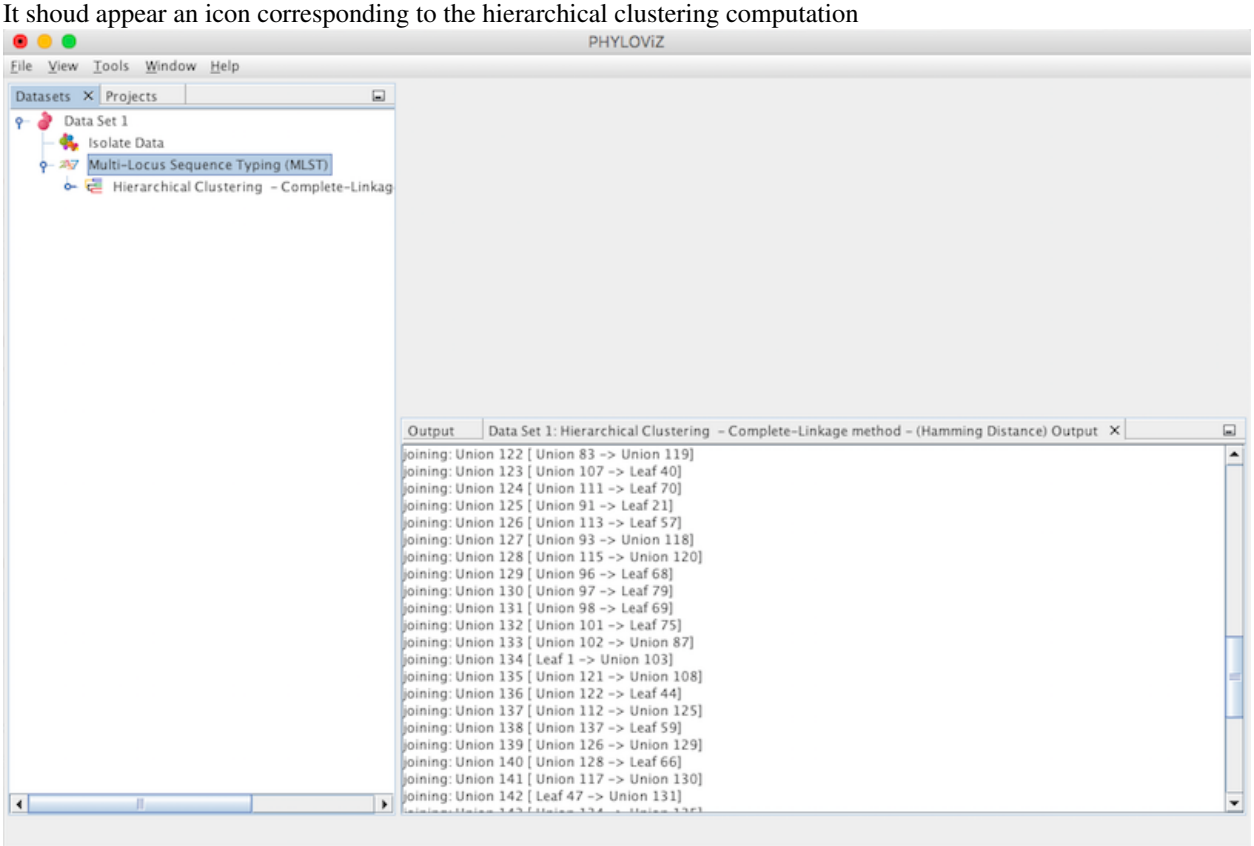
A Hierarchical Clustering *Output Tab* will appear and display the results of the application of the chosen method. A *Leaf* represents a Sequence Type and a *Union* represents a group that results of joining Leafs or Unions with Leafs. This process of joining is displayed step by step by the algorithm in the *Output's Tab*. Finally we have the number of ties occured. The tie break applied is to always choose the first one found.



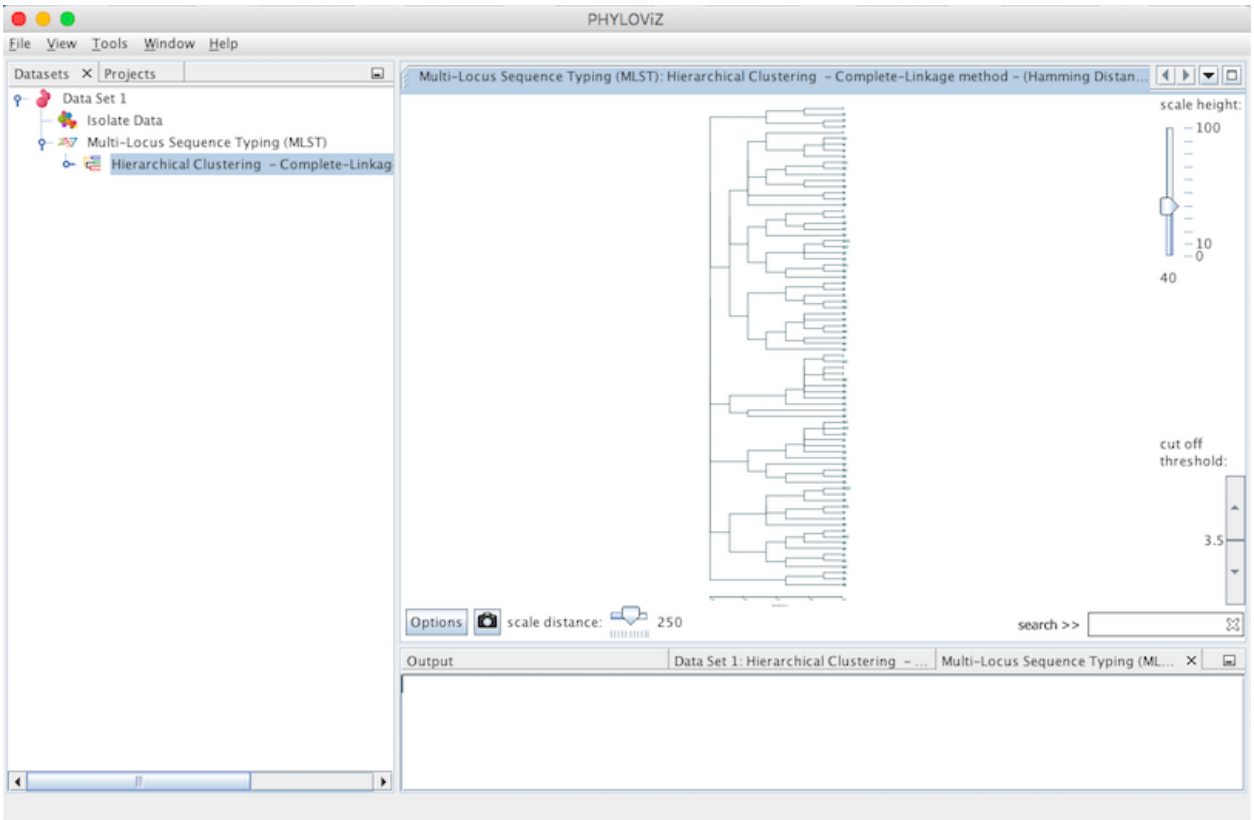
In order to display the dendogram view, it is necessary to expand the typing data on the Datasets' tab, if it is not already expanded.



It shoud appear an icon corresponding to the hierarchical clustering computation



Double clicking on the Hierarchical Clustering item will show the display. This type of clustering is represented in the format of a dendogram. The following screenshot summarizes the output for the previous dataset. Sometimes it is necessary to fit the image to see all the display at once. To do this, please right click on the mouse over the display.



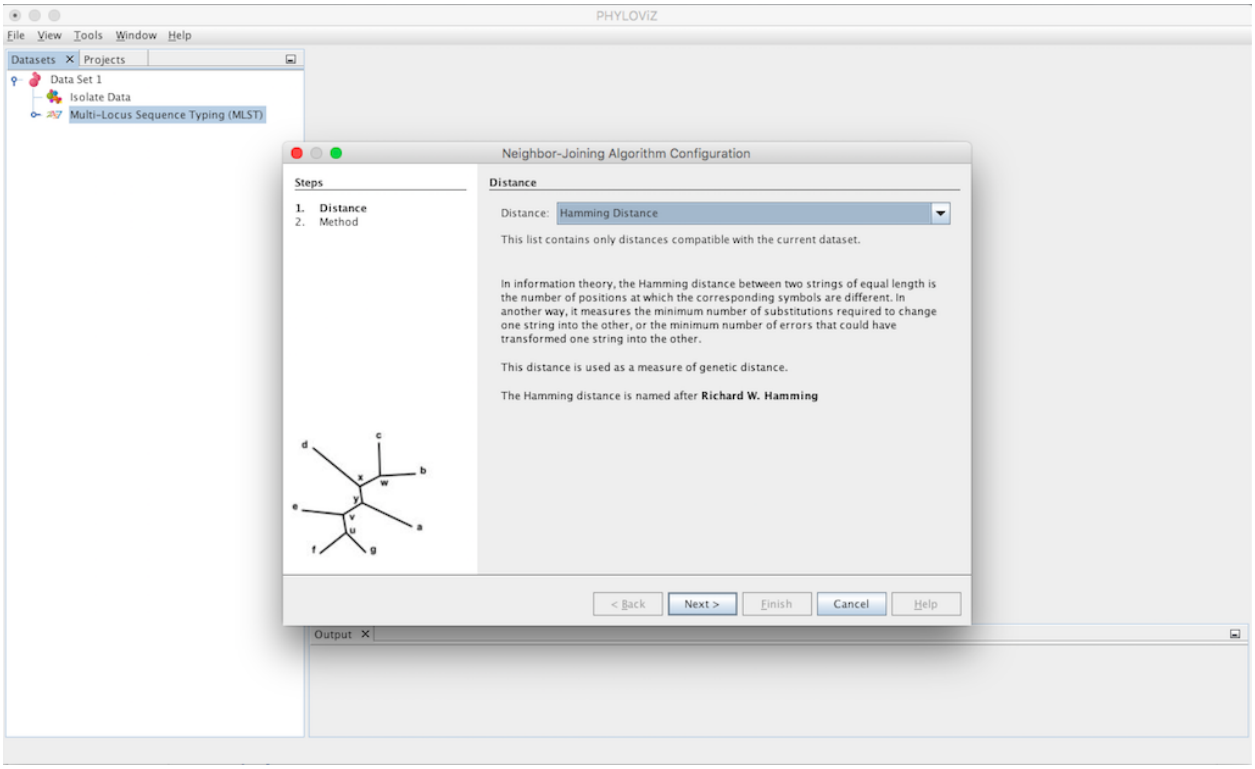
Some features were added to the visualization to improve and facilitate the analysis. These features are the following:

1. Height scale
2. Width scale
3. Options Panel
4. Search ST
5. Filter by distance (cut off threshold)
6. Export image

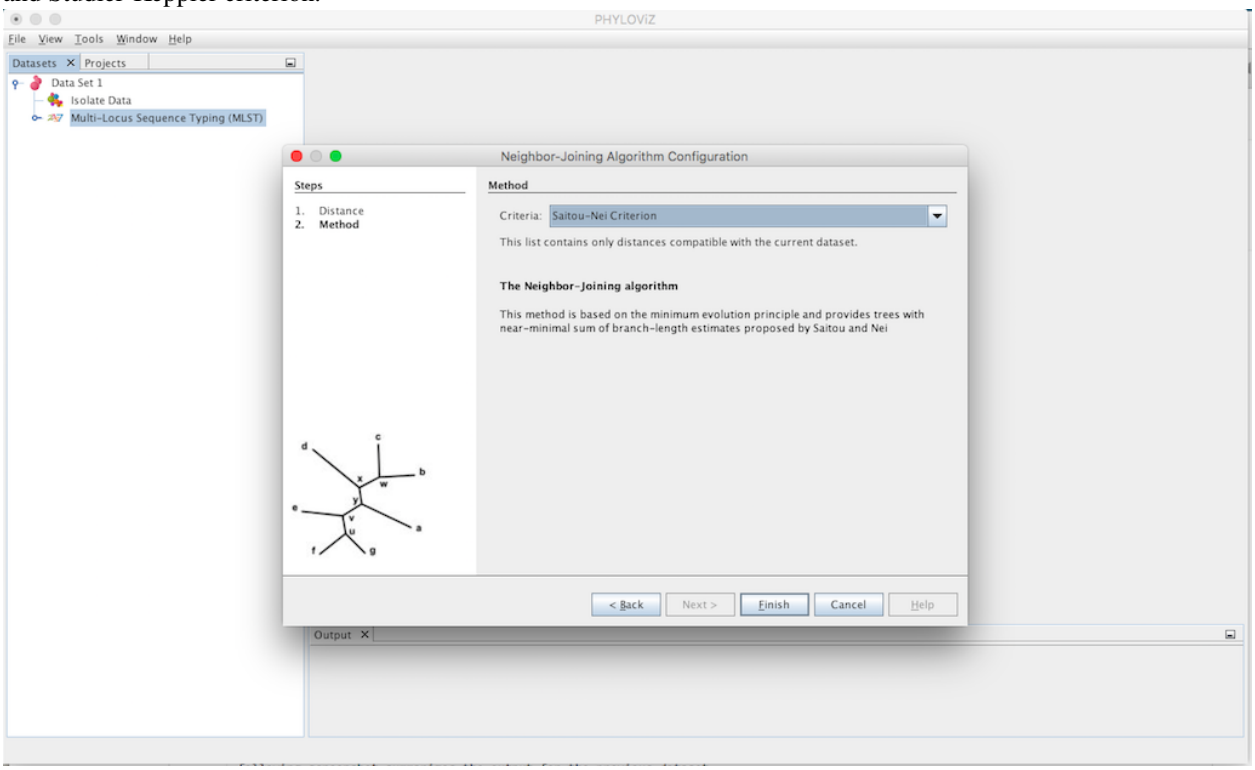
See section display and visualization for more information on these features.

3.4 Neighbor Joinning

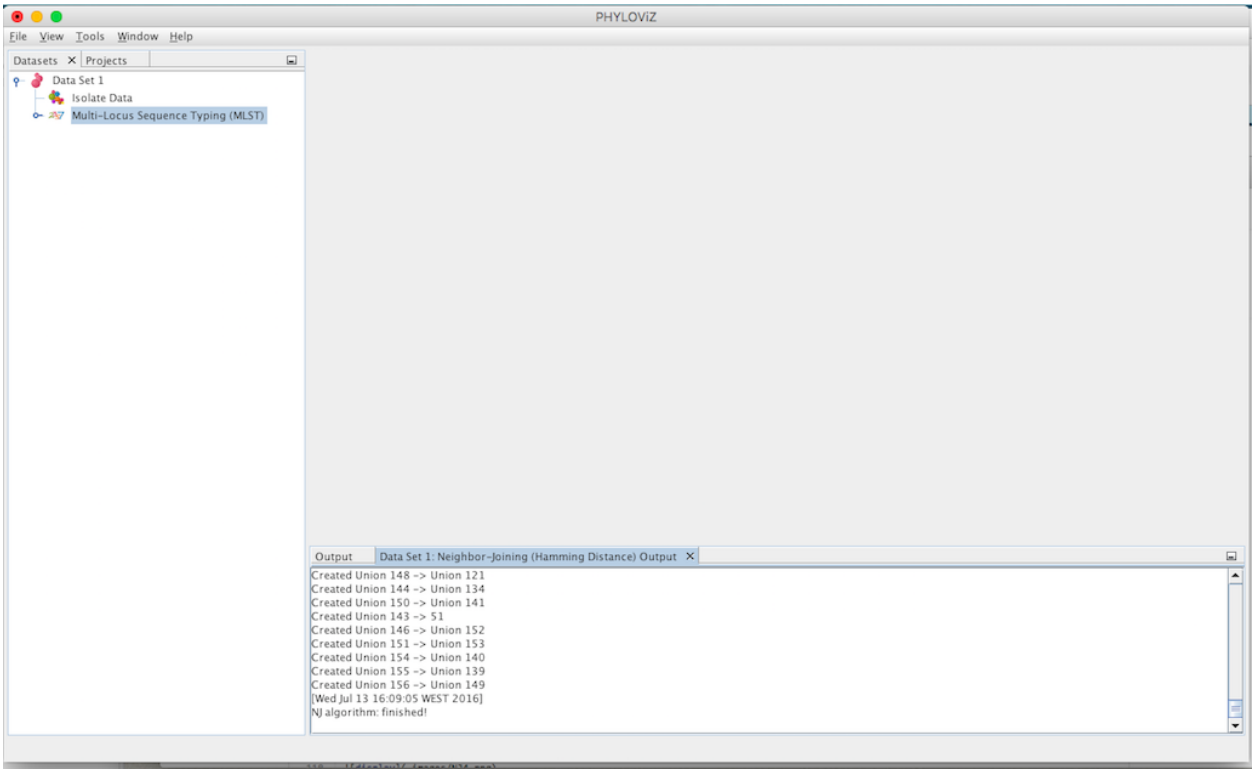
Selecting the Neighbor Joinning algorithm opens the dialog where you can select what method you want to apply. The first step is choosing the *Distance* to be used.



