# **PHYLOViZ Documentation**

Release 2.0

**PHYLOViZ** Team

September 07, 2016

1	Download and install         1.1       Binaries         1.2       Source	<b>3</b> 3 4
2	2.2 Loading a Dataset	5 6 10
3	3.1       goeBURST algorithm       1         3.2       goeBURST Full MST algorithm       1         3.3       Hierarchical Clustering       1	<b>15</b> 15 19 22 26
4	4.1 Interface features	<b>31</b> 31 39
5	5.1       The isolate data tab       4         5.2       The typing data tab       4         5.3       Regular expression primer       4         5.4       Queries using the table view       4         5.5       Queries using the tree view       4	<b>41</b> 42 42 43 46 49
6	6.1 Saving	<b>51</b> 51 52

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.

#### **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.

#### 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.

00	0				🖹 san	pleAPfi	le.txt	<u> </u>	00	0			samp	leADfile.tx	t		1
ST	gki	gtr	murI	mutS	recP	xpt	ygiZ	-	Strain	emm typ	e	Group	carbohydrate		Location	Collection	6
1	10	6	6	6	12	13	8		168554	stG485	G	47	Portugal	UL			
2	5	4	4	1	2	15	2		171712	stG480	G	38	Portugal	UL			
3	5	3	4	1	6	2	1		220269	stG2078	G	15	Portugal	UL			
4	2	2	4	1	8	7	2		223754	stC839	С	3	Portugal	UL			- 11
5	2	2	4	1	12	12	7		230631	stG480	G	8	Portugal	UL			
6	1	3	1	1	1	1	4		231995	stC74a	G	29	Portugal	UL			
7	1	1	1	1	1	1	3		241940	stC36	С	50	Portugal	UL			
8	1	1	1	1	1	1	4		273600	stG166b	G	65	Portugal	UL			
9	1	1	1	1	1	1	2		299298	stG643	G	8	Portugal	UL			
10	10	4	7	7	12	13	8		313247	stG6	G	25	Portugal	UL			
11	11	3	4	1	2	7	5			stG2078	G	17	Portugal	UL			
12	4	4	5	2	17	6	2				G	15	Portugal	UL			- 11
13	10	5	6	6	12	13	9		380870	stG480	G	41	Portugal	UL			
14	10	4	7	6	12	13	8		386041	stC839	С	3	Portugal	UL			
15	3	3	2	2	9	8	2			stG2078	G	72	Portugal	UL			
16	4	4	1	2	17	1	2		423738	stG6264		C		rtugal	UL		
17	4	4	1	2	17	6	2		450784	stG10	G	15	Portugal	UL			
18	4	2	4	1	8	7	2		460880	stG10	Ğ	15	Portugal	UL			
19	3	8	4	1	8	7	2		493188	stG485	Č.	69	Portugal	UL			
20	3	3	ż	8	9	6	6		542567	stG6	Ğ	62	Portugal	ŰL			
21	3	8	2	2	9	8	2		618280	emm57	Ğ	57	Portugal	UL			
22	3	3	2	8	1	11	6		SH0004	stG6792		4	Portugal	UL			
23	3	3	4	1	3	1	1		SH0015	stG6	G	25	Portugal	UL			
24	3	2	1	5	15	4	3		SH0032	stG166b		15	Portugal	UL			
25	3	2	1	5	7	4	3			stG2078		17	Portugal	UL			_
26	3	2	1	1	2	10	2		SH0107	stG643		52	Portugal	UL			
20	2	2	4	1	13	12	7		SH0110	stG6	G	25	Portugal	UL			
28	3	3	4	2	15	14	2			st66792		4	Portugal	UL			
20 29	3	2	4	2	7		3			stG6792		4	Portugal	UL			
29 30	3	2	4	1	2	1 10	2		SH0218		G	15	Portugal	UL			
30 31	3	2	4	1	8	10	2		SH0210		C	69	Portugal	UL			
31	3	2	4		o 4	10	5		SH0254	stC6979		80	Portugal	UL			
33		2		1 8	9				SH0259		G	71	Portugal	UL			
	3 3	7	2 4	0 1		6	6		SH0255		G	55	Portugal	UL			
34		2			14	15	10		SH0275	st0405 stC36	C		Portugal	UL			
35	3	4	4	1	14	15	2			stG5420		49		UL			
36	4		1	2	17	6				st05420 st074a		25	Portugal				
37	3	2	1	1	4	1	3		G121 G122		G	29 29	Australia Australia	QIMR QIMR			
38	1	1	1	1	1	1	1								OTHD		
39	1	1	1	1	1	21	4		GCS1012		stC1400			stralia	QIMR		
40	1	1	1	4	1	1	4			stG6264		C		stralia	QIMR		
41	1	1	1	9	1	1	1			stG6264		C		stralia	QIMR		
42	1	1	4	1	1	1	4			stG6264		C		stralia	QIMR		
43	2	2	4	1	1	1	2			stG166b		56	Australia	QIMR			
44	2	2	4	2	3	7	1				G	12	Australia	QIMR			
45	2	2	4	10	8	19	2		GGS10b		G	44	Australia	QIMR	0.1 MD		
46	2	4	4	1	19	17	6		GGS1117		stC74a	G		stralia	QIMR		
47	3	2	1	1	20	1	3		GGS1154			G		stralia	QIMR		
48	3	2	1	2	10	4	2			stG4831		74	Australia	QIMR			
49	3	2	3	1	5	5	2	4	GGS19	stC1400		64	Australia	QIMR			4
50	3	2	3	1	5	18	2		GGS2	stG10	G	15	Australia	QIMR			A. V
51	3	2	4	1	11	1	2		GGS24	stG6	G	44	Australia	QIMR			- 1
52	3	2	4	1	11	3	5	1	GGS430	stG643	G	12	Australia	OIMR			

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.

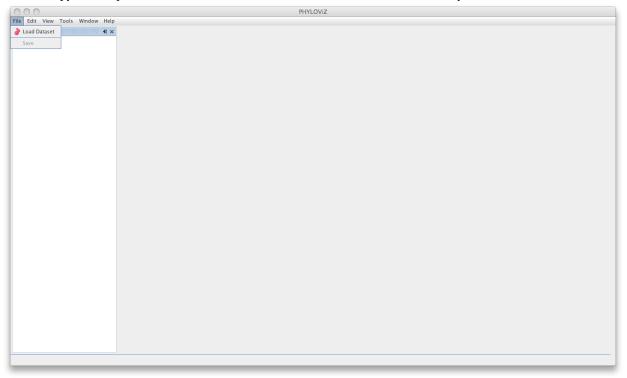
You can find an example of MLST data correctly formatted here. Note that in this file several STs are represented by more than one isolate (e.g. ST3 was found in 6 isolates).

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.

000		PHYLOViZ	
File Edit View Tools Window Help			
Datasets Carteria Car	000	Load a Dataset	
			_
	Steps  1. Dataset 2. Typing Data 3. Isolate Data	Create a Dataset name and choose the Dataset type depending on the typing data to be used. The options are Multi-Locus Sequence Typing (MLST) and Multi-Locus Variable Number of Tandem Repeats Analysis (MLVA).	
		< Back Next > Finish Cancel Help	

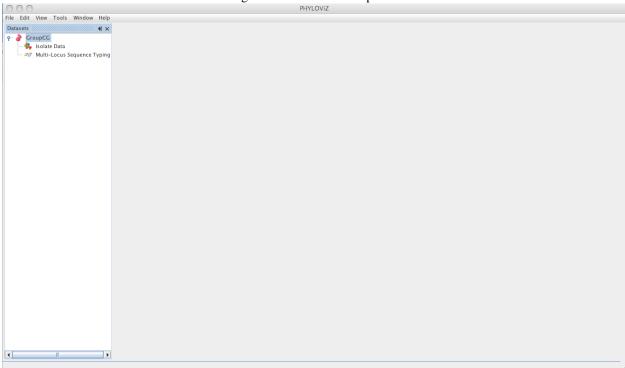
The next step is loading the allelic profile data for the method you selected.

000		PHYLOVIZ
File Edit View Tools Window Help		
Datasets		
	000	Load a Dataset
	Steps	Typing Data
	1. Dataset 2. Typing Data	File: //Users/jcarrico/Desktop/Phyloviz/sampleAPfile.txt Browse
	3. Isolate Data	Load a tab separated file with the first row containing column headers (field names). The first column should be a Sequence Identifier, such as the ST for MLST data. Other columns should be the different loci used in the typing scheme. Example of a MLST dataset for <i>Streptococcus dysgalactiae subspecies equisimilis.</i> $\begin{array}{cccccccccccccccccccccccccccccccccccc$
	2357	This files should not contain missing data. If any allele missing data is found the entire line is discarded.
		< Back Next > Finish Cancel Help

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.

0.0.0			
000		PHYLOVIZ	
File Edit View Tools Window Help Datasets	000	Load a Dataset	
	Steps  1. Dataset 2. Typing Data 3. Isolate Data	Isolate Data File: //Users/jcarrico/Desktop/Phyloviz/sampleADfile.txt Key: Strain If you train If you train try ou abserf croup carbohydrate algor ST heade (Location Chool Collection allelic Prome data.rt une metra dues not appear in the drop duown Update button. If you don't have any available data or don't wish leave this fields blank an press Finish.	phic ene presence i to n the y with column entifier. Jata to the menu press the
		< Back Next > Finish Cance	1 Help

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.