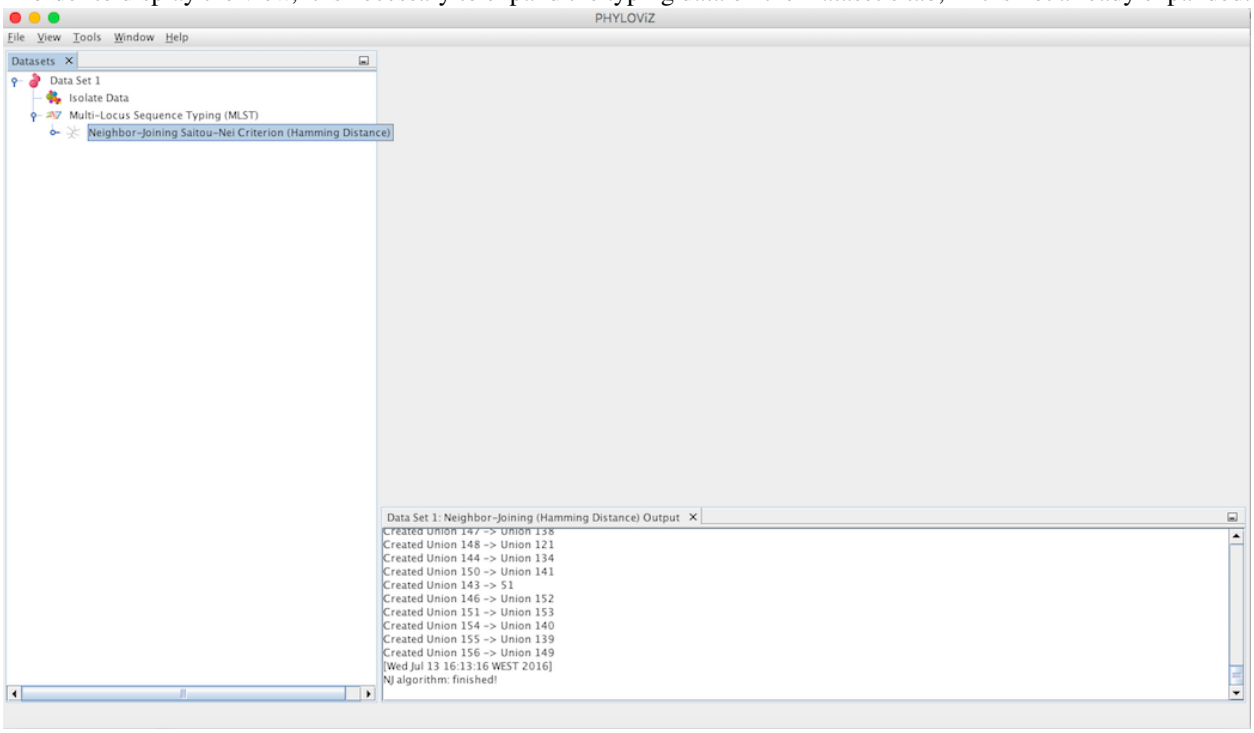
The second step is to select the *Criteria* of the tree branch-length minimization. You can choose between Saitou-Nei and Studier-Keppler criterion.



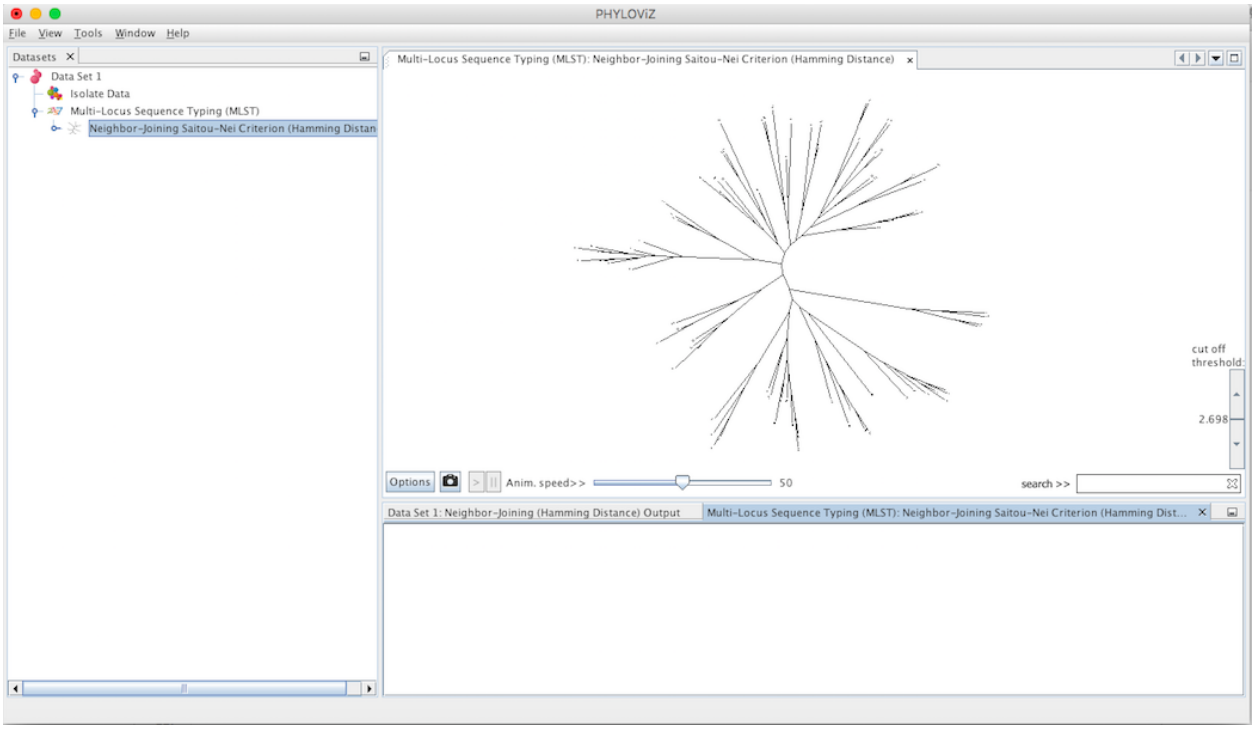
A Neighbor Joinning *Output Tab* will appear and display the results of the application of the chosen method. The information displayed represents the same as the Hierarchical Clustering *Output Tab*.



In order to display the view, it is necessary to expand the typing data on the Dataset's tab, if it is not already expanded.



Double clicking on the Neighbor Joinning item will show the display. By default it is represented in the format of a radial tree. The following screenshot summarizes the output for the previous dataset.



Some features were added to the visualization to improve and facilitate the analysis. These features are the following:

1. Options Panel that includes changing the tree layout
2. Search ST
3. Filter by distance
4. Export image

CHAPTER 4

Display and visualization

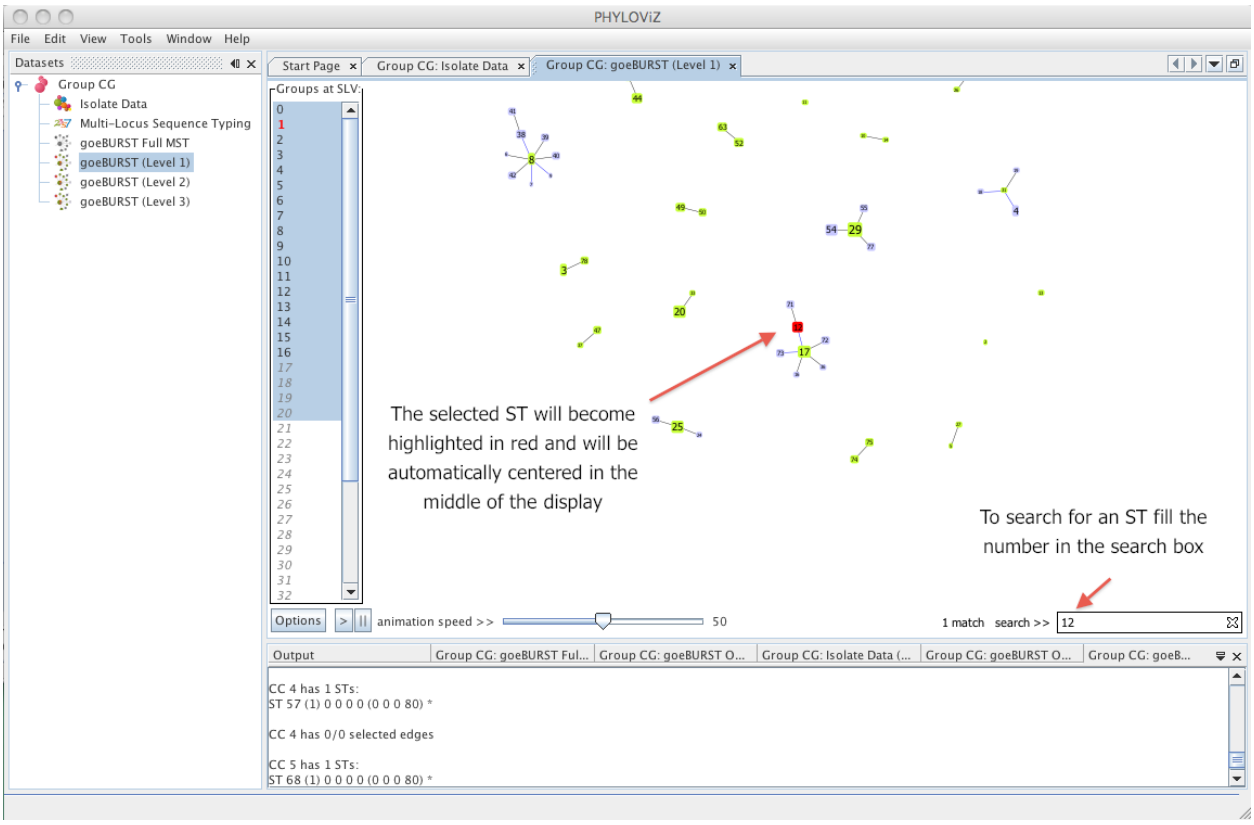
4.1 Interface features

After running the selected algorithm, you will notice that the program then tries to optimize the display of the group with the largest number of elements in the data set. You can change the speed at which this occurs by moving the animation speed slider.

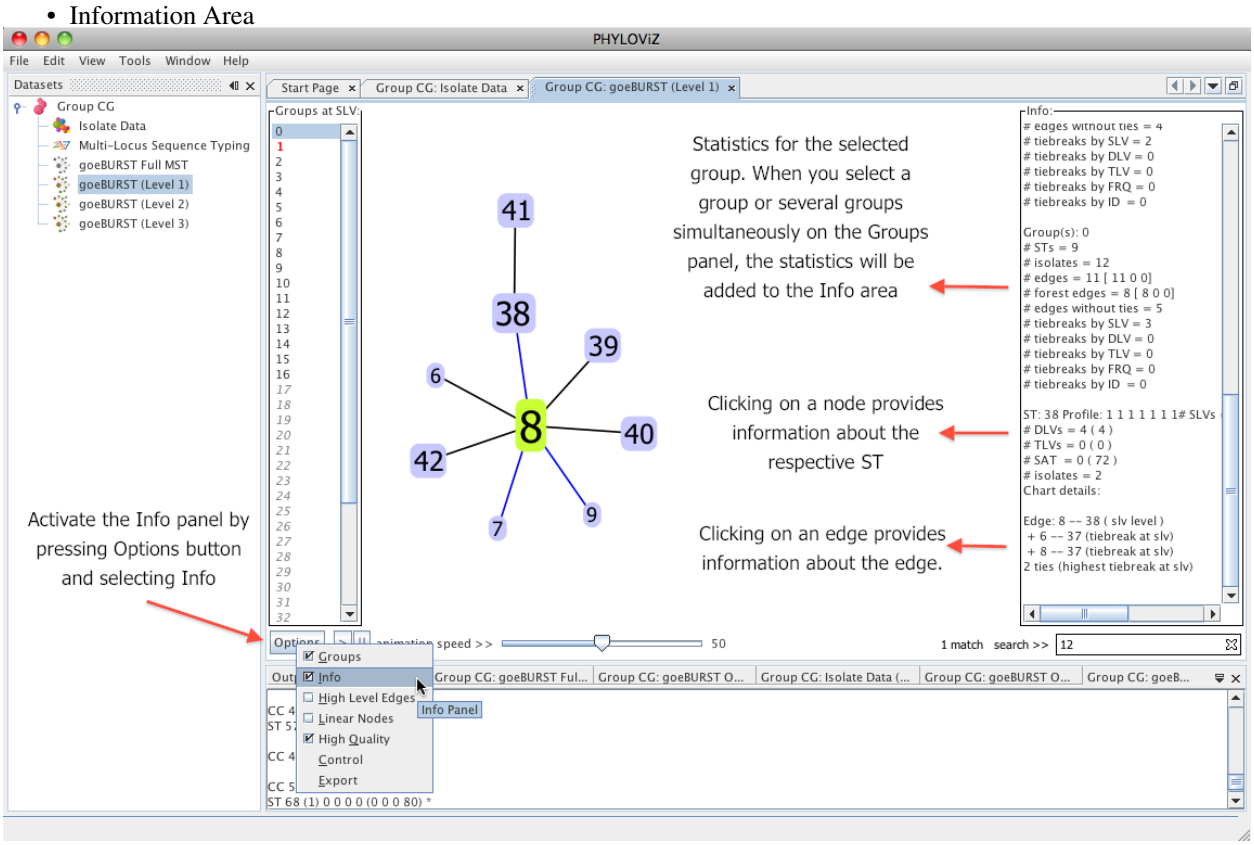
The Display tab offers the user the ability to search for an isolate, Highlight the SLVs and DLVs, control the animation speed, select diferent diferent or multiple groups. You can fit any displayed graphs to the window by right-clicking any open space (i.e. with no link or ST node) on the window.