dit View Tools Window Help													
ts 40 ×	GroupC	G: Isolate Data 🤉	¢										
GroupCG		ible 🔾 tree Re										Select	View Res
🛼 Isolate Data	Churcher		Group carb.	ST	Location	Collection						Select	view Kes
Multi-Locus Sequence Typing	168554	stG485	G	47	Portugal	UL							
	171712 220269	stG480 stG2078	G	38 15	Portugal Portugal	UL							
	223754	stC839	C	3	Portugal	UL							
	230631	stG480	G	8	Portugal	UL							
	231995	stC74a	G	29	Portugal	UL	_						
	241940 273600	stC36 stG166b	C G	50 65	Portugal Portugal	UL							
	299298	stG643	G	8	Portugal	UL	-						
	313247	stG6	G	25	Portugal	UL							
	363962 378119	stG2078	G	17	Portugal	UL							
	380870	stC839 stG480	G	41	Portugal Portugal	UL	-						
	386041	stC839	C	3	Portugal	UL							
	394314	stG2078	G	72	Portugal	UL	-						
	423738 450784	stG62647 stG10	C G	20	Portugal Portugal	UL	-						
	460880	stG10	G	15	Portugal	UL							
	493188	stG485	C	69	Portugal	UL							
	542567 618280	stG6	G	62 57	Portugal	UL	-						
	SH0004	emm57 stG6792	G	4	Portugal Portugal	UL	-						
	SH0015	stG6	G	25	Portugal	UL							
	SH0032	stG166b	G	15	Portugal	UL							
	SH0102 SH0107	stG2078 stG643	G	17 52	Portugal	UL	_						
	SH0107 SH0110	stG643	G	25	Portugal Portugal	UL	-						
	SH0113	stG6792	G	4	Portugal	UL	-						
	SH0124	stG6792	G	4	Portugal	UL							
	SH0218 SH0254	stG245 stG485	G C	15 69	Portugal Portugal	UL							
		Isolate Data (Sel											
	Grouped.	isolate Data (Sel	ection view)									Legend	
												I	
						PHYLOVi	z					Load	Save
ां View Tools Window Help s अधि x		G: Isolate Data →	د) GroupCG	: Multi–Locus S	Sequence Typin		z					Load	
rit View Tools Window Help s 40 × GroupCG	GroupCO			: Multi–Locus S	Sequence Typir		z					Load	
rit View Tools Window Help s 40 × GroupCG Isolate Data	GroupCO View:  ta	ible 🔾 tree Re		: Multi–Locus S	Sequence Typin			yqiZ					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	ible 🔾 tree Re gki 10	gex filter: gtr 6	murl 6	mutS 6	recP	xpt 13	yqiZ 8					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	uble 🔾 tree Re gki 10 5	gex filter: gtr 6 4	murl 6 4	mutS 6	recP 12 2	xpt 13 15	8					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	ible 🔾 tree Re gki 10	gex filter: gtr 6 4 3	murl 6	mutS 6	recP	xpt 13	8					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	ble C tree Re gki 10 5 5 5	gex filter: gtr 6 4 3 2 2	6 4 4	mutS 6 1 1	recP 12 2 6	xpt 13 15	8 2 1					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	ble _ tree Re gki 10 5 5 2 2 2 1	gex filter: gtr 6 4 3 2 2 3	murl 6 4 4 4 4 1	mutS 6 1 1 1 1 1 1	recP 12 2 6 8 12 1	xpt 13 15 2 7 12 1	8 2 1 2 7 4					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCO View:  ta	ble C tree Re gki 10 5 2 2 2 1 1 1	gex filter: gtr 6 4 3 2 2 2 3 1	murl 6 4 4 4 4 1 1	mutS 6 1 1 1 1 1 1 1 1	recP 12 2 6 8 12 12 1	xpt 13 15 2 7 12 1 1	8 2 1 2 7 4 3					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:      ta ST 1 2 3 4 5 6 7	ble _ tree Re gki 10 5 5 2 2 2 1	gex filter: gtr 6 4 3 2 2 3	murl 6 4 4 4 4 1	mutS 6 1 1 1 1 1 1	recP 12 2 6 8 12 1	xpt 13 15 2 7 12 1	8 2 1 2 7 4					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:      ta ST 1 2 3 4 5 6 7 8 9 10	ble C tree Re gki 10 5 2 2 1 1 1 1 1 1 1 0	gex filter: gtr 6 4 3 2 2 3 1 1 1 4	murl 6 4 4 4 1 1 1 1 1 7	mutS 6 1 1 1 1 1 1 1 1 1 7	recP 12 2 6 8 12 1 1 1 12	xpt 13 15 2 7 12 1 1 1 1 13	8 2 1 2 7 4 3 3 4 2 8					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:	ble C tree Re gki 10 5 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	gex filter: gtr 6 4 3 2 2 3 1 1 1 4 3 3	murl 6 4 4 4 1 1 1 1 7 4	mutS 6 1 1 1 1 1 1 1 1 1 7 7	recP 12 2 6 8 12 1 1 1 1 1 1 2 2	xpt 13 15 2 7 12 1 1 1 1 1 3 7	8 2 1 2 7 4 3 4 2 8 5					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:      ta ST 1 2 3 4 5 6 7 8 9 10	ble C tree Re gki 10 5 5 2 2 1 1 1 1 1 1 1 1 4	gex filter: gtr 6 4 3 2 2 3 1 1 1 4 4 3 4	murl 6 4 4 4 1 1 1 1 7 4 5	mutS 6 1 1 1 1 1 1 1 1 1 1 7 7 1 2	recP 12 2 6 8 12 1 1 1 12	xpt 13 15 2 7 12 1 1 1 1 13	8 2 1 2 7 4 3 4 2 2 8 5 5 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:  ta ST 1 2 3 4 5 6 7 8 9 10 11 12	ble C tree Re gki 10 5 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	gex filter: gtr 6 4 3 2 2 3 1 1 1 4 3 3	murl 6 4 4 4 1 1 1 1 7 4	mutS 6 1 1 1 1 1 1 1 1 1 7 7	recP 12 2 6 8 12 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	xpt 13 15 2 7 12 1 1 1 1 1 1 3 7 6	8 2 1 2 7 4 3 4 2 8 5					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC           View: @ ta           ST           1           2           3           4           5           6           7           8           9           10           12           13           14           15	ble $\bigcirc$ tree Re gki 10 5 2 2 1 1 1 1 1 1 1 1 4 10 3 3	gex filter: gtr 6 4 3 2 2 2 3 1 1 1 4 4 3 4 5 4 3 3 1 1 3 4 3 3 1 1 4 3 3 4 3 3 4 5 5 5 5 5 5 5 5 5 5 5 5 5	murl 6 4 4 1 1 1 1 1 7 7 5 6 6 7 2	mutS 6 1 1 1 1 1 1 1 1 1 2 6 6 2	recP 12 2 6 12 12 1 2 6 1 1 1 1 1 1 2 1 7 12 2 12 2	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 1 2 7 7 4 3 4 2 8 5 5 2 9 9 8 8 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View: @ ta 5 5 6 7 7 8 9 10 11 12 13 14 15 16	ble C tree Re gki 10 5 5 2 2 1 1 1 1 10 10 10 10 3 4	gtr         gtr           6         4           3         2           2         3           1         1           4         3           5         4           3         4	murl 6 4 4 4 1 1 1 1 1 1 5 6 6 7 7 2 2 1	mutS 6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 6 6 6 2 2 2	recP 12 2 6 8 12 1 1 1 1 2 1 1 1 2 2 12 12 12 12 12 9 9 17	xpt 13 15 2 7 12 1 1 1 1 1 1 1 3 7 6 13 13 8 1	8 2 1 1 7 4 3 4 4 2 2 8 5 5 2 9 8 8 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC           View: @ ta           ST           1           2           3           4           5           6           7           8           9           10           12           13           14           15	ble $\bigcirc$ tree Re gki 10 5 2 2 1 1 1 1 1 1 1 1 4 10 10 3	striker:           gtr           6           3           2           3           1           1           4           3           4           5           4           3           4           4           4           4           4	murl 6 4 4 1 1 1 1 1 7 7 5 6 6 7 2	mutS 6 1 1 1 1 1 1 1 1 1 2 6 6 2	19 x 12 2 6 8 12 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 1 2 7 7 4 3 4 2 8 5 5 2 9 9 8 8 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View: @ ta 5 7 8 9 10 11 12 13 14 15 16 17 18 19	ble () tree Re gki 5 5 2 2 2 1 1 1 1 1 1 1 4 4 4 4 4 4 4 3 3	2 c filter: 6 4 3 2 2 2 3 3 1 1 1 4 3 4 4 3 4 4 3 4 4 4 2 2 4 4 3 4 4 2 8 8	murl         6           4         4           4         1           1         1           7         4           5         6           7         2           1         1	mut5 6 1 1 1 1 1 1 1 1 1 2 6 6 2 2 2 2 1 1	19 x 12 2 6 8 12 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 7 6 1 3 8 1 1 3 7 7 7 7 7 7	8 2 1 2 7 4 4 2 2 8 5 5 5 5 2 9 9 8 8 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View: @ ta 5 7 8 9 10 11 12 13 14 15 16 17 18 19 20	ble () tree Re gki 10 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 filter: gtr 6 4 3 2 2 3 1 1 4 4 3 4 4 4 4 4 4 4 4 2 8 8 3	murl           6           4           4           4           1           1           1           7           6           7           2           1           1           1           1           1           1           2           1           4           2	muts         6           1         1           1         1           1         1           1         1           2         6           2         2           2         2           1         1           3         1           4         1           5         2           2         2           1         1           6         2           2         2           1         3	recP           12           2           6           8           1           1           1           1           12           13           14           15           16           17           17           17           8           8           9	13 15 2 7 7 12 1 1 1 1 1 3 7 6 1 3 13 13 8 8 1 1 3 6 7 7 7 6 6	8 2 2 7 4 4 3 3 4 4 2 2 8 8 5 2 9 9 8 8 2 2 2 2 2 2 2 2 6 6					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View: (a) ta 5 7 8 9 10 11 12 13 14 15 16 17 16 16 17 18 19 20 21	ble () tree Re gki 10 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 c filter: 6 gtr 6 gtr 2 2 2 3 1 1 1 1 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 8 8	murl           4           4           4           1           1           1           1           6           7           2           1           4	mut5           6           1           1           1           1           1           1           1           1           1           1           1           1           2           2           2           1           1           2           2           1           3           2	19 x 12 2 6 8 12 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 3 8 1 6 7 7 6 8	8 2 2 7 4 4 2 3 3 4 4 2 2 9 9 8 8 5 2 9 9 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View: (a) ta 5T 1 2 3 4 4 5 6 7 8 9 10 11 12 13 16 11 12 13 16 17 18 19 20 21 22	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 c filter: gtr 4 2 2 3 1 1 1 4 4 3 3 4 4 4 4 4 4 4 3 3 4 4 5 5 4 4 3 3 4 4 5 5 6 8 8 8 8 3 3 5 6 6 6 7 8 8 8 8 8 8 8 8 8 8 8 8 8	murl           6           4           4           4           1           1           1           7           6           7           2           1           1           1           1           1           1           2           1           4           2	mutS           6           1           1           1           1           1           1           2           6           2           2           2           1           1           1           1           1           1           2           6           2           2           1           8	recP           12           2           6           8           12           1           1           1           12           13           14           15           16           17           17           17           8           9           17           8           9           1           17 </td <td>13 15 2 7 12 1 1 1 1 1 1 3 7 6 13 13 6 7 7 6 7 7 6 8 11</td> <td>8 2 2 7 4 4 3 3 4 4 2 2 8 8 5 2 9 9 8 8 2 2 2 2 2 2 2 2 6 6</td> <td></td> <td></td> <td></td> <td></td> <td></td>	13 15 2 7 12 1 1 1 1 1 1 3 7 6 13 13 6 7 7 6 7 7 6 8 11	8 2 2 7 4 4 3 3 4 4 2 2 8 8 5 2 9 9 8 8 2 2 2 2 2 2 2 2 6 6					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 c filter: 6 gtr 6 gtr 2 2 2 3 1 1 1 1 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 8 8	murl           6           4           4           1           1           1           5           6           7           2           1           4           5           6           7           1           4           2           2           2           2	mut5           6           1           1           1           1           1           1           1           1           1           1           1           1           2           2           2           1           1           2           2           1           3           2	recP           12           2           6           8           12           1           1           1           12           13           14           15           16           17           17           17           8           9           17           8           9           1           17 </td <td>xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 3 8 1 6 7 7 6 8</td> <td>8 2 2 3 4 4 3 3 4 4 2 2 8 5 2 2 9 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td> <td></td> <td></td> <td></td> <td></td> <td></td>	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 3 8 1 6 7 7 6 8	8 2 2 3 4 4 3 3 4 4 2 2 8 5 2 2 9 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG 5 Isolate Data	GroupCC           View: ● ta           ST           1           2           3           4           5           6           7           8           10           11           12           13           14           15           16           17           18           19           20           21           22           23           24           25	ble () tree Re gki 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 cx filter: 6 gtr 6 gtr 2 2 2 3 1 1 1 4 4 4 4 5 4 4 4 4 4 4 2 8 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           4           4           4           1           1           1           1           1           6           7           6           2           1           2           2           2           2           2           2           1           1	mutS           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           6           6           2           2           1           2           8           2           8           8           8           5	recP           12           2           6           8           12           1           1           1           12           1           12           13           14           15           7	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 2 2 2 4 3 4 4 4 4 2 2 8 5 2 2 5 2 2 2 2 2 2 2 2 6 6 6 1 3 3					
Tit View Tools Window Help s 40 × GroupCG Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	gtr           6           3           2           2           3           1           1           4           3           4           4           4           3           4           4           8           8           3           3           2           2           2           2           3           3           3           3           3           2           2	murl           6           4           4           1           1           1           7           4           5           5           7           2           1           4           2           1           2           1           2           1           2           1           2           1           1           1           1	mutS           6           1           1           1           1           1           1           1           1           2           6           2           2           1           2           2           1           2           2           1           2           2           1           2           1           5           1	recP           12           2           6           8           12           1           1           1           12           1           1           12           17           12           9           17           8           9           15           7	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 1 1 4 4 3 4 4 2 2 8 5 2 2 2 2 2 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG 5 Isolate Data	GroupCC           View: ● ta           ST           1           2           3           4           5           6           7           8           10           11           12           13           14           15           16           17           18           19           20           21           22           23           24           25	ble () tree Re gki 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 cx filter: 6 gtr 6 gtr 2 2 2 3 1 1 1 4 4 4 4 5 4 4 4 4 4 4 2 8 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           4           4           4           1           1           1           1           1           1           6           7           6           2           1           2           2           2           2           2           1           1	mutS           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           6           6           2           2           1           2           8           2           8           8           8           5	recP 12 2 6 8 12 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 2 2 2 4 3 4 4 4 4 2 2 8 5 2 2 5 2 2 2 2 2 2 2 2 6 6 6 1 3 3					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:  table to the term of the term of term o	ble () tree Re gki 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 c filter: gtr d	murl           4           4           4           1           1           7           6           7           6           2           1           1           2           2           2           2           1           1           1           1           2           2           1           1           1           3	mutS           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           1           2           2           1           2           2           1           2           2           1           5           5           1	recP           12           2           6           8           12           1           1           1           12           1           1           12           17           12           9           17           8           9           15           7	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 2 2 4 4 4 4 2 2 4 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC           View: ● ta           ST           1           2           3           4           5           6           7           8           9           10           11           12           13           14           15           16           17           18           19           20           23           24           25           26           27           28           29           30	ble () tree Re gki 10 5 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 c filter: gtr d	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt 13 15 2 7 12 1 1 1 1 1 1 1 1 1 1 1 1 1	8 2 2 1 3 4 4 3 2 2 3 3 2 2 5 5 2 9 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 4 4 4 4 2 2 5 5 2 2 2 2 2 2 2 2 2 2 2 2 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           4           1           1           1           7           4           5           6           7           2           1           2           2           2           2           2           1           1           4           2           2           1           1           4           4	mutS           1           2           2           1           8           5           5           1           2           1           2           2	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2					
t View Tools Window Help s <b>4 x</b> GroupCG s Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 5 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2	- - - - - - - - - - - - - - - - - - -	quence Typing (S	election view)	Select	
Tit View Tools Window Help s 40 × GroupCG Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2	Multi-Locus Se	quence Typing (S	election view)		
rit View Tools Window Help s 40 × GroupCG Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2	Multi-Locus Se	quence Typing (S	:lection view)	Select	
rit View Tools Window Help s 40 × GroupCG	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2		quence Typing (S	election view)	Select	
Tit View Tools Window Help s 40 × GroupCG Isolate Data	GroupCC View:  table to the second se	ble () tree Re gki 10 5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1	2 Filter: 9 gtr 4 3 2 2 2 2 3 1 1 1 4 4 3 4 4 4 4 4 4 4 5 4 4 4 3 3 4 4 2 2 2 2 2 2 2 2 2 2 2 2 2	murl           6           4           4           1           1           1           7           4           5           6           7           4           1           1           1           1           2           2           2           2           4           1           1           4           2           2           4           4           4           4           4           4           4	mult           6           1           1           1           1           1           1           2           6           2           6           2           6           2           2           4           1           2           6           2           1           2           8           9           1           2           5           1           2           2           1           2           1           2           1           2           2           1           2      2           2           1	I2           12           2           6           8           12           1           1           1           12           1           12           1           12           1           12           13           16           7	xpt           13           15           2           7           12           1           1           1           1           1           1           13           6           13           6           7           6           7           6           7           6           13           13           14           10           12           14           100           7	8 2 2 2 2 4 3 4 4 5 2 2 7 7 7 4 4 8 8 8 7 2 9 8 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 2		quence Typing (S	election view)	Select	

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.

$\odot \odot \odot$	PHYLOVIZ			
File Edit View Tools Window Help				
Datasets 40 ×	GroupCG: Isolate Data 🗙			
	GroupCG: Isolate Data ×         View:       table <ul> <li>tree Regex filter:</li> <li>Group carbohydrate</li> <li>G</li> <li>C</li> <li>ST</li> <li>C collection</li> </ul>	Select	View	Reset
	GroupCC: Isolate Data (Selection view)	rgend Load	Save	<b>₹</b> x

# 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*.

<u>File</u> ⊻iew <u>T</u> ools <u>W</u> indow <u>H</u>	Help			
嵾 Load Dataset		Start Page 🗙		
惧 Load Dataset from MLST D	)Bs			
🔡 Open Proj <u>e</u> ct	Ctrl+Shift-O			
Open Recent Project				
<u>C</u> lose Project		Learn & Discover	My PHYLOViZ	What's New
C <u>l</u> ose Other Projects		Learn & Discover	My PHILOVIZ	What sinew
Close All Projects				
<u>S</u> ave		Take a Tour	Demos & Tut	Featured Plugin:
E <u>x</u> it				i outur ou r lugini
		Load a Dataset	Loading data	•
		What's New	Running the data ana	
		what shew	The PHYLOViZ display	
		Bug Tracking	Querying and visualiz	
			Project management	
			All Online Documen	
				PHYLOVIZ 2.0 is avail
			X	
			Show On Startup	45.j.

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

Steps	Database								
<ol> <li>Database</li> <li>Typing Data</li> </ol>	Dataset Name: spneumo								
3. Isolate Data 4. Sequence Data	Public DB Name: pubmlst.org - Streptococcus pneumoniae 💌 Update								
	pubmlst.org - Stenotrophomonas maltophilia         This plugin create         isolate data from         . MLST.net         pubmlst.org - Streptococcus canis         . MLST.net         pubmlst.org - Streptococcus dysgalactiae equis         . mlst.OCC.ie         . www.pasteur.fr/         . www.shigatox.net								
	Choose a dataset name and an available online database for a given microorganism to continue. An internet connection is necessary to communicate with the available webservices. This is a BETA version of the plugin. Please contact us if some bug is detected.								
	< <u>B</u> ack Next > <u>F</u> inish Cancel <u>H</u> elp								

The next step is to download the dataset.

Steps	Typing Data							
<ol> <li>Database</li> <li>Typing Data</li> <li>Isolate Data</li> <li>Sequence Data</li> </ol>	Dataset Name: Streptococcus pneumoniae Profile: aroE gdh gki recP spi xpt ddl Dataset Size: 10061 STs							
	Start/Stop       Done!         Loading from public databases the typing data of a specific dataset.         Each entry in a dataset is composed by a Sequence Identifier, followed by the different loci in the typing scheme.         Press the Start/Stop button to initiate the download.         If you press the Back button, you will have to restart the download.							
2357								
	< <u>Back</u> Next > <u>Finish</u> Cancel <u>H</u> elp							

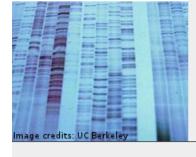
In the next window we can load ancillary data on isolates. In this example we choose to not load any data.

Steps	Isolate Data
<ol> <li>Database</li> <li>Typing Data</li> <li>Isolate Data</li> <li>Sequence Data</li> </ol>	File:       Browse         Key:       Update         If you have any type of ancillary data about the isolates (Demographic information, Epidemiological information, Antibiotic Resistance, Gene presence absence, etc) you can load a tab separated file to further represent it on the algorithm. This file should be a tab separated file with the first row with column headers (field names). One of the fields should be the Sequence identifier. Choose this field in the Key drop down menu to link the ancillary data to the allelic profile data. If the field does not appear in the drop down menu press the Update button. If you don't have any available data or don't wish do load any leave this fields blank an press Finish.
	< <u>Back</u> Next > <u>Finish</u> Cancel <u>H</u> elp

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

Steps	Sequer	nce Data								
1. Database 2. Typing Data 3. Isolate Data	-	<ul> <li>No sequence data</li> <li>Load sequence data from public databases</li> </ul>								
4. Sequence Data	aroE	Done!	or	Browse						
	gdh	Download	or	Browse						
	gki	Download	or	Browse						
	recP	Download	or	Browse						
	spi	Download	or	Browse						
	×pt	Download	or	Browse						
	ddl	Download	or	Browse						

Choose this option to download all sequence data available on the loci from the public database for the selected microorganism.



To import each locus independently, you can either:

. Press Download, to transfer the sequences from a public database;

. Press Browse, to load a Fasta file containing the sequences.

Finally press Finish to create the dataset and proceed with the analysis.

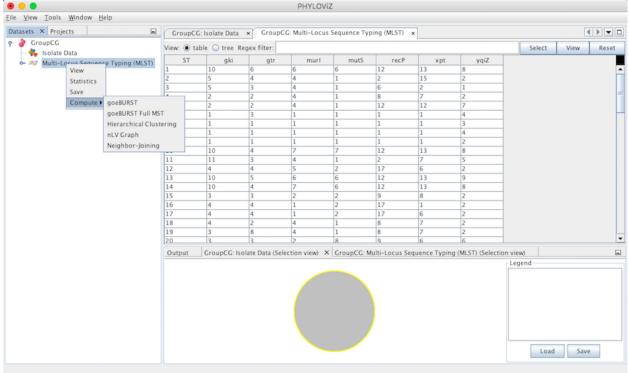
< <u>B</u> ack	Next >	<u>F</u> inish	Cancel	<u>H</u> elp

#### At the end we have seven typing data items to explore and analyze. <u>F</u>ile ⊻iew <u>T</u>ools <u>W</u>indow <u>H</u>elp

- 🍦 spneumo				us sequence			1			
🕶 🌌 Multi-Locus Sequence Typing (MLST)		ible 🔾 tree	Regex filter:					Select	View	Rese
🕶 🜌 aroElocus sequence	ID	s[0]	s[1]	s[2]	s[3]	s[4]	s[5]	s[6		s[7]
∽ 🖅 gdh locus seguence	aroE_1	G	A	A	G	С	G	A	G	
∽ 247 gkilocus sequence	aroE_2	G	A	A	G	С	G	A	G	
► 257 recP locus sequence	aroE_3	G	A	A	G	С	G	A	G	
	aroE_4	G	A	A	G	с	G	A	G	
🕶 🌌 spilocus sequence	aroE_5	G	A	A	G	С	G	A	G	
🖙 🎭 xpt locus sequence	aroE_6	G	A	A	G	С	G	A	G	
🖕 🌌 ddl locus sequence	aroE_7	G	A	A	G	С	G	A	G	
	aroE_8	G	A	A	G	С	G	A	G	
	aroE_9	G	A	A	G	С	G	A	G	
	aroE_10	G	A	A	G	С	G	A	G	
	aroE_11	G	A	A	G	С	G	A	G	
	aroE_12	G	A	A	G	С	G	A	G	
	aroE_13	G	A	A	G	С	G	A	G	
	aroE_14	G	A	A	G	С	G	A	G	
	aroE_15	G	A	A	G	С	G	A	G	
	aroE_16	G	A	A	G	С	G	A	G	
	aroE_17	G	A	A	G	С	G	A	G	
	aroE_18	G	A	A	G	С	G	A	G	
	aroE_19	G	A	A	G	С	G	A	G	
	aroE_20	G	A	A	G	С	G	A	G	
	aroE_21	G	A	A	G	С	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			A	G	C C	G		G	
	aroE_26	G	A		-	-	G	A	G	
	aroE_27	G	A	A	G	C C	G	A	G	
	aroE_28	G		A	G	C	G		G	
	aroE_29	G	A	A	G	c	G	A	G	
	aroE_30									
	aroE_31	G	A	A	G	C	G	A	G	
	aroE_32	G	A	A	G	С	G	A	G	