4.1.1 Common features

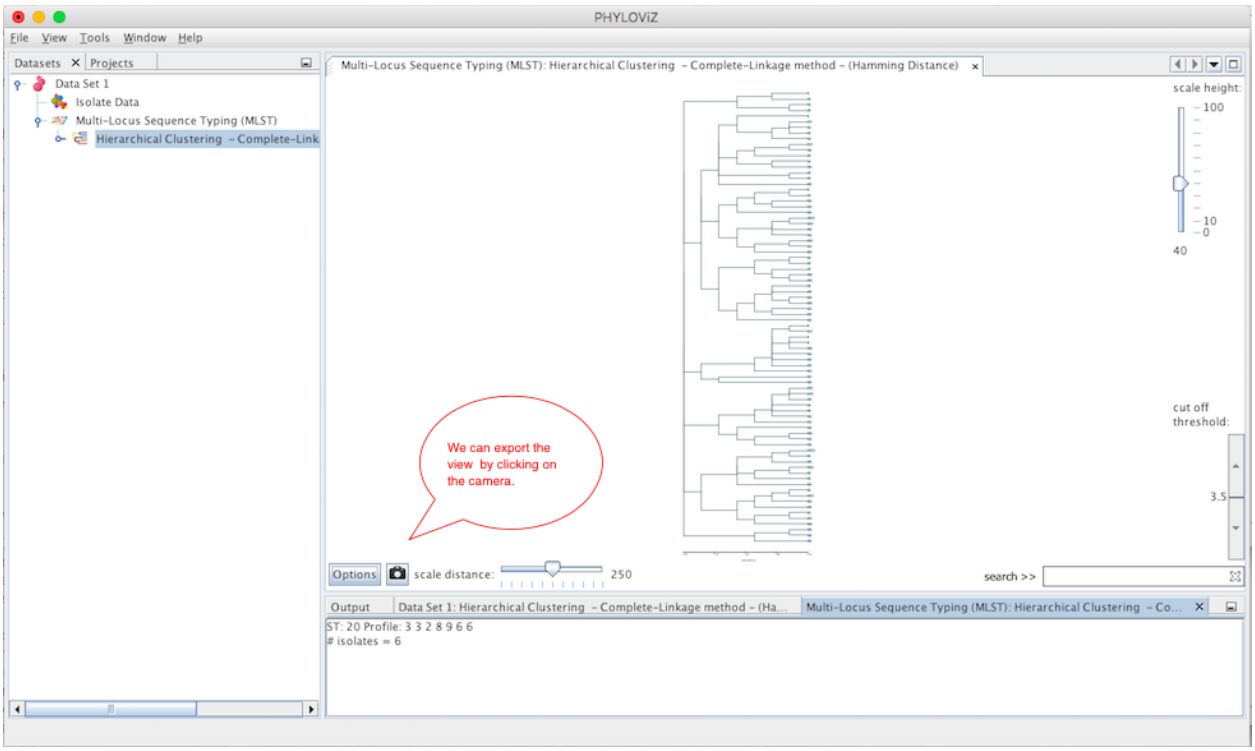
- ST search



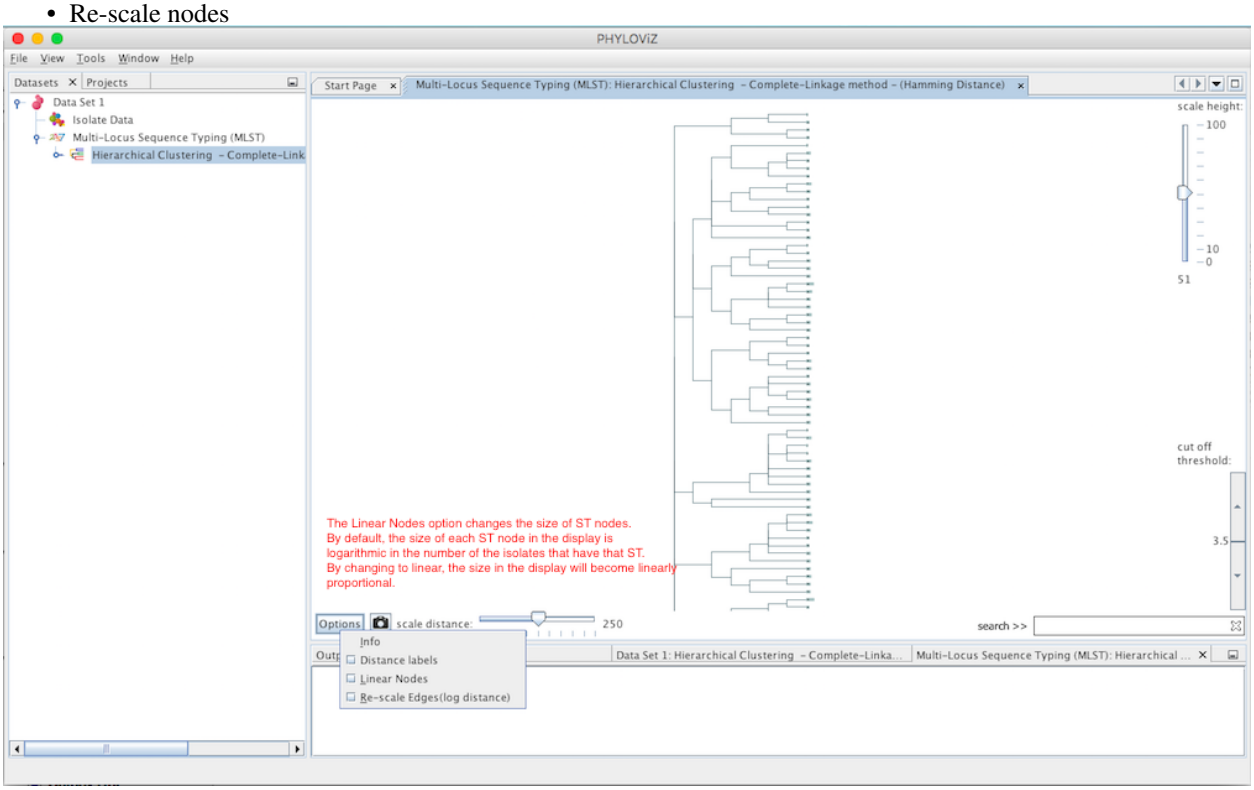
- Information Area



- Exporting Image

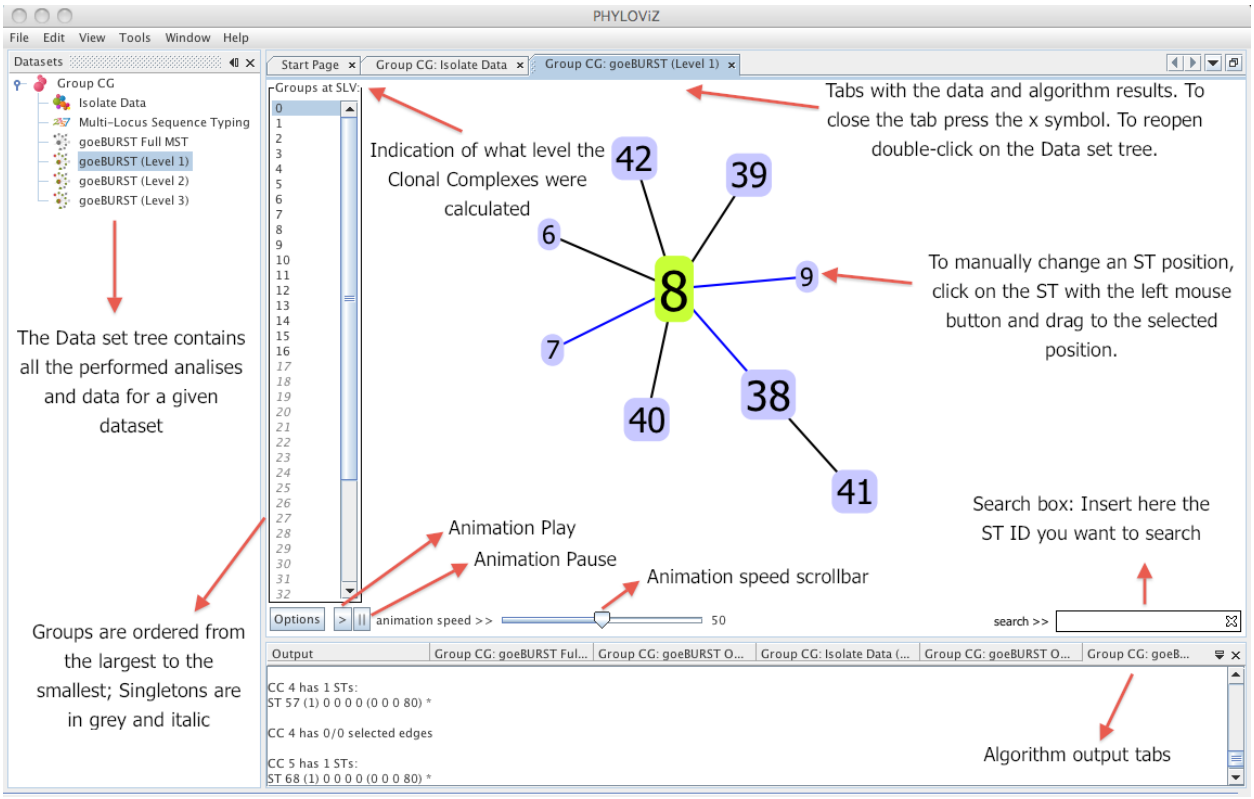


- Re-scale nodes

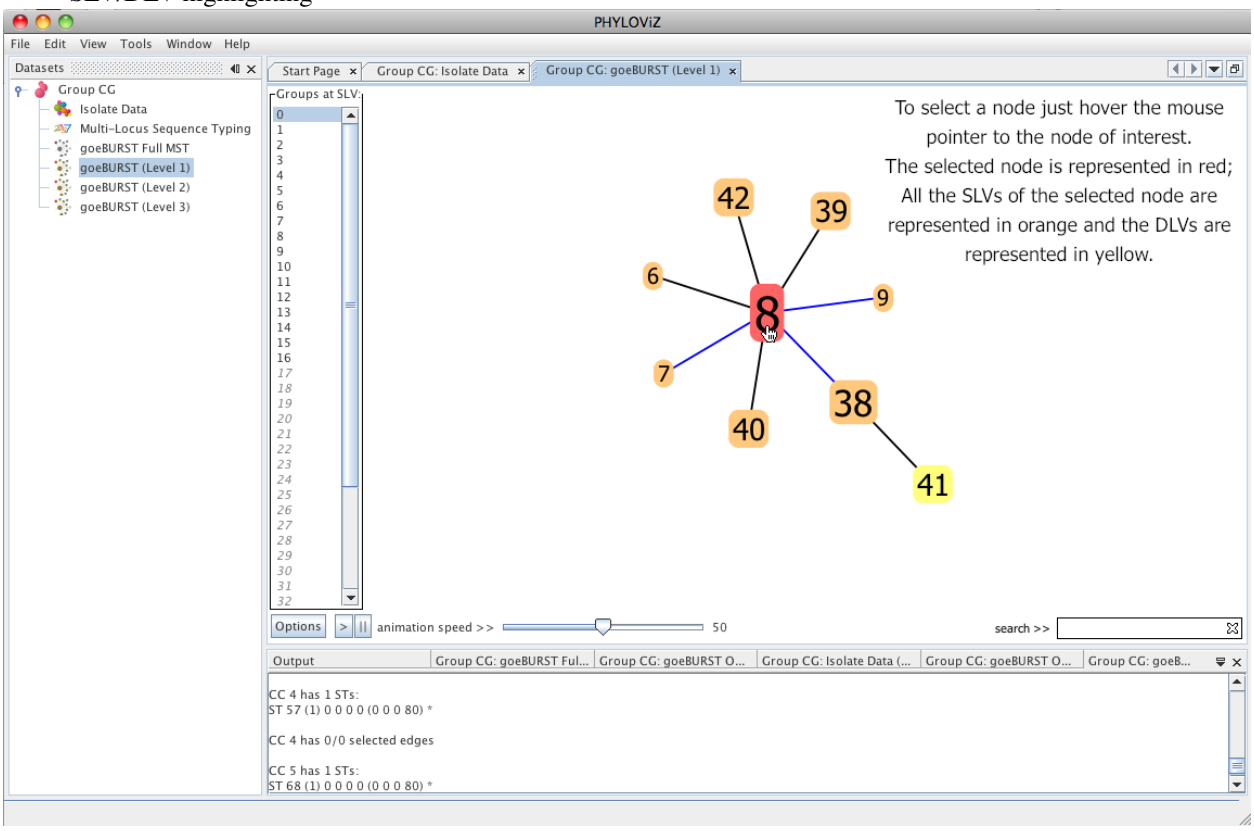


4.1.2 GoeBURST and GoeBURST Full MST features