#### 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.

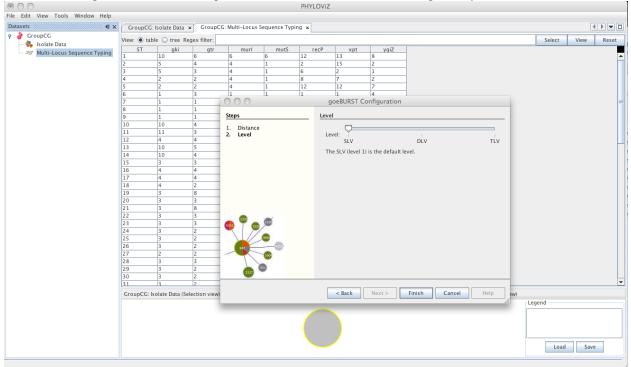


#### 3.1 goeBURST algorithm

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.

$\odot$ $\bigcirc$						PHYLC	OViZ											
File Edit View Tools Window Help																		
Datasets 40 ×	GroupC	5: Isolate Data	K Group(	G: Multi–Locus S	equence Typ	ing x										2		
👇 🧳 GroupCG	View:  tr	ble 🔾 tree Re	nev filter											s	elect	View	Reset	
— 鵫 Isolate Data	ST	gki	gex mer.	murl	mutS	ree	cD V	pt	yqiZ	1					elect	view	Kese	-
- 🌌 Multi-Locus Sequence Typing	1	10	6	6	6	12	13	.pt 8	yqız	-								
	2	5	4	4	1	2	15	2		1								
	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	11	1	1	1	4										
	7	1	1	000			goeBUR	ST Confi	guration				_					=
	8	1	1	Steps			Distance											
	10	10	4				Distance						_					
	10	11	3	1. Distance			Distance:	eBURST [	Distance				-					
	12	4	4	2. Level			This list contains only distances compatible with the current dataset.						_					
	13	10	5	-			This list co	ontains on	ly distances	s compatible	with the cu	irrent datas	et.					H
	14	10	4															
	15	3	3	1														
	16	4	4	1														
	17	4	4															
	18	4	2	_														
	19	3	8															
	20	3	3															
	21 22	3	8															
	22	3	3		110													
	23	3	2	🤓 🛛 💬	~													
	25	3	2		3968													
	26	3	2		410.													
	27	2	2		-													
	28	3	3		-													
	29	3	2		21													
	30	3	2	•		_							_					
	31	3	2															
	GroupCG:	Isolate Data (Se	lection view	>			< Back	< N	ext >	Finish	Cancel	Help	ew)					
				1									11.	Legend				
						-												
														1	Load	Sav	e	
															2544	Juv		
<u>.</u>																		-

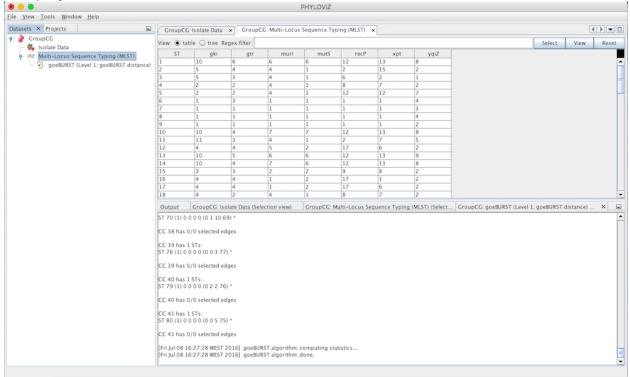
The second step is the choice of the level to which clonal complexes will be formed. The usual default for MLST analysis is SLV Level. Choosing DLV or TLV level will take longer calculation times, but could provide some insight to the relationships between clonal complexes formed at SLV and DLV level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

Group	CG: Isolate Dat	a × Group(	G: Multi-Locus	s Sequence Typ	oing (MLST)	×				< > -								
View:	View:      table      tree Regex filter:										Reset							
	ST gki gtr murl mutS recP xpt yqiZ																	
(MLST) 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											
10	11	3	4	1	2	7	5											
12	4	4	5	2	17	6	2											
12	4	5	6	6	17	13	9											
13	10	2 4	7	6	12	13	8											
14	3		2		9		2											
	4	3		2		8	2											
16		4	1		17	1	-											
17	4	4	1	2	17	6	2											
18	4	2	4	1	8	7	2											
Output	GroupCG:	Isolate Data (Sel	ection view)	GroupCG: M	ulti-Locus See	quence Typing	(MLST) (Selecti	GroupCG: goeBURST (Level 1	; goeBURST dis	tance) Out.	× [							
	0000(011)	) 69) *																
ST 70 (1)	ST 70 (1) 0 0 0 0 (0 1 10 69) *																	
		CC 38 has 0/0 selected edges																
	s 0/0 selected	edges																
CC 38 has		edges																
CC 38 has CC 39 has	s 1 STs:																	
CC 38 has CC 39 has																		
CC 38 has CC 39 has ST 76 (1)	s 1 STs:	77) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha:	s 1 STs: 0 0 0 0 0 (0 0 3 s 0/0 selected	77) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha:	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs:	77) * edges																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha:	s 1 STs: 0 0 0 0 0 (0 0 3 s 0/0 selected	77) * edges																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha: ST 79 (1)	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2	77) * edges 76) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha: ST 79 (1)	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs:	77) * edges 76) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha: ST 79 (1)	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2 s 0/0 selected	77) * edges 76) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha: ST 79 (1) CC 40 ha: CC 41 ha:	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2 s 0/0 selected	77) * edges 76) * edges																
CC 38 hat ST 76 (1) CC 39 hat ST 77 (1) CC 40 hat ST 79 (1) CC 40 hat ST 80 (1)	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2 s 0/0 selected s 1 STs: 0 0 0 0 (0 0 5	77) * edges 76) * edges 75) *																
CC 38 hat ST 76 (1) CC 39 hat ST 77 (1) CC 40 hat ST 79 (1) CC 40 hat ST 80 (1)	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2 s 0/0 selected s 1 STs:	77) * edges 76) * edges 75) *																
CC 38 ha: CC 39 ha: ST 76 (1) CC 39 ha: CC 40 ha: ST 79 (1) CC 41 ha: ST 80 (1) CC 41 ha: Fri Jul 08	s 1 STs: 0 0 0 0 (0 0 3 s 0/0 selected s 1 STs: 0 0 0 0 (0 2 2 s 0/0 selected s 1 STs: 0 0 0 0 (0 0 5 s 0/0 selected 16:27:28 WES	77) * edges 76) * edges 75) *	RST algorithm:	computing sta	tistics													

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.

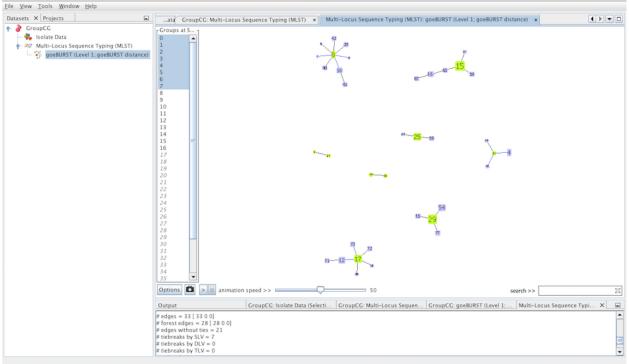


Double clicking on the goeBURST item that is now on the Dataset tree menu will show the display. The clonal complexes will be arbitrarily numbered starting from 0 (for the CC with most STs) and contains all the data relevant to the goeBURST analysis (STs in each group and the drawn SLVs edges). The following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

File Edit Vew Tools Window Help   Datasets   CroupCC   CroupCC Control   CroupCC Induct Contr <tr< th=""><th>000</th><th>PHYLOVIZ</th><th></th></tr<>	000	PHYLOVIZ	
• OroupCC       • OroupC C         • • State Data       • • • • • • • • • • • • • • • • • • •	File Edit View Tools Window Help		
P CoupCC For CoupC C For CoupC A </td <td>Datasets 🛛 🕯 🗙</td> <td>GroupCG: Isolate Data x GroupCG: Multi-Locus Sequence Typing x GroupCG: goeBURST (Level 1) x</td> <td></td>	Datasets 🛛 🕯 🗙	GroupCG: Isolate Data x GroupCG: Multi-Locus Sequence Typing x GroupCG: goeBURST (Level 1) x	
GroupCG: Isolate Data (Selection view)       GroupCG: Multi-Locus Sequence Typing (Selection view)       GroupCG: goeBURST Output	♀-	Croups at SLV	
[Fri Mar 25 11:22:20 WET 2011] goeBURST started       Image: Computing groups         [Fri Mar 25 11:22:20 WET 2011] goeBURST algorithm: integrating data       Image: Computing groups         #CC = 42       CC 0 has 9 STs:         ST 38 (2) 44 00 (4 4 0 72)       ST 38 (2) 44 00 (4 4 0 72)         ST 9 (0) 35 00 (3 5 2 70)       ST 8 (6) 71 00 (7 1 0 72) *         ST 8 (6) 71 00 (7 1 0 72) *       ST 39 (1) F61 01 (61 172)         ST 40 (1) 17 70 01 (7 0 72)       ST 40 (1) 17 00 (17 0 72)		Options >    animation speed >> 50 search >>	×
goeBURST algorithm: computing groups       Image: Cross and Cro		GroupCG: Isolate Data (Selection view) GroupCG: Multi-Locus Sequence Typing (Selection view) GroupCG: goeBURST Output	× ₹ (888888
		goeBURST algorithm: computing groups [Fri Mar 25 11:22:20 WET 2011] goeBURST algorithm: integrating data #CC = 42 CC 0 has 9 STs: ST 38 (2) 44 00 (44 072) ST 98 (0) 35 00 (35 171) ST 78 (0) 35 00 (35 171) ST 78 (0) 51 00 (35 2 70) ST 88 (6) 71 00 (71 072) ST 84 (0) 14 30 (14 372) ST 40 (1) 17 00 (17 072)	

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

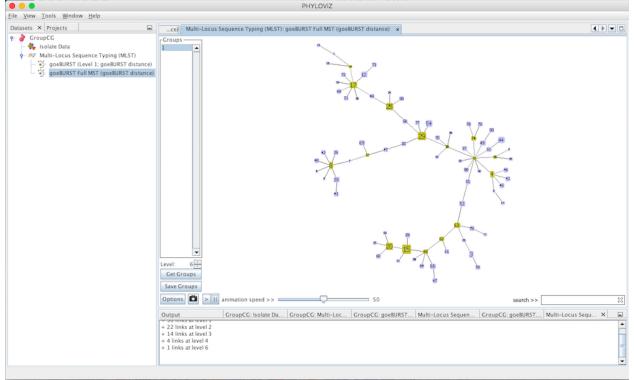
Multiple groups can be displayed simultaneously by selecting them, using the CTRL /CMD and/or SHIFT keys.



### 3.2 goeBURST Full MST algorithm

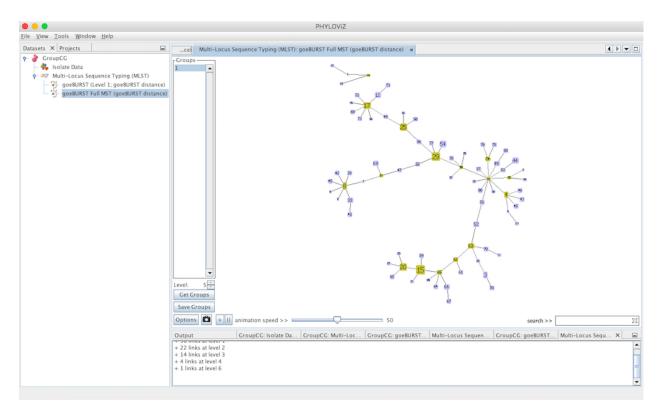
Using an extension of the goeBURST rules up to (n)LV level (where (n) equals to the number of loci your dataset uses), a Minimum Spanning Tree-like structure can be drawn. This is typicially used for SNP or cg/wgMLST datasets with dozens to thousand of loci.

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.

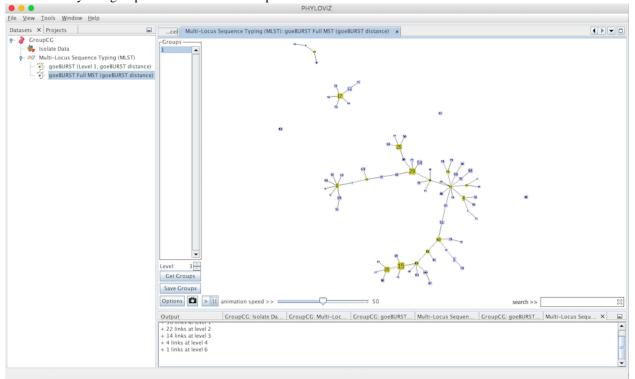


New options appear on the display: The *Level* selector and two new buttons *Get Groups* and *Save Groups*. The *Level* represents the *Locus Variant* level and allows the removal of all the links greater than the number represented. The user can use the up and down arrows or directly edit the number 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.

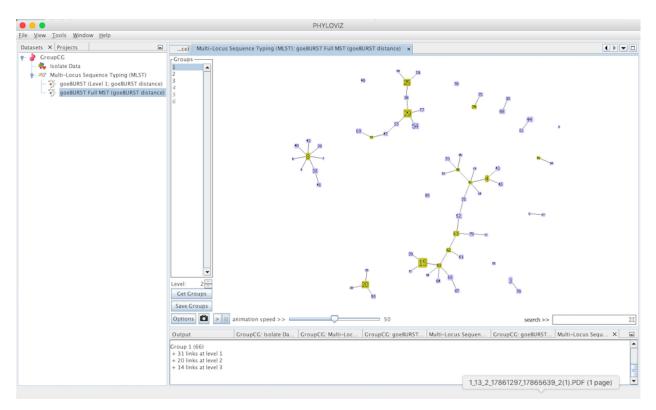
Decreasing the *Level* selector, allows the user to see how clonal complexes would relate to each other at a certain level. Level 1, 2 and 3 are equivalent to calculating goeBURST at those levels (SLV,DLV and TLV respectively). The following images shows what happens to the dataset when you decrease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



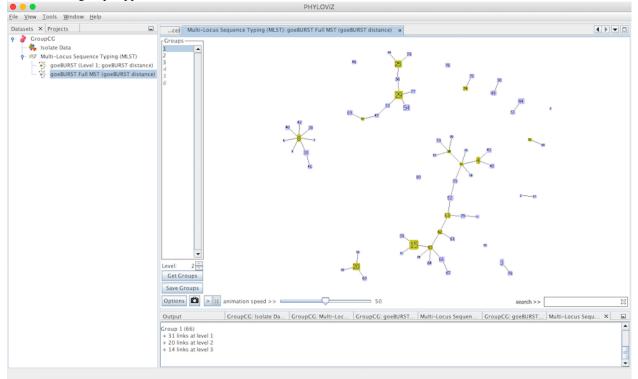
At level 5 only two groups are formed in the sample dataset.



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

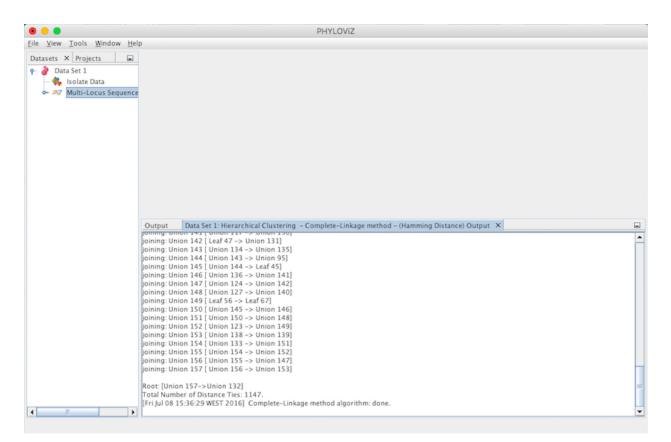
Selecting the Hierarchical Clustering opens the dialog where you can select what method you want to apply. The first step is choosing the *Distance* to be used. Currently the hamming distance is the only one available, but others could be implemented.

		PHYLOVIZ
<u>File View T</u> ools <u>W</u> indow	Help	
Datasets × Projects	-	
💡 🍦 Data Set 1		
— 🐥 Isolate Data	• • •	Hierarchical Clustering Configuration
⊶ 🛷 Multi-Locus Sequ		
	Steps	Distance
	1. Distance 2. Method	Distance: Hamming Distance
	z. Method	This list contains only distances compatible with the current dataset.
		The let containe only ansances comparise min the carrier addeed
		In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In another way, it measures the minimum number of substitutions required to change one string into the other, or the minimum number of errors that could have transformed one string into the other. This distance is used as a measure of genetic distance. The Hamming distance is named after <b>Richard W. Hamming</b>
	2	
1	09 08 07 06 05 04 03 02	
1		< Back Next > Finish Cancel Help
•	•	

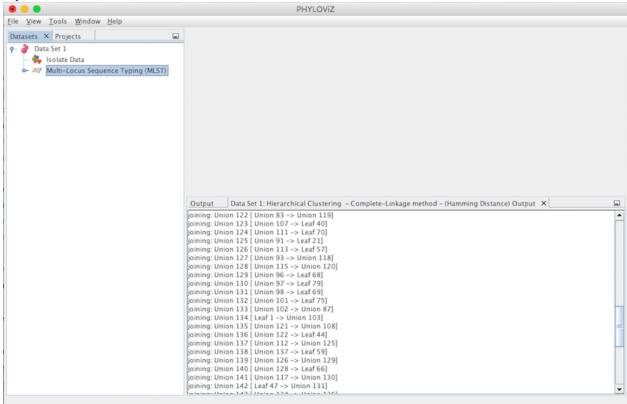
The second step is to select the *Method*. You can choose between complete-linkage, single-linkage, UPGMA (Unweighted Pair Group Method with Arithmetic mean) and WPGMA (Weighted Pair Group Method with Arithmetic mean). Selecting the method corresponds to selecting the criterion of minimal dissimilarity.

		PHYLOVIZ	
<u>File View Tools W</u> indow	<u>H</u> elp		
Datasets × Projects			
♀- ∂ Data Set 1 - 4 Isolate Data			
🗠 🛷 Multi-Locus Sequ	• • •	Hierarchical Clustering Configuration	
	Steps	Method	
	1. Distance 2. Method	Method: Complete-Linkage method	
		This list c Complete-Linkage method Single-Linkage method	
		UPGMA method CL (CompWPGMA method	
		clustering method. It is also known as farthest neighbour clustering	
		CL is used for the creation of <b>phenetic trees</b> (phenograms) and in a phylogenetic context, it assumes a constant rate of evolution (molecular clock hypothesis), and is not a well-regarded method for inferring relationships unless this assumption has been tested and justified for the data set being used. This algorithm constructs a <b>rooted tree</b> (dendrogram) that reflects the structure present in a pairwise <b>dissimilarity matrix</b> . At each step, the nearest two clusters are combined into a higher-level cluster. The <b>distance</b> between any two clusters <b>A</b> and <b>B</b> is the <b>farthest</b> distance between elements of each cluster. The method is attributed to <b>D. Defays</b> .	
	03 08 07 06 05 0A 03 02		
		< Back Next > Finish Cancel Help	
	•		

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.



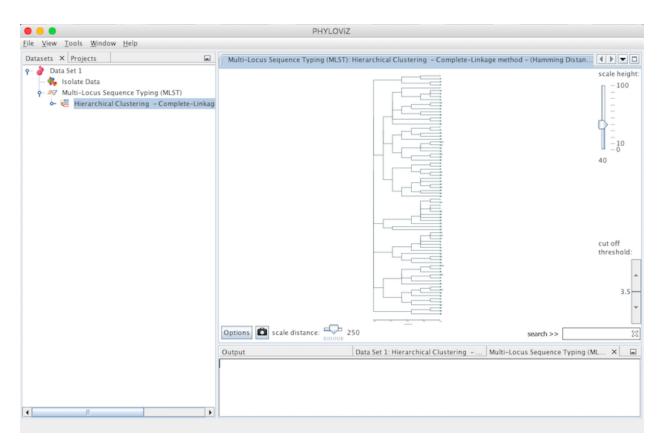
In order to display the dendogram view, it is necessary to expand the typing data on the Datasets' tab, if it is not already expanded.