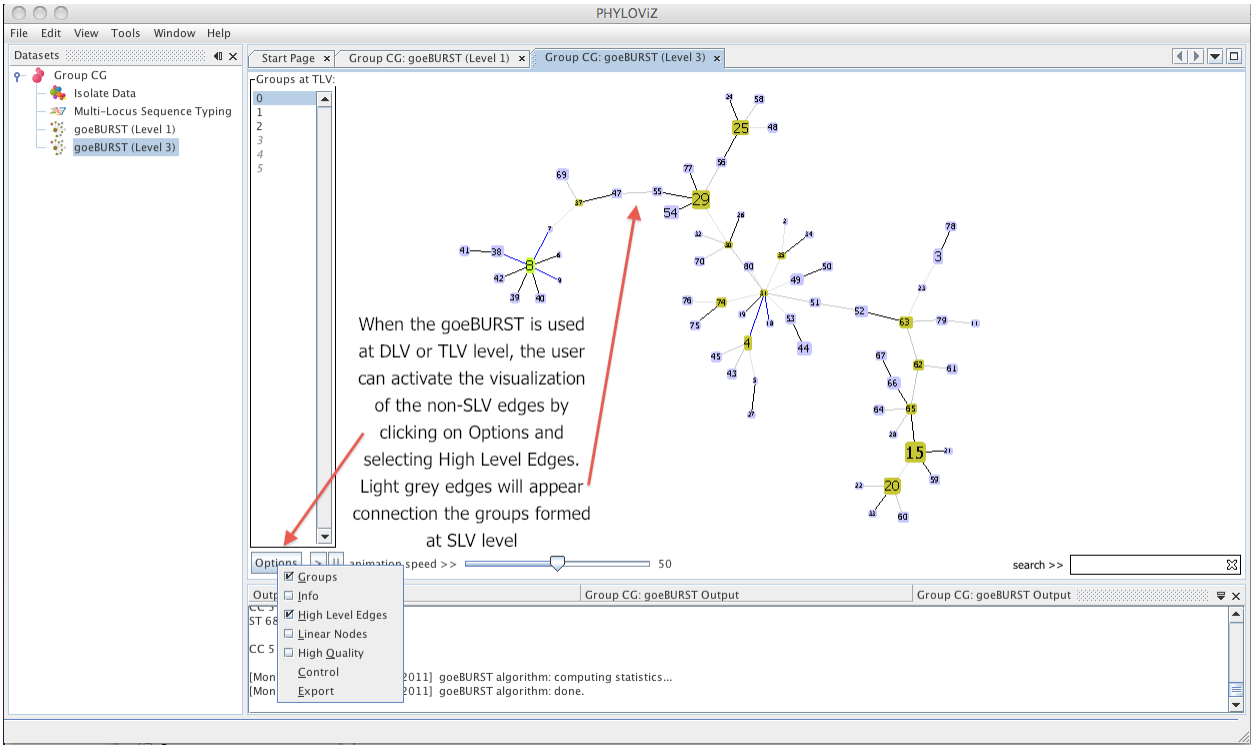
- Basic Interface



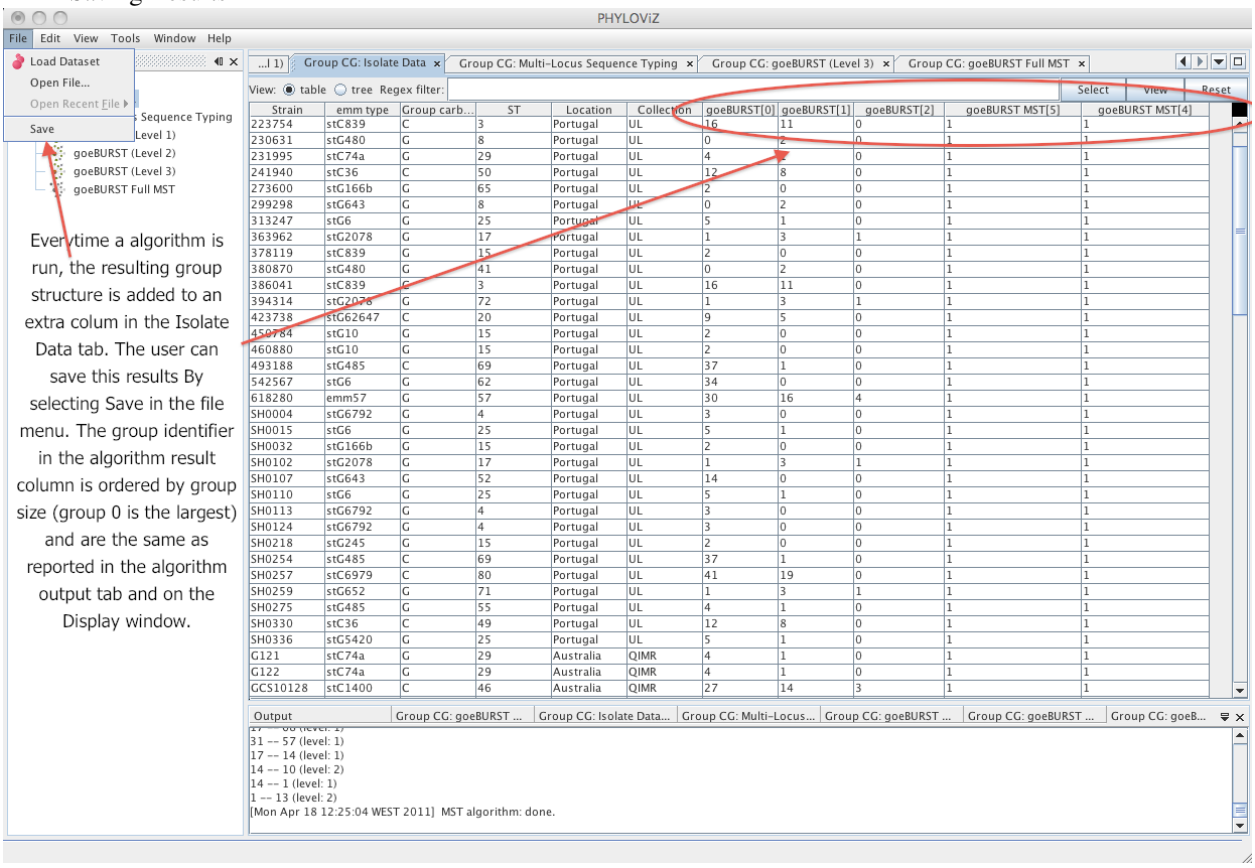
- SLV/DLV highlighting



- High Level Edges

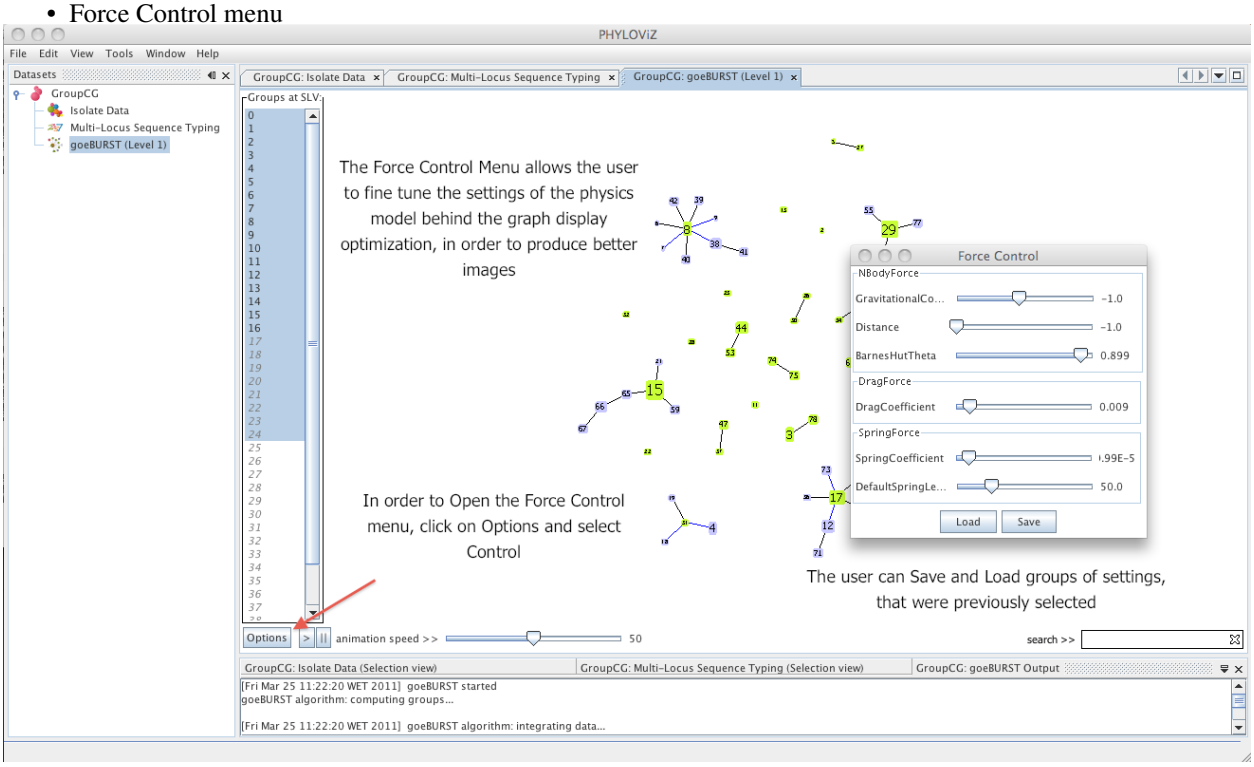


- Saving Results



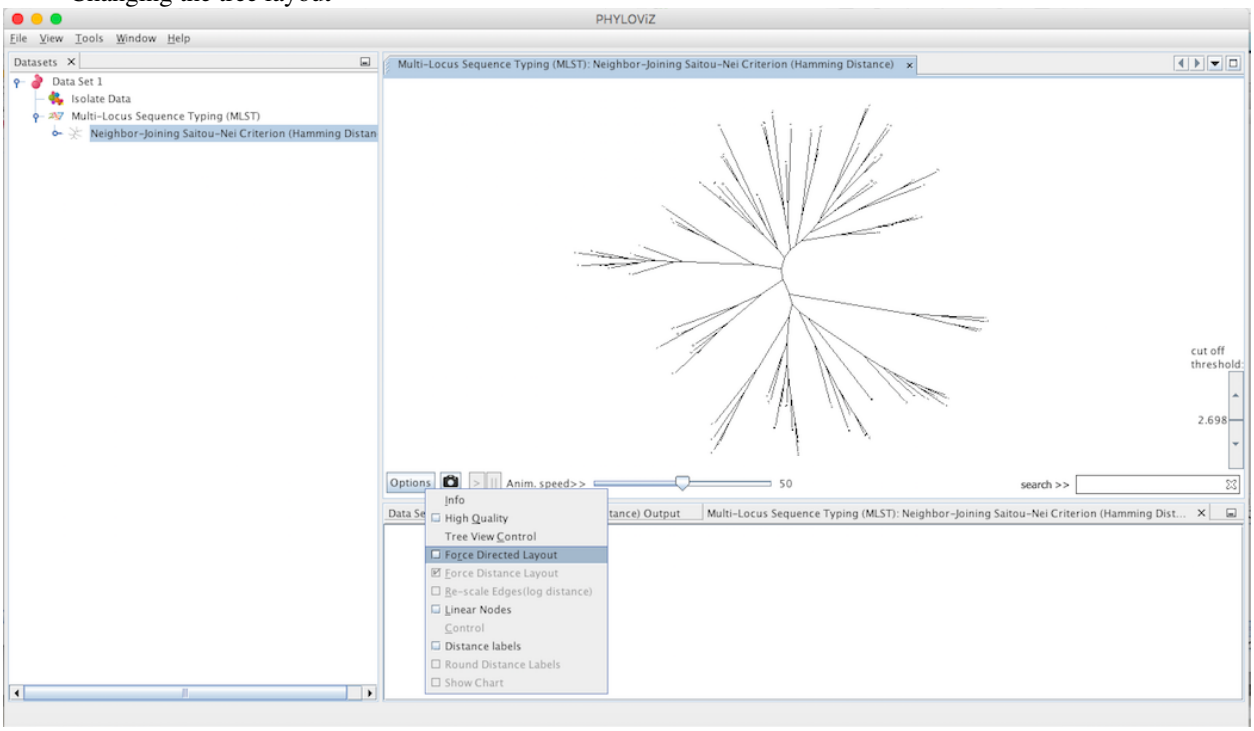
4.1.3 GoeBURS, GoeBURST Full MST and Neighbor Joinning features