) 😑 🕒	PHYLOVIZ
le ⊻iew <u>T</u> ools <u>W</u> indow <u>H</u> elp	
atasets × Projects	
ata Set 1	
- 🐥 Isolate Data	
- AV Multi-Locus Sequence Typing (MLST)	
- Hierarchical Clustering - Comple	
<ul> <li>Complete Contracting - Complete</li> </ul>	at sinnag
	Output Data Set 1: Hierarchical Clustering – Complete–Linkage method – (Hamming Distance) Output X
	joining: Union 122 [ Union 83 -> Union 119] joining: Union 123 [ Union 107 -> Leaf 40]
	joining: Union 124 [ Union 11] -> Leaf 70]
	joining: Union 125 [ Union 91 -> Leaf 21]
	joining: Union 126 [ Union 113 -> Leaf 57]
	joining: Union 127 [ Union 93 -> Union 118]
	joining: Union 128 [ Union 115 -> Union 120]
	joining: Union 129 [ Union 96 -> Leaf 68]
	joining: Union 130 [ Union 97 -> Leaf 79]
	joining: Union 131 [ Union 98 -> Leaf 69]
	inizing Union 122 (Union 101 ) Losf 75]
	joining: Union 132 [ Union 101 -> Leaf 75]
	joining: Union 133 [ Union 102 -> Union 87]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44] joining: Union 137 [ Union 112 -> Union 125]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44] joining: Union 137 [ Union 112 -> Union 125] joining: Union 138 [ Union 137 -> Leaf 59]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44] joining: Union 137 [ Union 112 -> Union 125] joining: Union 138 [ Union 117 -> Leaf 59] joining: Union 139 [ Union 126 -> Union 129]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 134 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44] joining: Union 137 [ Union 112 -> Union 125] joining: Union 138 [ Union 137 -> Leaf 59]
	joining: Union 133 [ Union 102 -> Union 87] joining: Union 133 [ Leaf 1 -> Union 103] joining: Union 135 [ Union 121 -> Union 108] joining: Union 136 [ Union 122 -> Leaf 44] joining: Union 137 [ Union 112 -> Union 125] joining: Union 138 [ Union 137 -> Leaf 59] joining: Union 138 [ Union 126 -> Leaf 66]

#### 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.



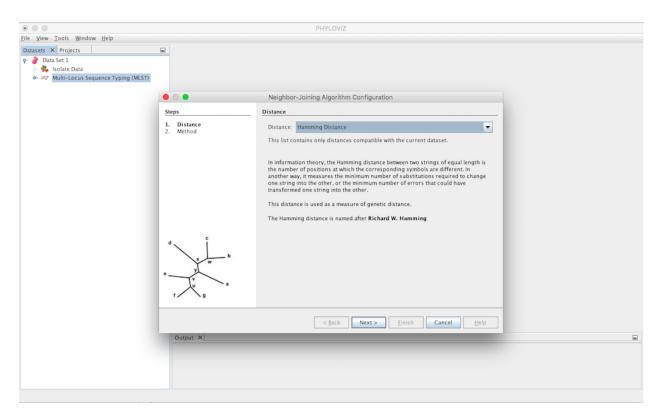
Some features were added to the visualization to improve and facilitate the analysis. These features are the following:

- 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.



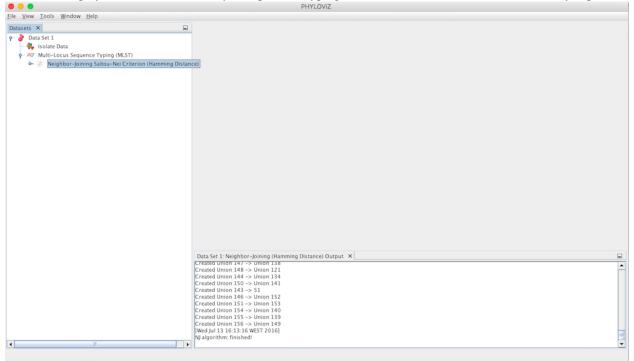
The second step is to select the *Criteria* of the tree branch-length minimization. You can choose between Saitou-Nei and Studier-Keppler criterion.

		PHYLOVIZ
Eile View Tools Window Help		
Datasets × Projects		
P ata Set 1     Hata Set 1     Hata Set 1     Solate Data     Multi-Locus Sequence Typing (MLST)		
	• • •	Neighbor-Joining Algorithm Configuration
	Steps	Method
	1. Distance 2. Method	Criteria: Saitou-Nei Criterion This list contains only distances compatible with the current dataset. The Neighbor-Joining algorithm This method is based on the minimum evolution principle and provides trees with near-minimal sum of branch-length estimates proposed by Saitou and Nei
	Output X	Sack Next > Finish Cancel Help

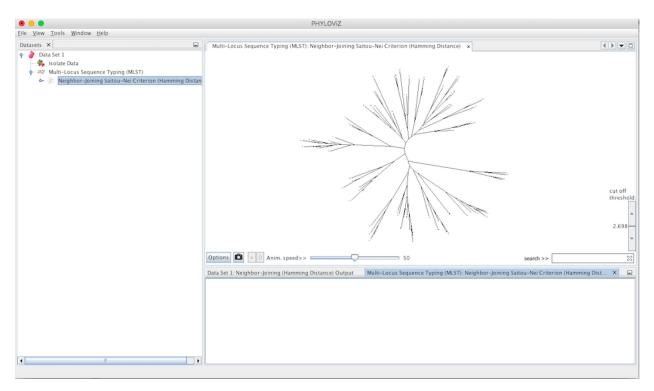
A Neighbor Joinning *Output Tab* will appear and display the results of the application of the chosen method. The information displayed represents the same as the Hierarchical Clustering *Output Tab*.

	PHYLOVIZ	
<u>File View Tools Window H</u> elp		
Pie give jours minute jep       Datasets X       Projects       P       Data Set 1       -		
	L'eated Union 144 -> Union 134 C'reated Union 150 -> Union 141 C'reated Union 135 -> S1 C'reated Union 154 -> Union 152 C'reated Union 154 -> Union 150 C'reated Union 154 -> Union 140 C'reated Union 155 -> Union 149	

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

#### **Display and visualization**

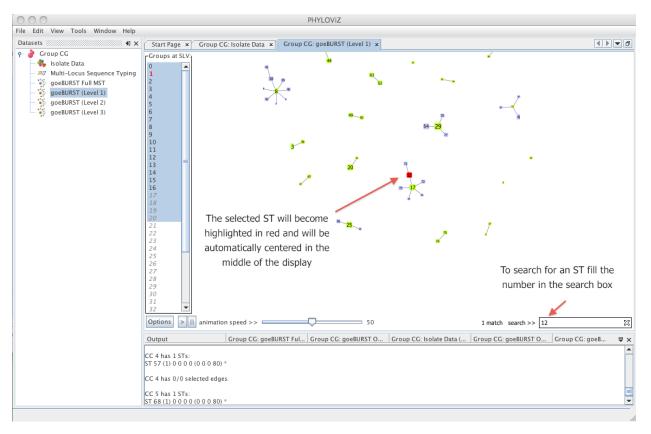
#### 4.1 Interface features

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.

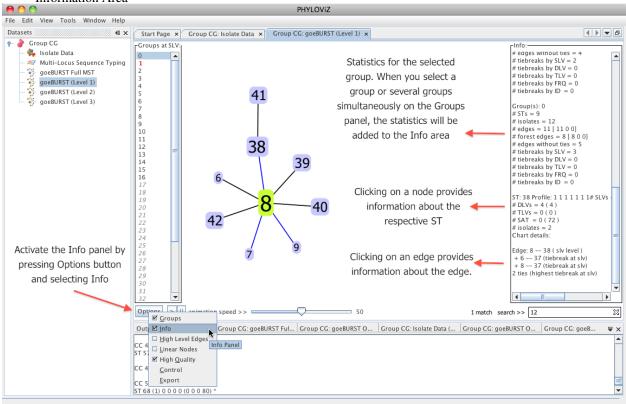
The Display tab offers the user the ability to search for an isolate, Highlight the SLVs and DLVs, control the animation speed, select different different or multiple groups. You can fit any displayed graphs to the window by right-clicking any open space (i.e. with no link or ST node) on the window.

#### 4.1.1 Common features

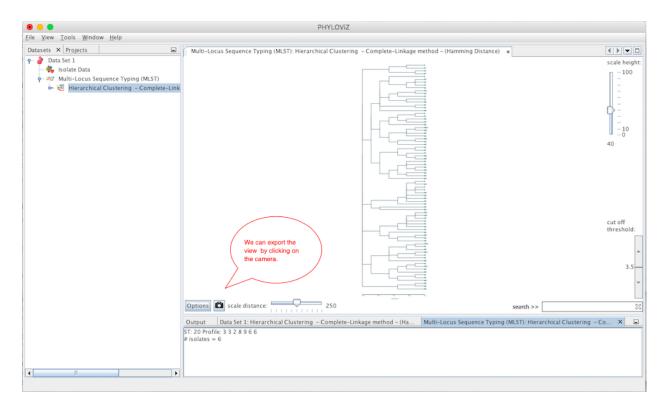
• ST search



Information Area



• Exporting Image



· Re-scale nodes ••• PHYLOVIZ <u>File View Tools Window H</u>elp Datasets × Projects Start Page x Multi-Locus Sequence Typing (MLST): Hierarchical Clustering - Complete-Linkage method - (Hamming Distance) x 4 🕨 🗖 P Data Set 1

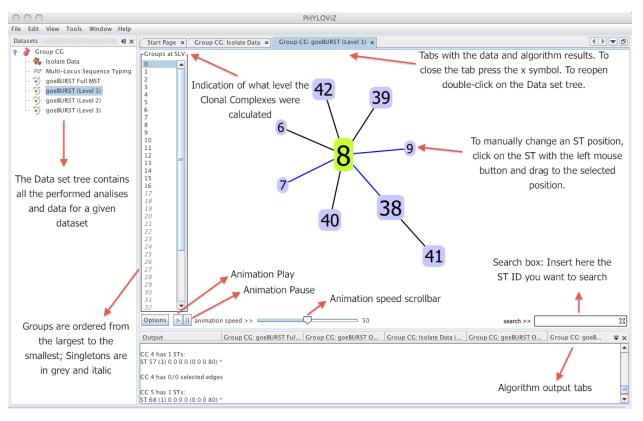
 → Data Set 1

 → Multi-Locus Sequence Typing (MLST)

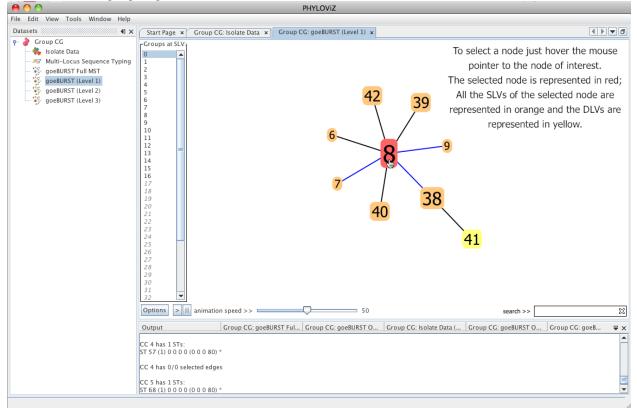
 → ₩ Hierarchical Clustering - Complete-Link
 scale height 100 -10 51 cut off threshold The Linear Nodes option changes the size of ST nodes. By default, the size of each ST node in the display is logarithmic in the number of the isolates that have that ST. By changing to linear, the size in the display will become lin proportional. 0 Options 🛱 scale distance: 250 search >> Info Data Set 1: Hierarchical Clustering – Complete–Linka... Multi–Locus Sequence Typing (MLST): Hierarchical ... 🗴 🗔 Outr Distance labels Linear Nodes Re-scale Edges(log distance) 4

#### 4.1.2 GoeBURST and GoeBURST Full MST features

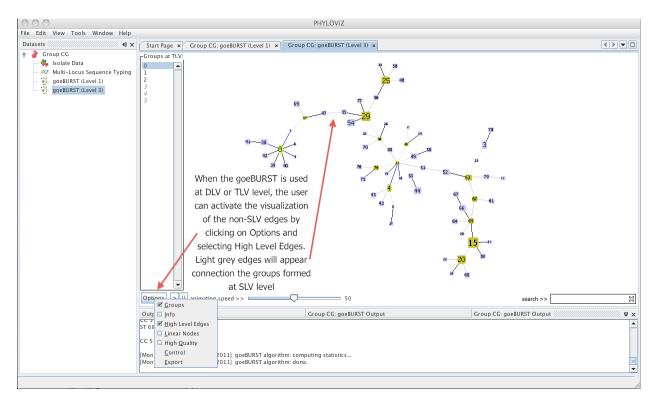
• Basic Interface



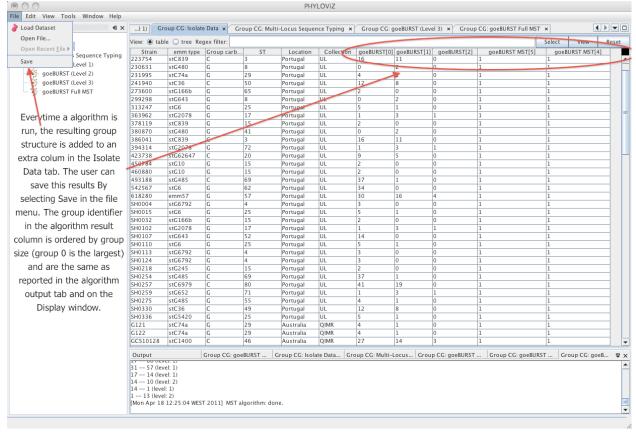
• SLV/DLV highlighting



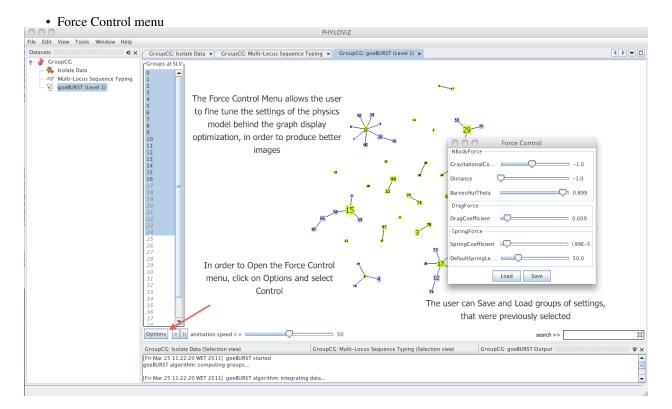
• High Level Edges



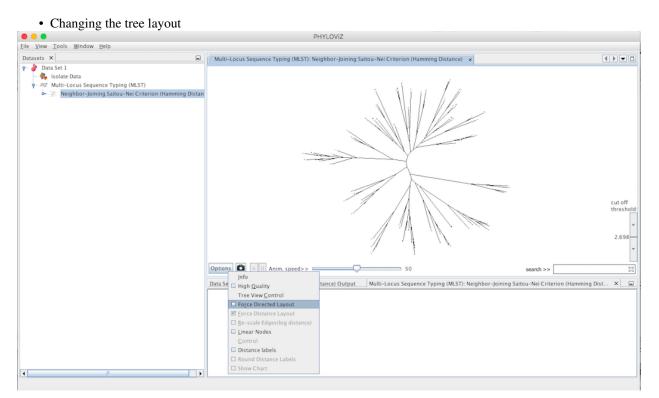
#### • Saving Results



### 4.1.3 GoeBURS, GoeBURST Full MST and Neighbor Joinning features

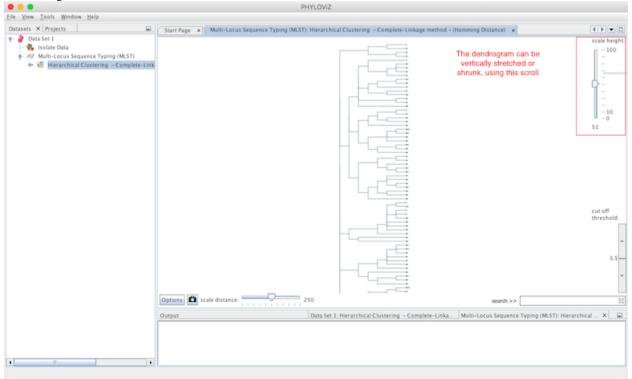


#### 4.1.4 Neighbor Joinning features

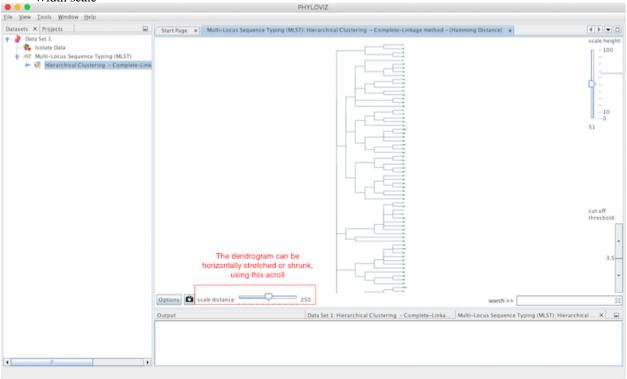


### 4.1.5 Hierarchical Clustering and Neighbor Joinning features

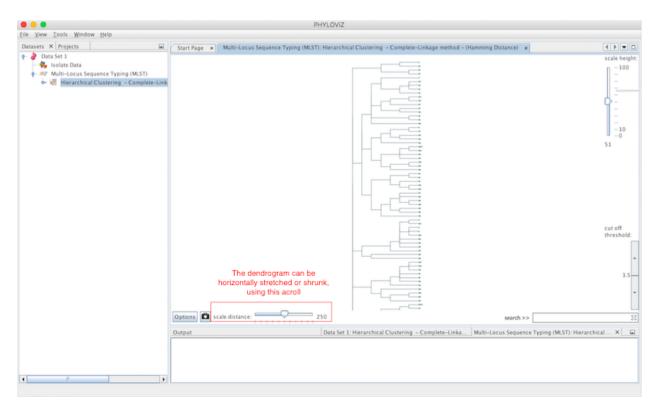
#### • Height scale



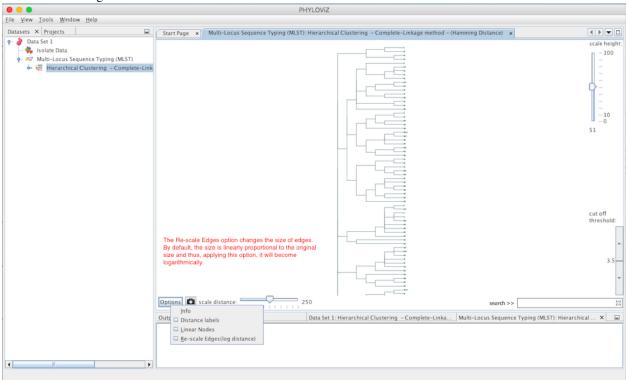
#### • Width scale



• 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

# 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

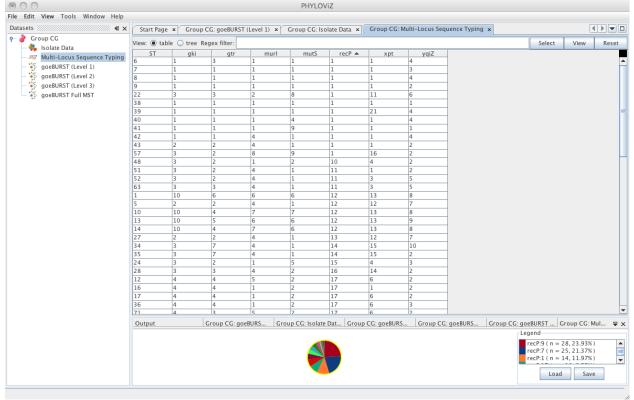
PHYLOViZ File Edit View Tools Window Help Datasets Start Page × Group CG: goeBURST (Level 1) × Group CG: Isolate Data × **4**0 × ♀- → Group CG → Solate Data /iew: 🖲 table 🔾 tree Regex filter: The Regex filter allows the creation of complex queries View Rese Select tC839 C 
 Location
 Collection
 goeBURST[0]
 goeBURST[1]
 goeBURST[2]

 Portugal
 UL
 16
 11
 0
 Strai. 223754 goeBURS Multicocus Sequence Typing goeBURST (Level 1) goeBURST (Level 2) 23063 stG480 Portugal tC74a Portugal Left clicking on a column header will sort the column in goeBURST (Level 3) stC36 Portuga goeBURST ull MST stG166b stG643 Portugal ascending or descending alphabetical order. 1 65 273 Portugal 24 ortugal st( ortugal Right clicking on a table header will select that field for Column headers Double click on the ortuga limiting the regex queries to it. By pressing the Select ortugal UL Isolate Data to active ortugal stCoss stG2078 buton will also select all entries in the selected column. the Isolate Data tab. 94314 Portugal stG62647 stG10 3738 Portugal After the selection is perfomed, the user can press The user can choose to Portugal View to plot the selection onto the resulting algorithm 60880 stG10 Portugal visualize the dataset as stG485 stG6 493188 Portugal 62 graphs. Reset will clear all the selections made. a table or a tree. 42567 Portugal emm57 stG6792 Portugal 618280 10004 Portugal SH0015 Portugal stG166b stG2078 Portugal SH0032 H0102 Portuga SH0107 stG643 Portugal 14 SH0110 stG6 Portuga stG6792 SH0113 Portuga stG6792 stG245 Portugal SH0124 H0218 Portuga SH0254 stG485 69 Portuga SH0257 stC6979 stG652 Portugal 80 71 41 SH0259 Portuga Output 🛛 🗶 Group CG: goeBURST Ou... 🛛 Group CG: Isolate Data (S... 🖉 Group CG: goeBURST Ou... 🖉 Group CG: goeBURST Ou... Group CG: goeBURST Full ...

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.



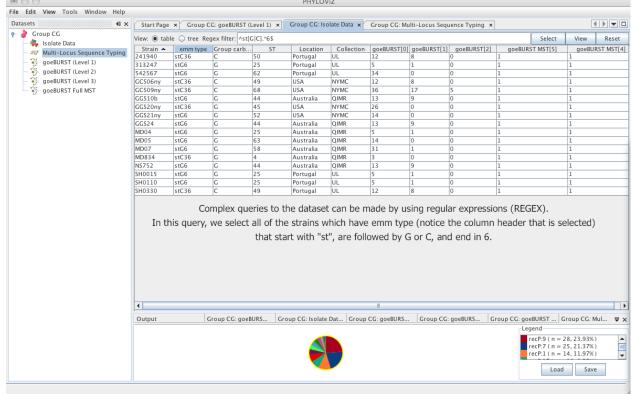
## 5.3 Regular expression primer

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, ATCG 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