- Force Control menu



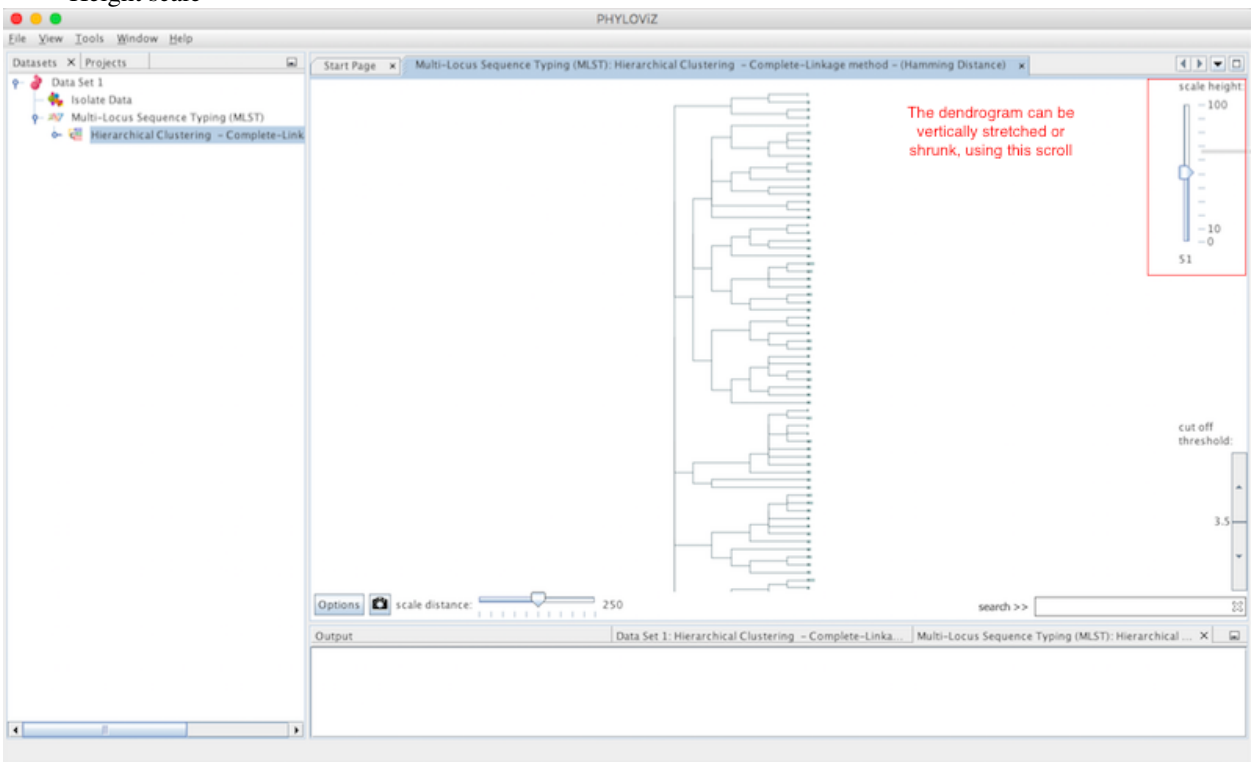
4.1.4 Neighbor Joinning features

- Changing the tree layout

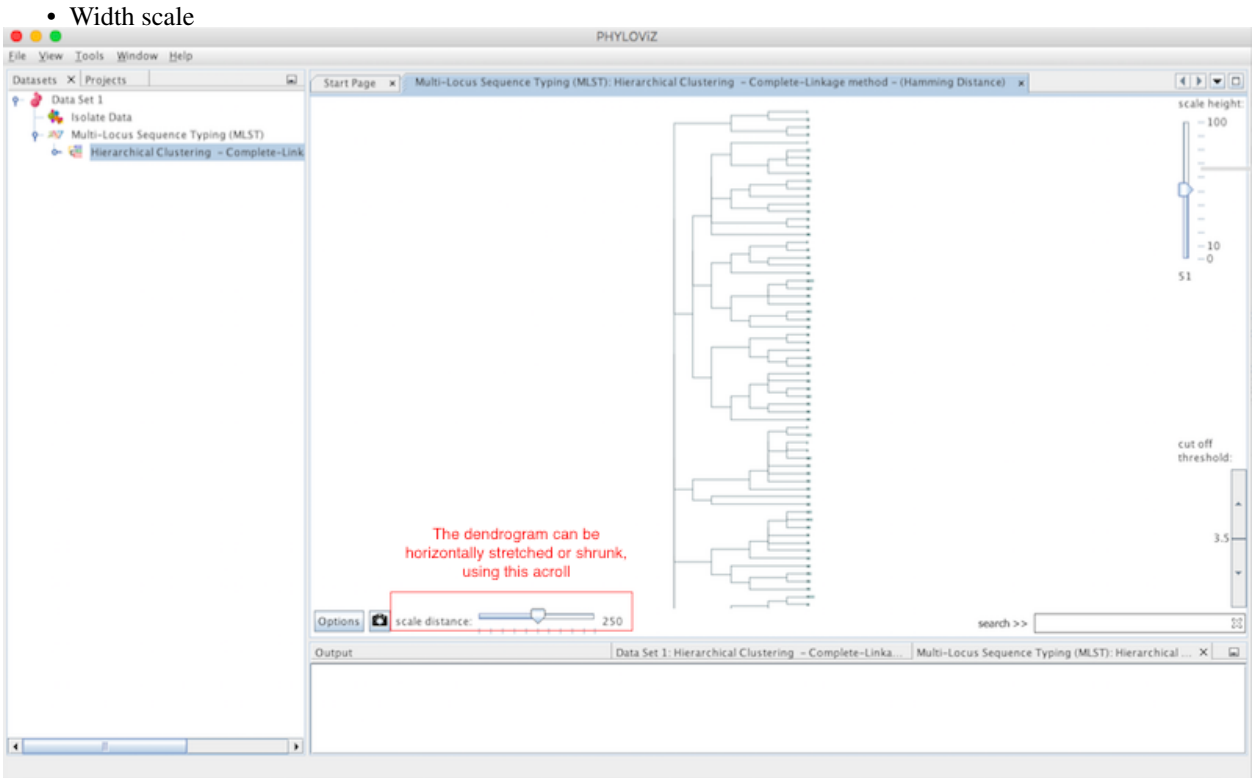


4.1.5 Hierarchical Clustering and Neighbor Joinning features

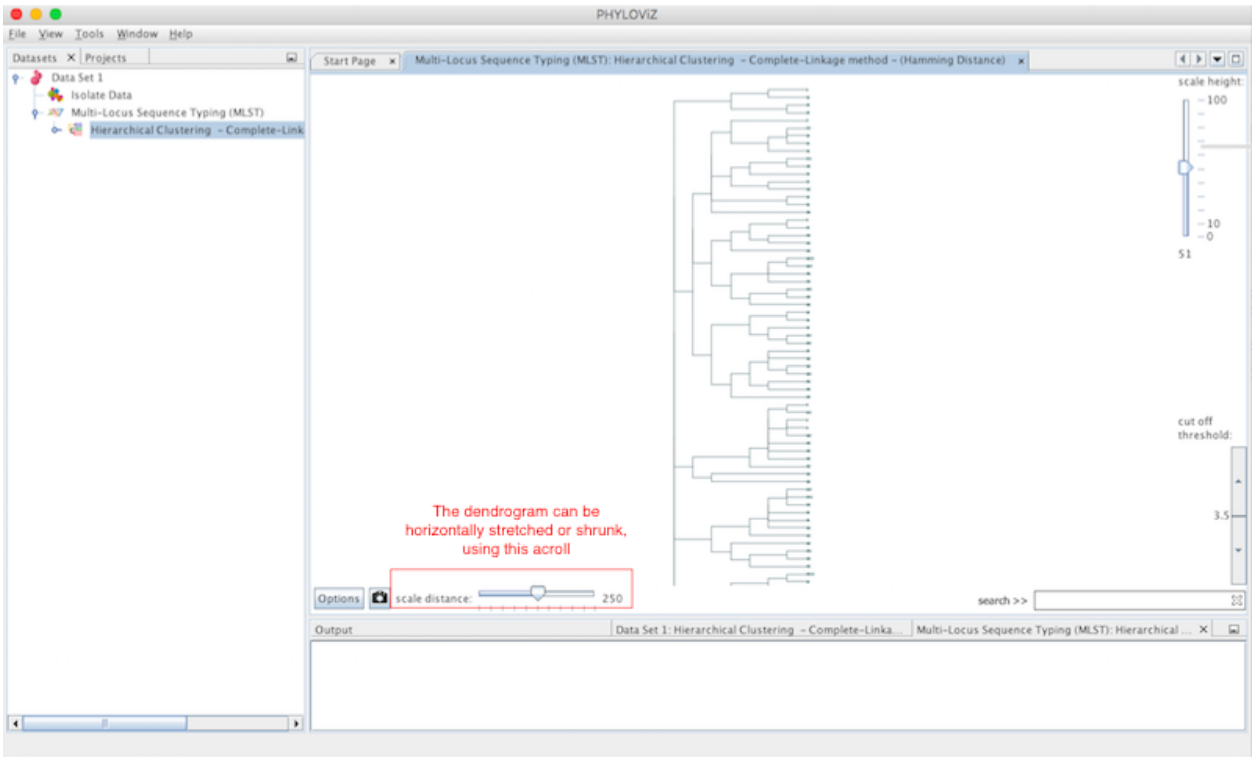
- Height scale



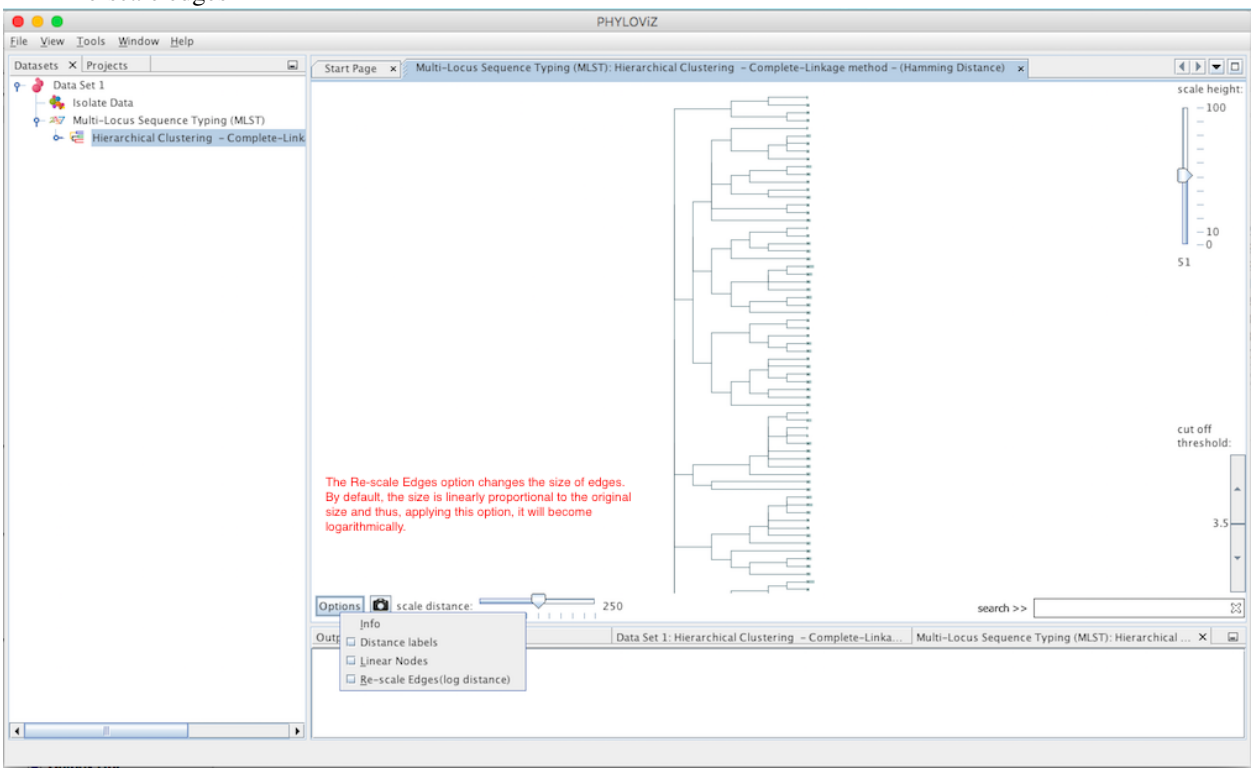
- Width scale



- Filter by distance (cut off threshold)



- Re-scale edges



4.2 Color conventions

Link colors for goeBURST results:

- Black - Link drawn without recourse to tiebreak rules,
- Blue - Link drawn using tiebreak rule 1 (number of SLVs),
- Green - Link drawn using tiebreak rule 2 (number of DLVs),
- Red - Link drawn using tiebreak rule 3 (number of TLVs),
- Yellow - Link drawn using tiebreak rule 4 or 5 (Frequency found on the data set and ST number , respectively),
- Gray - Links drawn at DLV (darker gray) or TLV (lighter gray) if the groups are constructed at DLV/TLV level.

Link colors for goeBURST Full MST results: The goeBURST Full MST algorithm links uses a grayscale with darker links having less differences between the profiles than the lighter gray links. To know the number of differences that the link represents click on the link in the Display window.

ST nodes colors:

- Light green - Group founder
- Dark green - Sub-group founder
- Light blue - Common node
- Red - Selected node

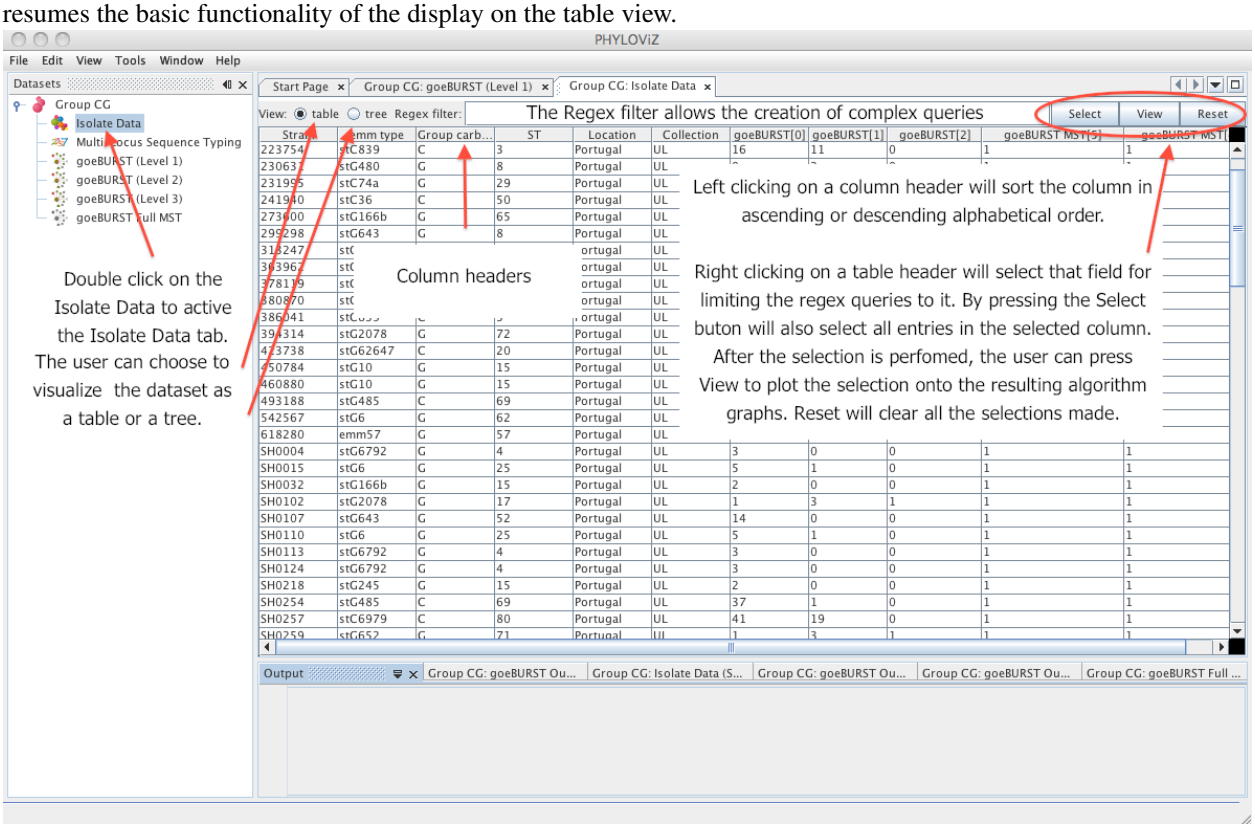
CHAPTER 5

Querying and visualizing the data

The main goal of PHYLOViZ is to provide a data visualization tool for overlaying accessory data on the data analysis algorithms result. This allows to test the method's adequacy to the data, or the proposal of novel hypothesis. This section will explain the basics on how this can be achieved in our software. The user can query the data using regular expressions, or simply manually selecting the desired fields from the table or, even just use the checkboxes in the tree view. Using your dataset and this instructions you should be able to create visualizations similar to the ones found in the PHYLOViZ website.

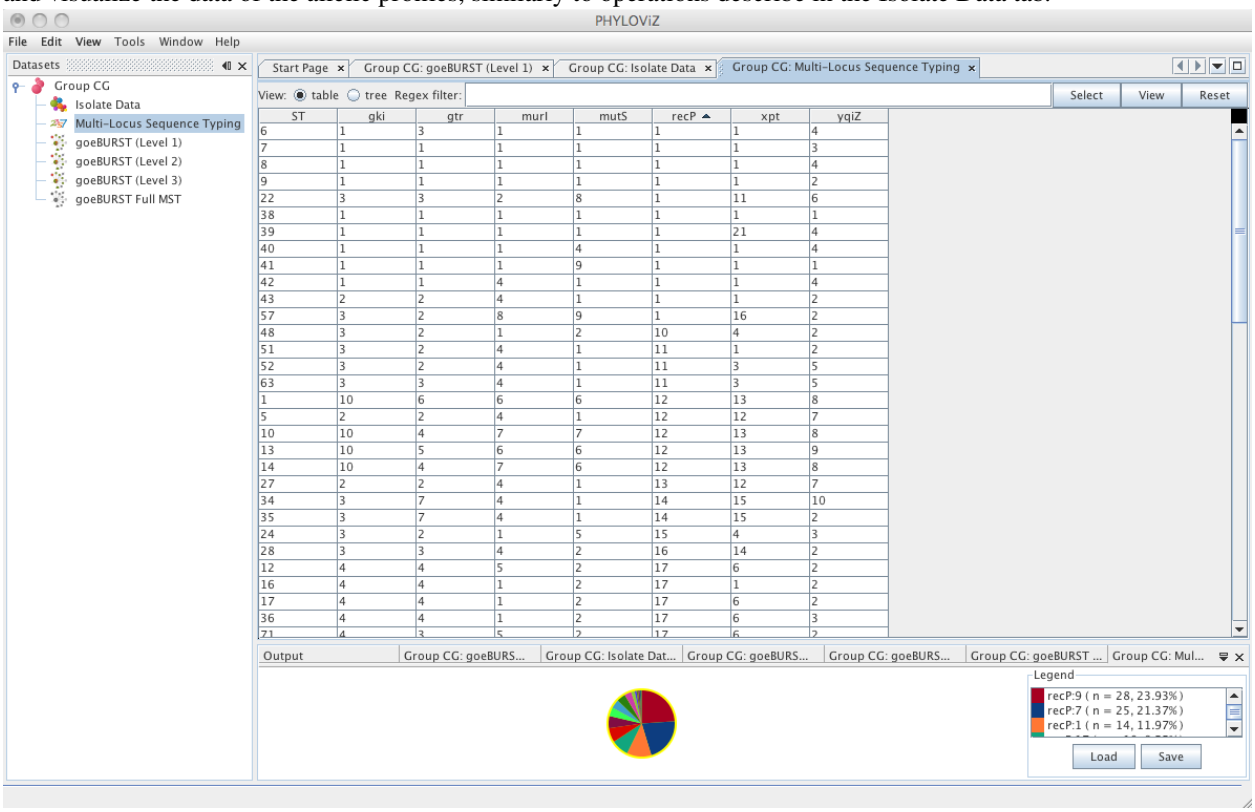
5.1 The isolate data tab

The *Isolate Data* tab is displayed by double clicking on the *Isolate Data* on the *Dataset* tree. The following screenshot resumes the basic functionality of the display on the table view.



5.2 The typing data tab

The *Typing Data* tab contains the allelic profiles loaded in the dataset. The name of displayed on the tab, and on the *Dataset* tree, is the name of the selected method during the *Load Dataset* procedure. The user can also query, select and visualize the data of the allelic profiles, similarly to operations describe in the Isolate Data tab.



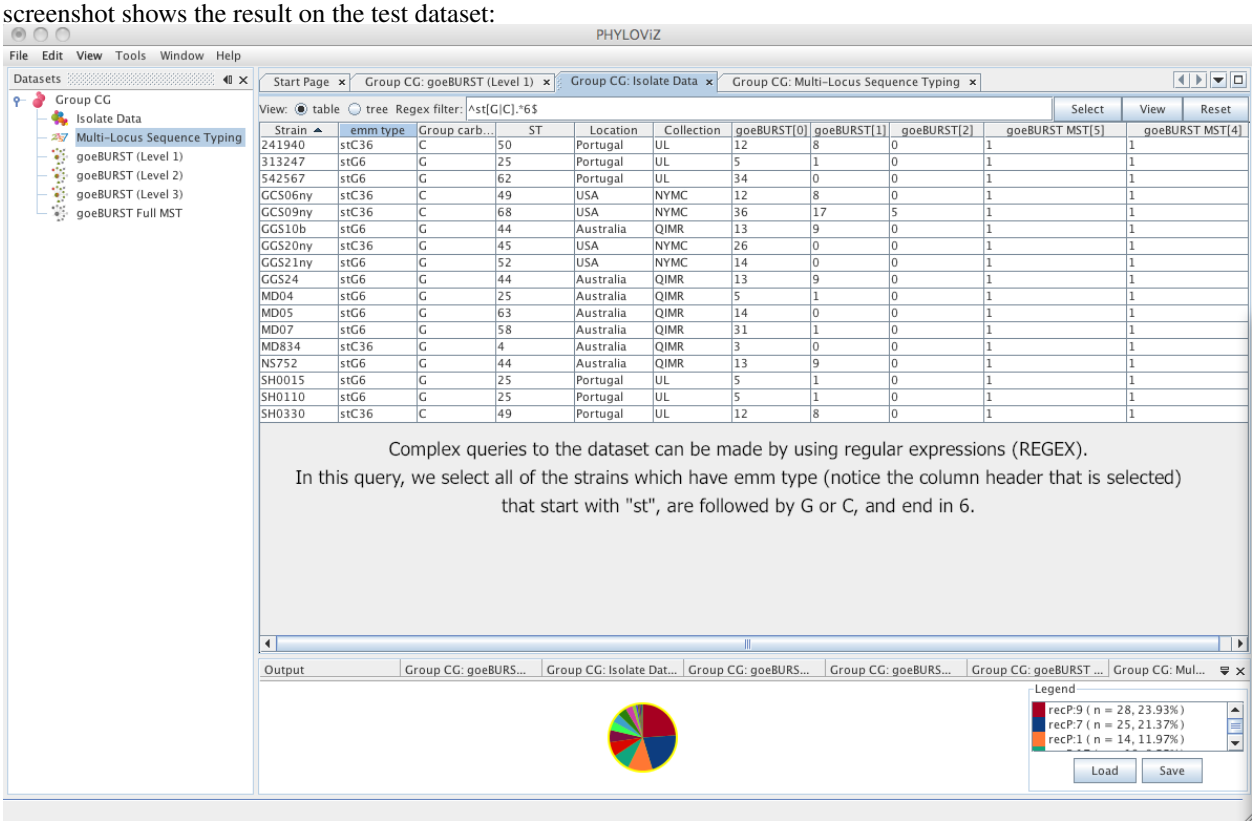
5.3 Regular expression primer

Some basic regular expressions that can be used in PHYLOViZ. For more complex expressions there are extensive tutorials on regular expressions online. Just search Regular Expression or regex.

- `.` (period mark) - represents any character.
- `[ ]` (square brackets) - Match anything inside the square brackets for one character position once and only once. Examples: `[40]` will match any field with `4` or `0`; `[7-9]` will match any field will `7`, `8` or `9` (`-` is the range separator).
- `^`(caret) - Starts with. Ex: `^P` will give you all the fields that start with a `P`. Inside the square brackets means negation. Example `[^a-c]` means anything not `a`, `b` or `c`.
- `$` (dollar sign) - Ends with. Ex. `7$` will give you all fields that end in a `7`.
- `?` (question mark) - Matches the preceding character 0 or 1 times only. Example: `colou?r` will find `color` and `colour`.
- `*` (asterisk) - Matches the preceding character 0 or more times. Example: `tre*` would find `tree`, `tread` and `trough`.
- `+` (plus) - Matches the preceding character 1 or more times. Example: `tre+` would find `tree`, `tread` but not `trough`.

- `{n}` (any integer between brackets) - Matches the preceding character exactly n times. Example: `AT[GC]{2}` would match `ATGC`, `ATCG`, `ATGG` or `ATCC` but not `ATGA`.

All these operators can be combined to create complex search expressions. For example : `^st[G|C].*6$` would find any field that starts with `st` followed by a `C` or a `G` then as 0 or more characters and ends with a `6`. The following screenshot shows the result on the test dataset:



5.4 Queries using the table view

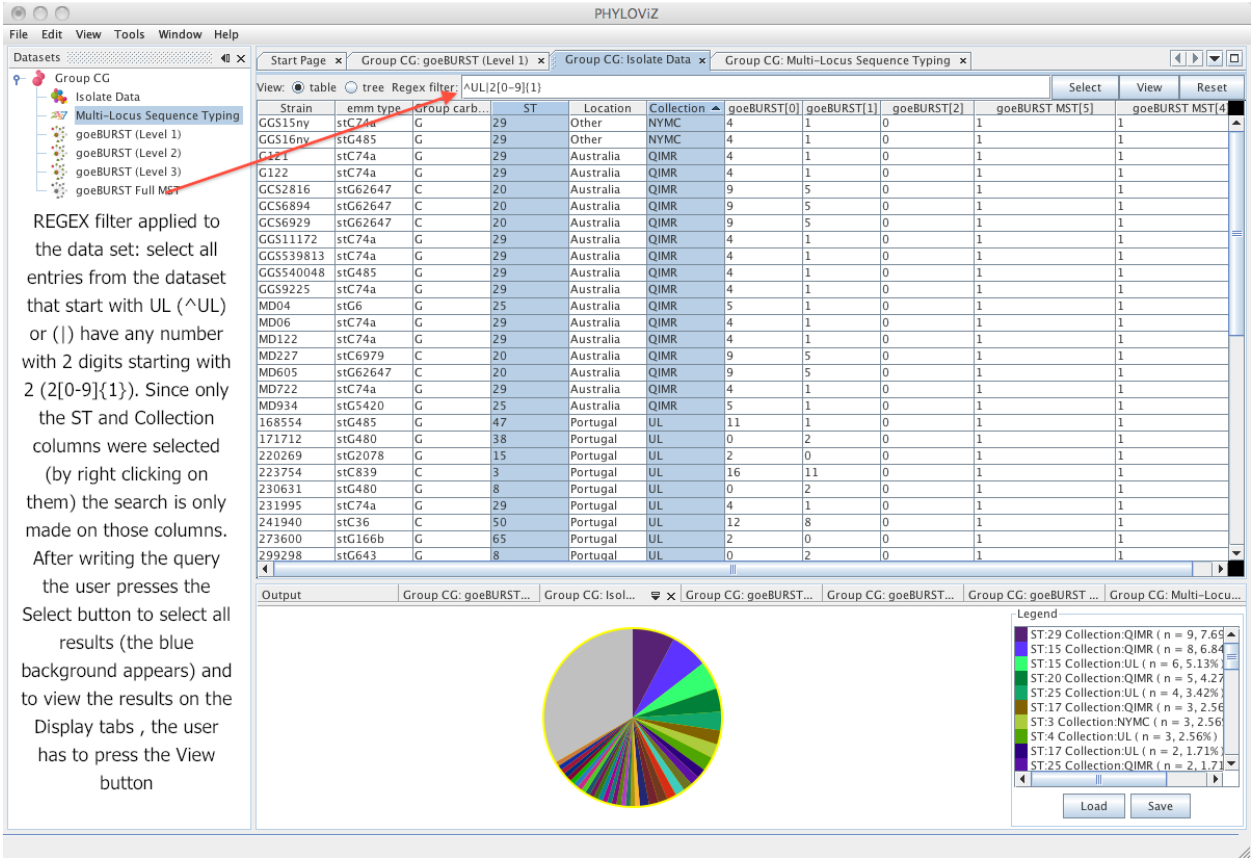
In the *Table* view of the *Data* tab you can manually select any field you want to represent by left clicking on it. That will automatically display all the entries with the selected value and not only the selected ones. To select multiple fields you can press the CTRL key (or CMD on Mac) while clicking on the desired fields. If you keep the SHIFT key pressed you can select ranges of cells.

You can also automatically select multiple columns by clicking with the right mouse button on the table headers and pressing the *Select* button.

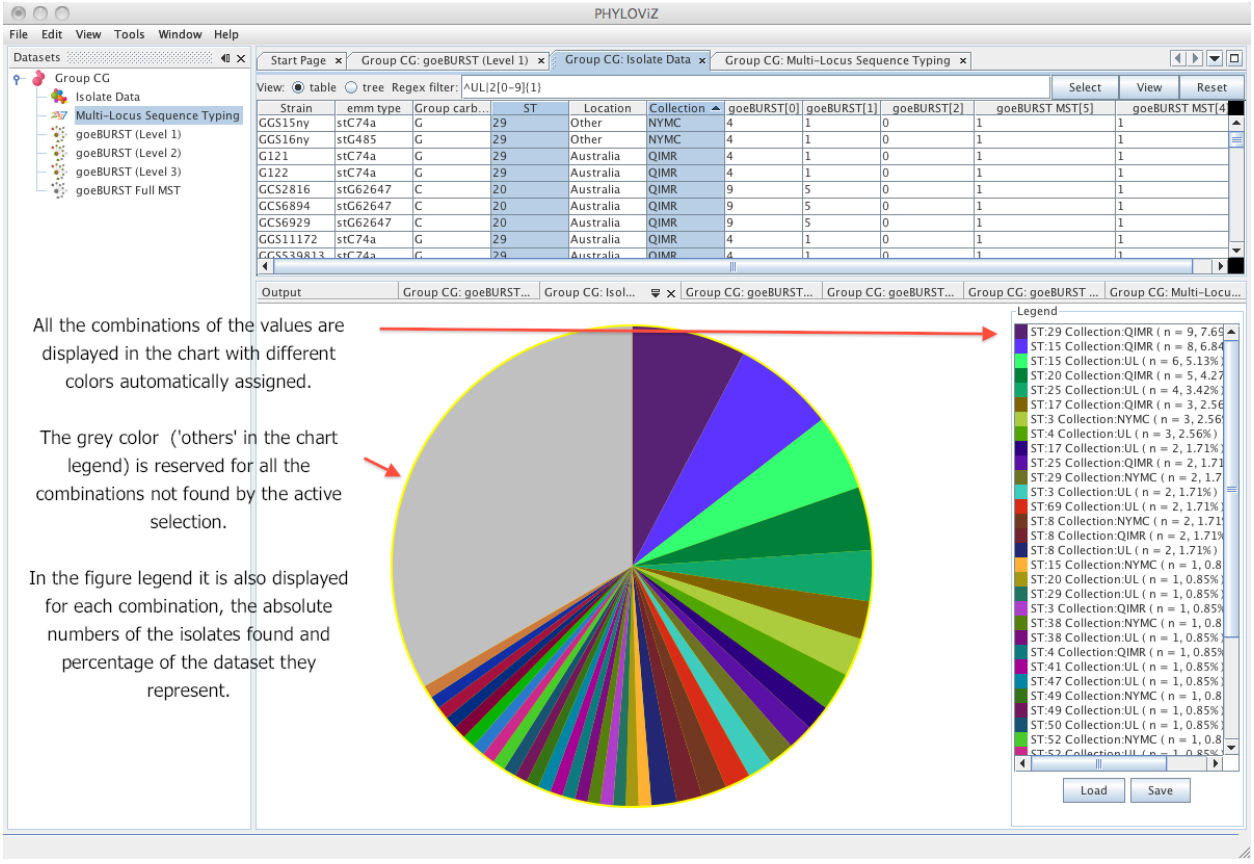
Finally to plot the data on the Display tab, press the View button, after all the desired selections are performed.

5.4.1 Query examples

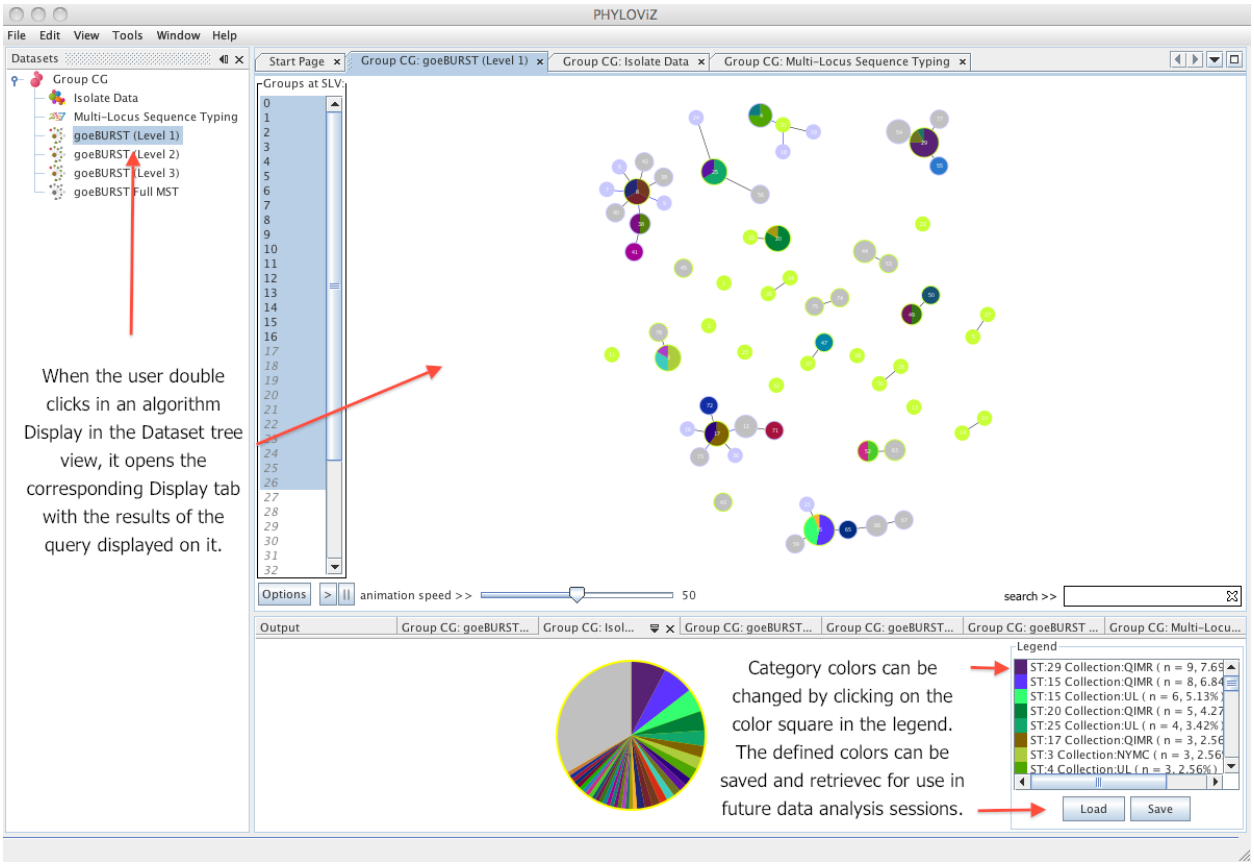
- Table view with selections



- Query results *Chart*



- Results on *Display Tab*

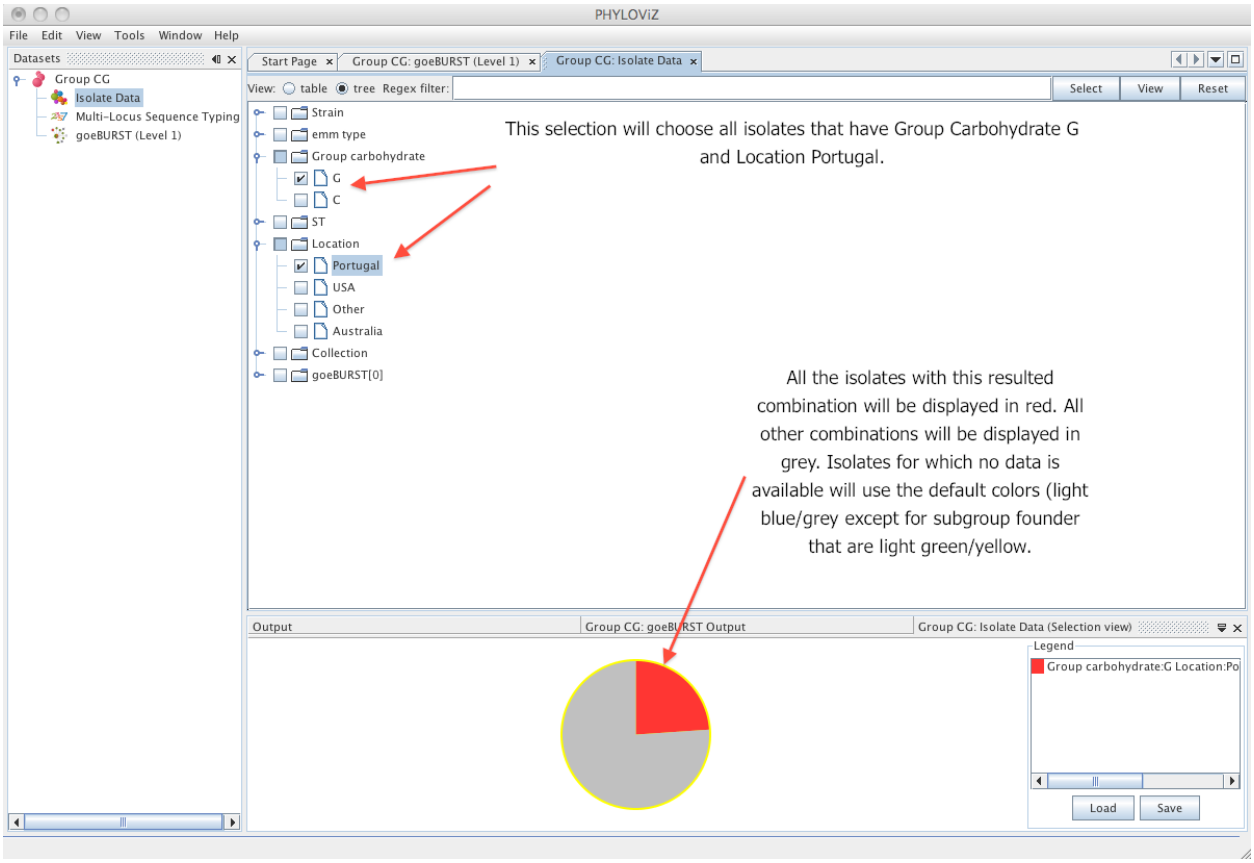


5.5 Queries using the tree view

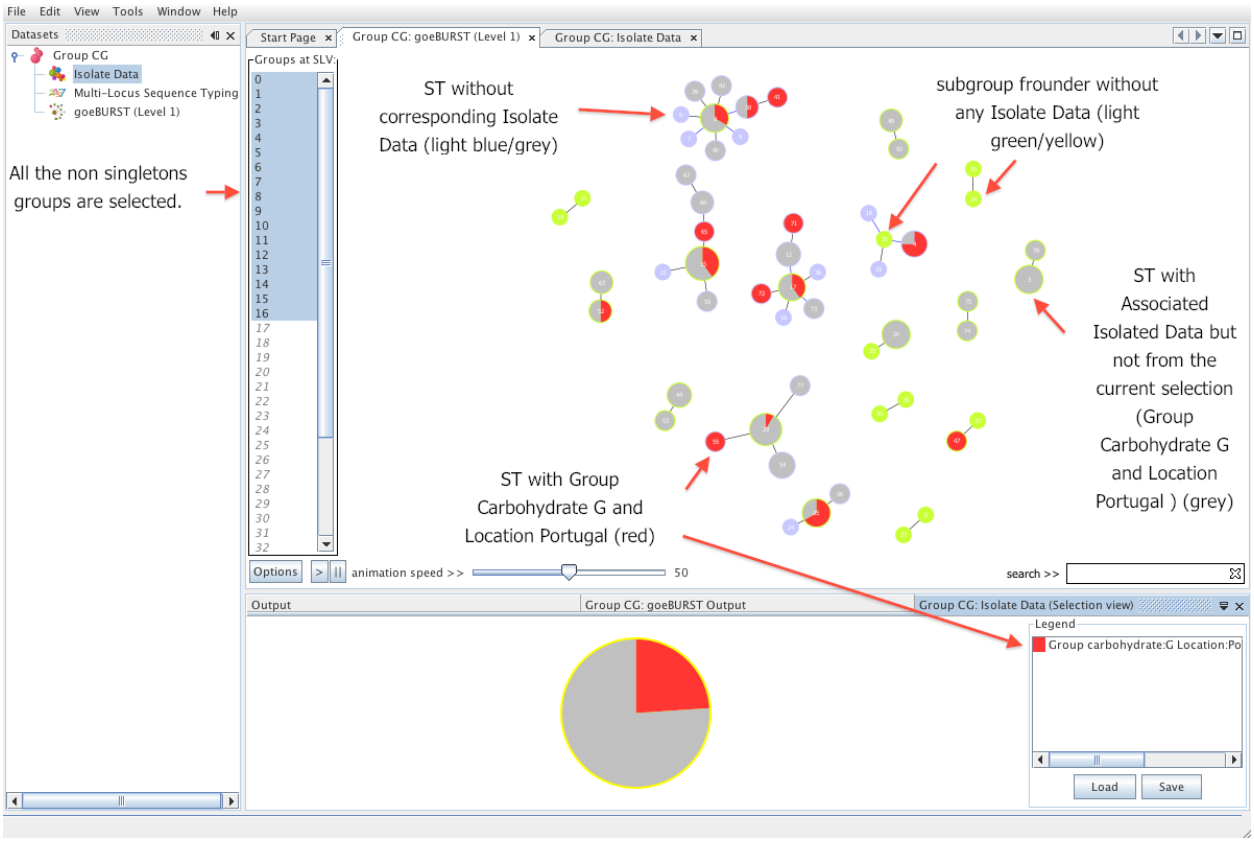
The *Tree* view offers a faster way to create simple queries. The user can also use the regex filter to search the dataset but all the possibilities for each dataset column are automatically indexed in a tree like manner. By pressing the *Select* button and switching to *Table* view the user can see the resulting selection. The users can alternate both views (Table and Tree) at will for creating the selection.

5.5.1 Query examples

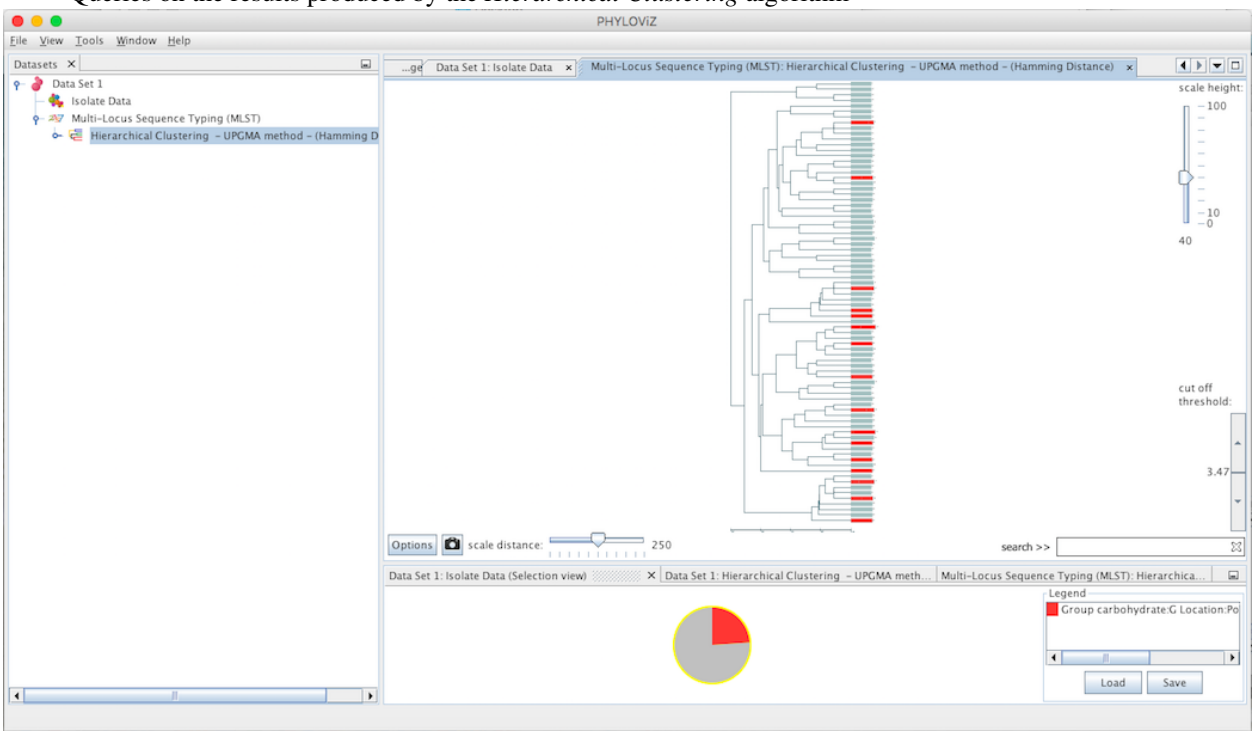
- Tree view with selections



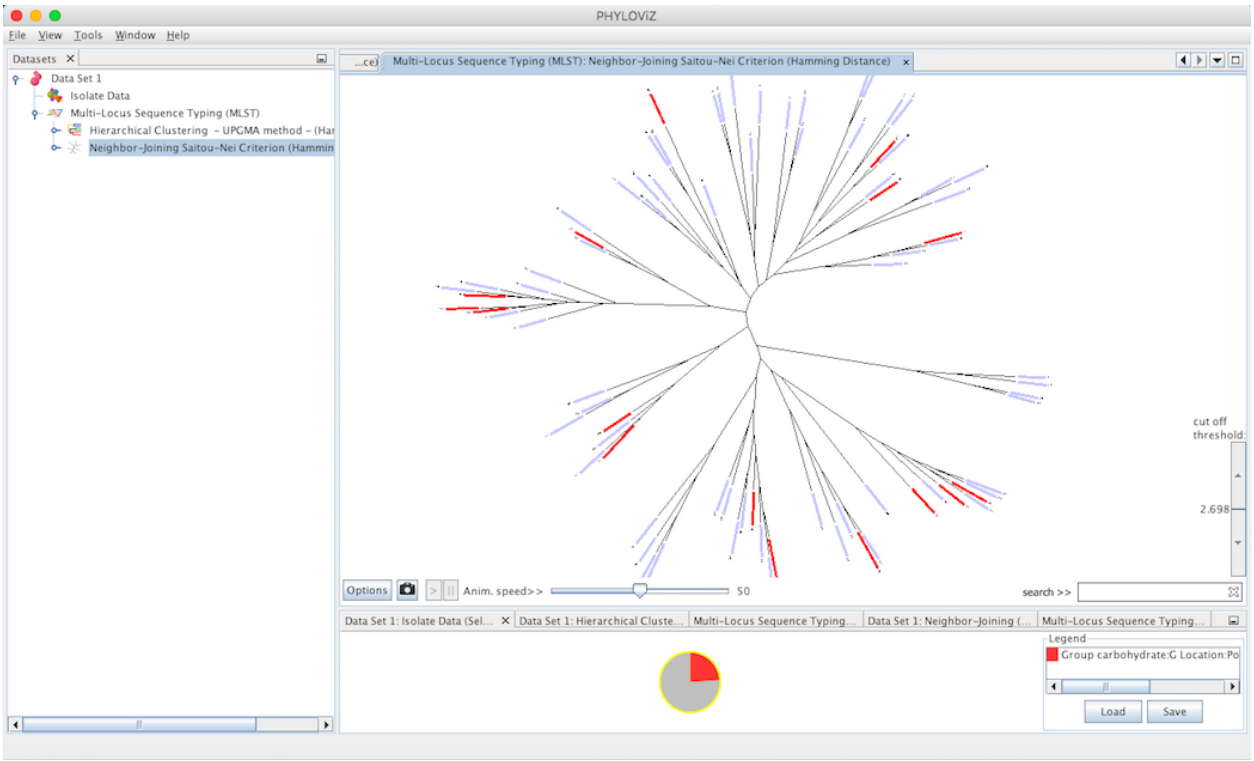
- Queries on the results produced by the *goeBURST* and *goeBURST Full MST algorithm*



- Queries on the results produced by the *Hierarchical Clustering* algorithm

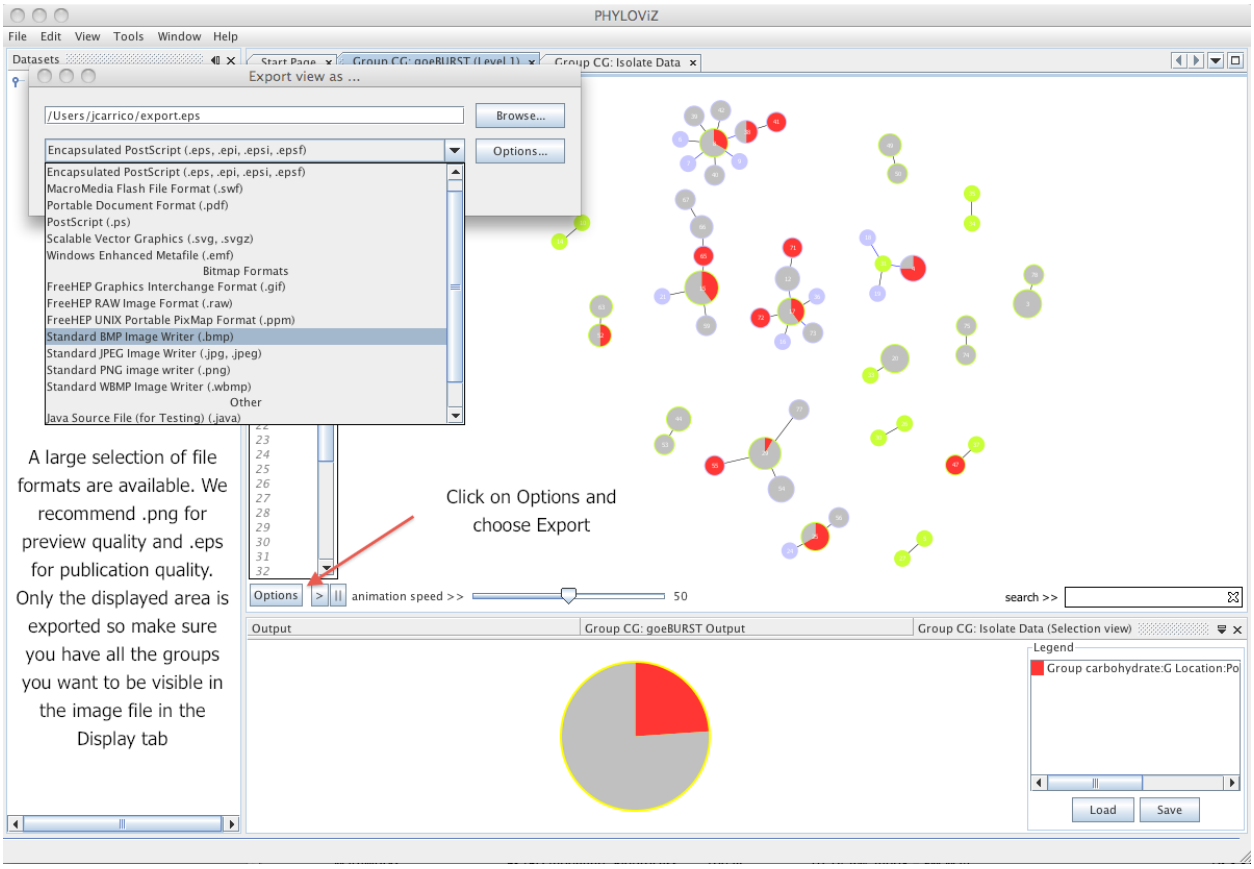


- Queries on the results produced by the *Neighbor Joinning* algorithm



5.6 Exporting the results to an image file

To export the resulting graphs to an image file. Click on the *Options* button and choose *Export*. Select the adequate file format for the intended purpose. We recommend the use of png images for presentation quality and eps for publication quality.


PHYLOViZ
File Edit View Tools Window Help
Datasets
Start Page
Group CG: goeBURST (Level 1)
Group CG: Isolate Data
Export view as ...
/Users/jcarrico/export.eps
Browse...
Encapsulated PostScript (.eps, .epi, .epsi, .epsf)
Options...
Encapsulated PostScript (.eps, .epi, .epsi, .epsf)
MacroMedia Flash File Format (.swf)
Portable Document Format (.pdf)
PostScript (.ps)
Scalable Vector Graphics (.svg, .svgz)
Windows Enhanced Metafile (.emf)
Bitmap Formats
FreeHEP Graphics Interchange Format (.gif)
FreeHEP RAW Image Format (.raw)
FreeHEP UNIX Portable PixMap Format (.ppm)
Standard BMP Image Writer (.bmp)
Standard JPEG Image Writer (.jpg, .jpeg)
Standard PNG image writer (.png)
Standard WBMP Image Writer (.wbmp)
Other
Java Source File (for Testing) (.java)
A large selection of file formats are available. We recommend .png for preview quality and .eps for publication quality. Only the displayed area is exported so make sure you have all the groups you want to be visible in the image file in the Display tab
Click on Options and choose Export
Options
animation speed >>
50
search >>
Output
Group CG: goeBURST Output
Group CG: Isolate Data (Selection view)
Legend
Group carbohydrate:G Location:Po
Load
Save

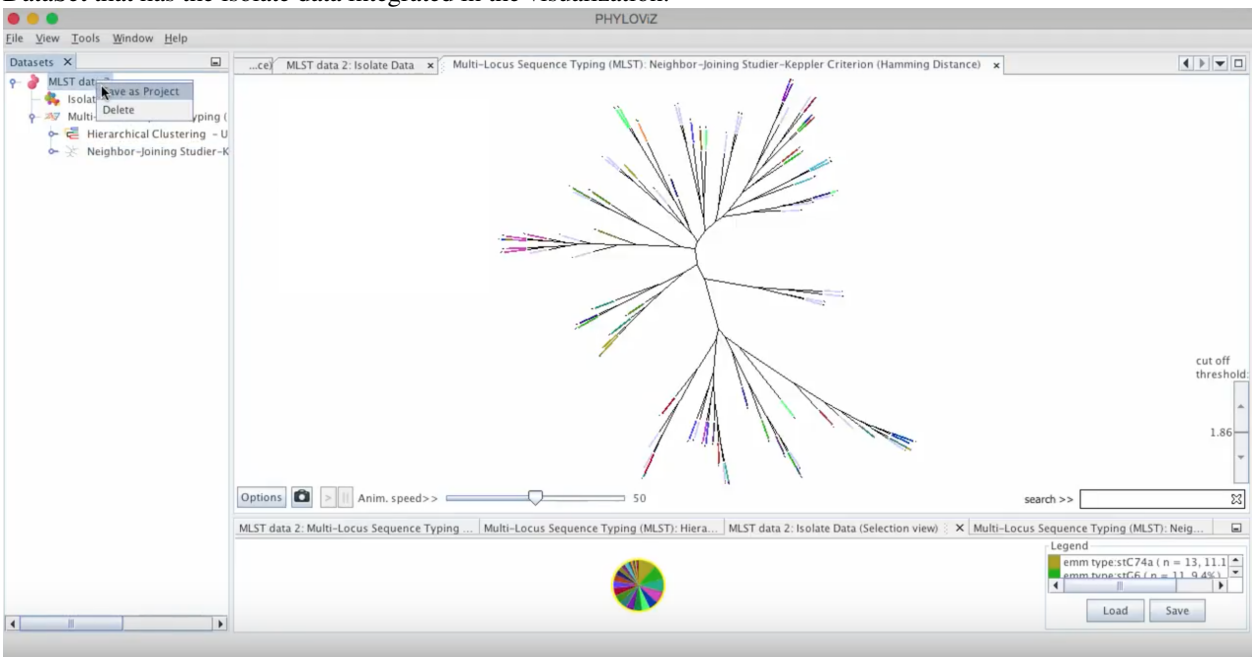
CHAPTER 6

Project management

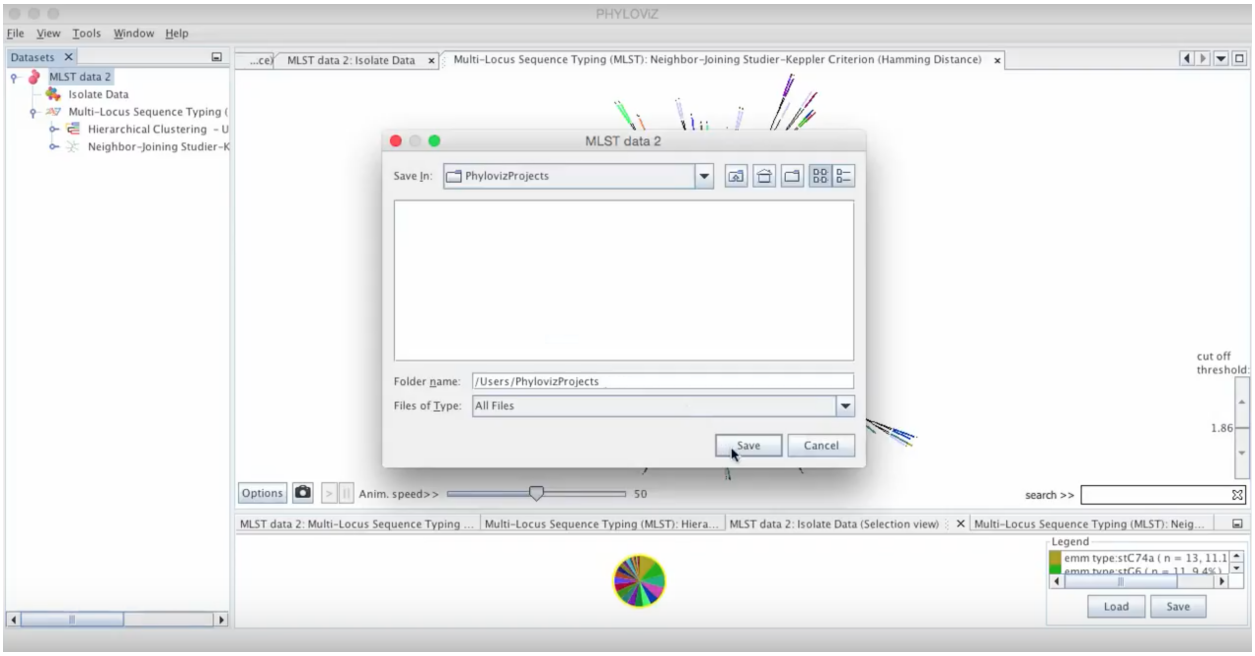
A PHYLOViZ Project allows users to save their ongoing studies and update them as needed. It is a time-saving feature when working with large data sets and essential for efficiently sharing results, since the saved projects can then be shared. Each project includes the data under analysis, results of inference algorithms, visualization serializations and related customizations.

6.1 Saving

Right click on the dataset you would like to save and choose the option *Save as Project*. As you can see we'll save a DataSet that has the isolate data integrated in the visualization.

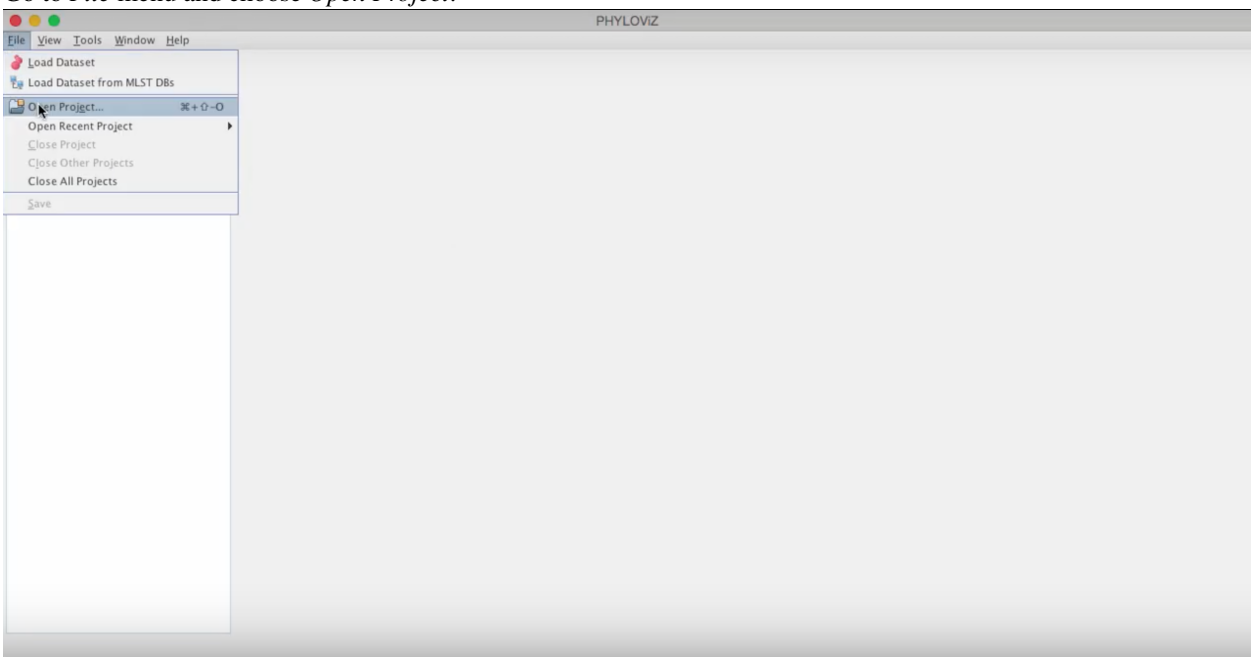


Finally you can choose where to save your project. A dialog appears if you are overriding an existing project or creating a new one with a name that was already taken in the chosen directory.

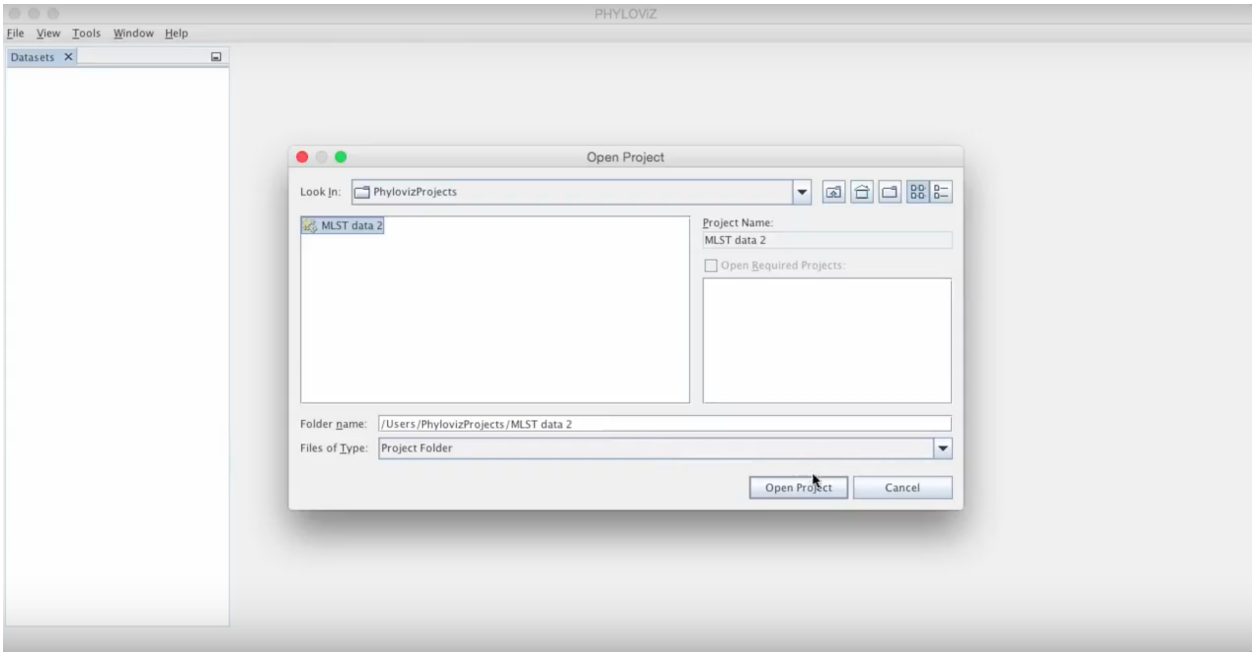


6.2 Loading

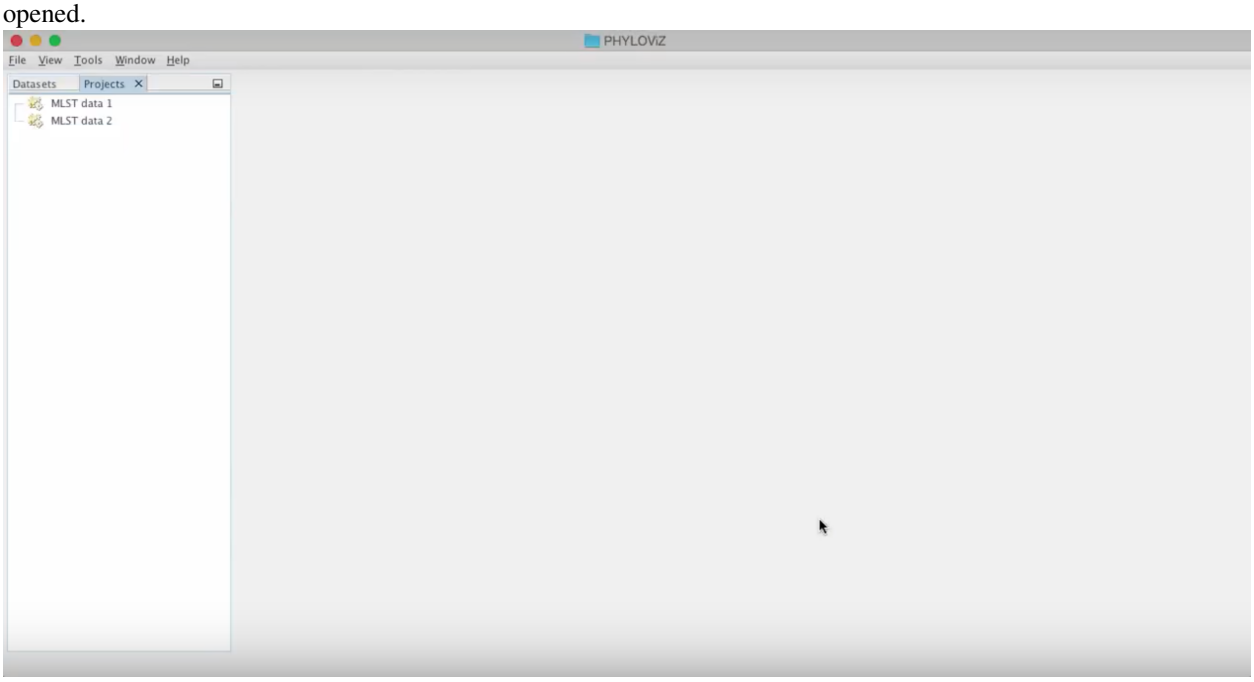
Go to *File* menu and choose *Open Project*.



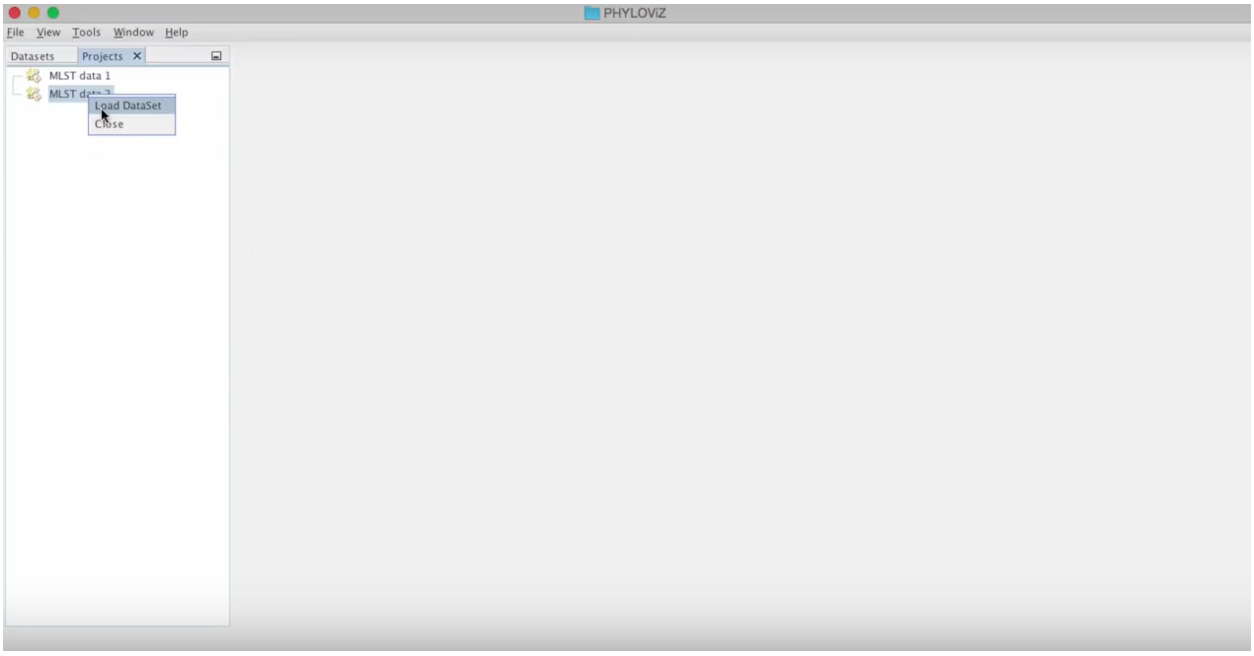
The next step is to find the project that you would like to load. After finding it click on *Open Project*.



This action will open the *Projects Tab* where you can see your project listed and many others that were previously opened.



Now for restoring the study just right click on the project and select *Load DataSet*. This will open the *Dataset's Tab* with your saved study.



The project is now loaded with all the study that was done before as we can see in the following screenshot. You can check that the isolated data integrated on saving was restored completely.