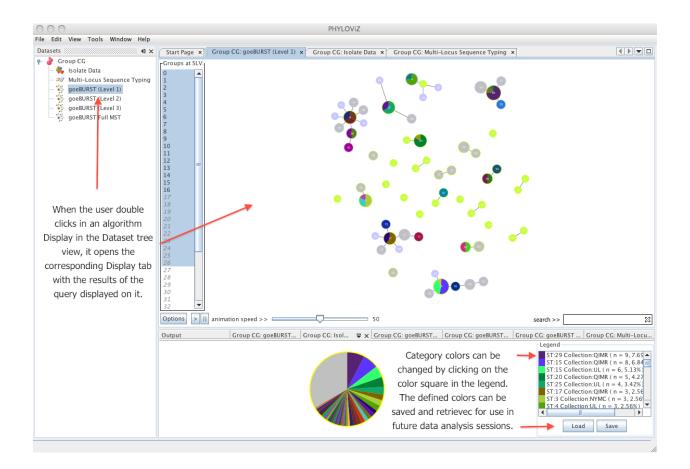
· Table view with selections

$\odot \odot \odot$					PHYLO	ViZ						
ile Edit View Tools Window Help												
Datasets 🛛 🔍 🗙	Start Page	× Group	CG: goeBURST (I	Level 1) ×	Group CG: Iso	olate Data 🗙	Group CG: Mu	lti-Locus Sequence Typing	×			
👇 🥜 Group CG	Views 🛞 tabl		egex filter: ^UL	2[0_0]/1)	_					Select	View	Reset
— 🎭 Isolate Data			-			C 11	BUDGT[0]		1 000			
- 257 Multi-Locus Sequence Typing	Strain GGS15ny	stC74a	Group carb G	. ST 29	Location Other	NYMC	goeBURST[0] 4	goeBURST[1] goeBURST[2 1 0	] goeвuk:	ST MST[5]	доевик: 1	ST MST[4]
<ul> <li>goeBURST (Level 1)</li> </ul>	GGS16ny	stG485	G	29	Other	NYMC	4	1 0	1		1	
goeBURST (Level 2)	C121	stC74a	G	29	Australia	QIMR	4	1 0	1		1	
— 🔅 goeBURST (Level 3)	G122	stC74a	G	29	Australia	QIMR	4	1 0	1		1	
goeBURST Full MST	GCS2816	stG62647	C	20	Australia	QIMR	9	5 0	1		1	
	GCS6894	stG62647	C	20	Australia	QIMR	9	5 0	1		1	
	GCS6929	stG62647	С	20	Australia	QIMR	9	5 0	1		1	
	GGS11172	stC74a	G	29	Australia	QIMR	4	1 0	1		1	
	GGS539813 GGS540048	stC74a stG485	G	29 29	Australia Australia	QIMR	4	1 0	1		1	
	GGS9225	stG485 stC74a	G	29	Australia	OIMR	4	1 0	1			
that start with UL (^UL)	MD04	stG6	G	25	Australia	OIMR	5	1 0	1		1	
· · /	MD06	stC74a	G	29	Australia	OIMR	4	1 0	1			
or ( ) have any number	MD122	stC74a	G	29	Australia	QIMR	4	1 0	1		1	
with 2 digits starting with	MD227	stC6979	С	20	Australia	QIMR	9	5 0	1		1	
	MD605	stG62647	C	20	Australia	QIMR	9	5 0	1		1	
2 (2[0-9]{1}). Since only	MD722	stC74a	G	29	Australia	QIMR	4	1 0	1			
	MD934	stG5420	G	25	Australia	QIMR	5	1 0	1			
	168554	stG485	G	47	Portugal	UL	11	1 0	1		1	
	171712 220269	stG480 stG2078	G	38 15	Portugal	UL	0	2 0	1		1	
	220269	stG2078 stC839	c	3	Portugal Portugal	UL	16	11 0	1		1	
	230631	stG480	G	8	Portugal	UL	0	2 0	1		1	
	231995	stC74a	G	29	Portugal	UL	4	1 0	1		1	
	241940	stC36	C	50	Portugal	UL	12	8 0	1		1	
	273600	stG166b	G	65	Portugal	UL	2	0 0	1		1	
After writing the guery	299298	stG643	G	8	Portugal	UL	0	2 0	1		1	
5 1 1	4											
the user presses the	Output		Group CG: goel	BURST	Group CG: Isol	₩ Group	cG: goeBURST	Group CG: goeBURST	Group CG: go	eBURST G	roup CG: N	lulti-Locu
Select button to select all									Lege	nd		
vegulte (the blue									ST	:29 Collection	QIMR ( n =	9,7.69
results (the blue									S1	:15 Collection	:QIMR (n =	= 8, 6.84
background appears) and										:15 Collection :20 Collection		
5 11 ,										:20 Collection		
to view the results on the										:17 Collection		
Display tabs , the user										:3 Collection:		
							/			:4 Collection: 17 Collection		
has to press the View										:17 Collection		
button									4			•
54000										Load	C	
										Load	Save	

• Query results Chart

					PHYLO	ViZ							
Edit View Tools Window Help													
sets 🛛 🗤 🕯	Start Page	× Group C	G: goeBURST (I	Level 1) ×	Group CG: Iso	olate Data 🗙	Group CG: Mu	lti-Locus Seq	uence Typing 🗙				
Group CG	View: 🖲 tab	le 🔾 tree Reg	gex filter: ^UL	2[0-9]{1}							Select	View	Res
- 🛼 Isolate Data	Strain	emm type	Group carb	. ST	Location	Collection A	goeBURST[0]	goeBURST[1]	goeBURST[2]	goeBURST	F MST[5]	goeBUR	ST MST
Multi-Locus Sequence Typing goeBURST (Level 1)	GGS15ny	stC74a		29	Other	NYMC	4	1	0	1		1	
goeBURST (Level 1)	GGS16ny	stG485	G	29	Other	NYMC	4	1	0	1		1	
goeBURST (Level 1) goeBURST (Level 2) goeBURST (Level 3)	G121 G122	stC74a stC74a	G	29 29	Australia Australia	QIMR	4	1	0	1		1	
goeBURST Full MST	GCS2816	stG62647	C	29	Australia	QIMR	9	5	0	1		1	
Sterorest Full Wat	GC56894	stG62647	C	20	Australia	QIMR	9	5	0	1		1	
	GCS6929	stG62647	С	20	Australia	QIMR	9	5	0	1		1	
	GGS11172	stC74a		29	Australia	QIMR	4	1	0	1		1	
	4	stC74a	lc.	29	Australia	OIMR	4	1	0	1		1	
	Output	c	iroup CG: goel	BURST Gr	oup CG: Isol	₹ × Group	CG: goeBURST	Group CO	: goeBURST			roup CG: I	Multi-L
II the combinations of the	values a	re			_					Legen			
				_							29 Collection 15 Collection		
displayed in the chart wit	h differer	nt									15 Collection		
colors automatically as	signed									ST:	20 Collection	n:QIMR ( n	= 5, 4.
colors automatically as	signeu.									ST:	25 Collection		
											17 Collection 3 Collection:		
The arev color ('others' ir	n the cha	rt								ST: ST:	3 Collection: 4 Collection:	NYMC ( n UL ( n = 3	= 3, 2.5
5, (		rt								ST: ST: ST:	3 Collection: 4 Collection: 17 Collection	NYMC (n UL (n = 3 h:UL (n =	= 3,2.5 ,2.56% 2,1.719
The grey color ('others' ir legend) is reserved for		rt								ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection	NYMC (n UL (n = 3 n:UL (n = n:QIMR (n	= 3, 2.5 , 2.56% 2, 1.719 = 2, 1.5
legend) is reserved for	all the									ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 3 Collection:	NYMC (n UL (n = 3 h:UL (n = n:QIMR (n h:NYMC (n UL (n = 2	= 3, 2.5 , 2.56% 2, 1.719 = 2, 1. n = 2, 1 , 1.71%
legend) is reserved for combinations not found by	all the									ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 3 Collection: 69 Collection	NYMC (n UL (n = 3 h:UL (n = n:QIMR (n h:NYMC (n UL (n = 2 h:UL (n =	= 3, 2.5 , 2.56% 2, 1.71 = 2, 1 n = 2, 1 , 1.71% 2, 1.71
legend) is reserved for	all the									ST: ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 3 Collection: 69 Collection 8 Collection:	NYMC ( n = 3 ::UL ( n = 3 ::UL ( n = ::QIMR ( n ::NYMC ( n UL ( n = 2 ::UL ( n = NYMC ( n	= 3, 2.5 , 2.56% 2, 1.71 = 2, 1. n = 2, 1 , 1.71% 2, 1.71 = 2, 1.7
legend) is reserved for ombinations not found by	all the									ST: ST: ST: ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 3 Collection: 69 Collection: 8 Collection: 8 Collection:	NYMC ( n = 3 uL ( n = 3 uUL ( n = 3 uQIMR ( n = 1 uUL ( n = 2 uUL ( n = 2 NYMC ( n = 2 QIMR ( n = 2 UL ( n = 2	= 3, 2.5 , 2.56% 2, 1.71 = 2, 1. 1 = 2, 1 , 1.71% 2, 1.71 = 2, 1.7 = 2, 1.7 , 1.71%
legend) is reserved for ombinations not found by selection.	all the the activ	ve								ST: ST: ST: ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 3 Collection: 69 Collection: 8 Collection: 8 Collection: 15 Collection	NYMC ( n = 3 11:UL ( n = 3 11:UL ( n = 3 11:UL ( n = 2 11:UL (	= 3, 2.5 , 2.56% 2, 1.71 = 2, 1. , 1.71% 2, 1.71 = 2, 1. , 1.71% 2, 1.71 = 2, 1.7 = 1, 1.7 = 1, 1.7 = 2, 1.7 = 1, 1.7 =
legend) is reserved for combinations not found by selection.	all the the activ odisplay	ve ved								ST ST ST ST ST ST ST ST ST ST ST ST ST	3 Collection: 4 Collection: 17 Collection 25 Collection 3 Collection: 69 Collection: 8 Collection: 8 Collection: 15 Collection 20 Collection	NYMC ( n = 3 n:UL ( n = 3 n:UL ( n = 3 n:UL ( n = 2 n:NYMC ( n UL ( n = 2 n:UL ( n = 2 n:UL ( n = 2 UL ( n = 2 n:NYMC ( n n:UL ( n = 2	= 3, 2.5 , 2.56% 2, 1.71 = 2, 1. , 1.71% 2, 1.71% 2, 1.71 = 2, 1.7 = 2, 1.7 = 2, 1.7 = 2, 1.7 = 2, 1.7 = 1.71% n = 1, 0
legend) is reserved for ombinations not found by selection.	all the the activ odisplay	ve ved								ST: ST: ST: ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 3 Collection: 69 Collection: 8 Collection: 8 Collection: 15 Collection 20 Collection 20 Collection 3 Collection	NYMC (n = UL (n = 3 h:UL (n = n:QIMR (n n:NYMC (n UL (n = 2 ):UL (n = QIMR (n = UL (n = 2 h:NYMC (n n:UL (n = h:UL (n = n:UL (n =	= 3, 2.5 , 2.56% 2, 1.71% = 2, 1. , 1.71% 2, 1.71% = 2, 1. , 1.71% = 2, 1.7 = 2, 1.7 = 2, 1.7 = 2, 1.7 = 1.71% 1 = 1, 0 1, 0.85% = 1, 0.8
legend) is reserved for ombinations not found by selection. I the figure legend it is als for each combination, the	r all the y the activ so display e absolute	ve ved e								517: 577: 577: 577: 577: 577: 577: 577:	3 Collection: 4 Collection: 17 Collection 25 Collection 3 Collection: 8 Collection: 8 Collection: 15 Collection 20 Collection 20 Collection 3 Collection: 38 Collection	NYMC (n = UL (n = 3 ::UL (n = n:QIMR (n = UL (n = 2 ::UL (n = NYMC (n = UL (n = 2 n:NYMC (n = UL (n = 2 n:NYMC (n = n:UL (n = n:UL (n = n:UL (n = n:UL (n = n:UL (n = n:NYMC (n = n:UL (n = n:NYMC (n = n))))))))))))))))))))))))))))))))))	= 3, 2.5 $, 2.56%$ $2, 1.71%$ $= 2, 1.71%$ $2, 1.71%$ $2, 1.71%$ $2, 1.71%$ $= 2, 1.7$ $, 1.71%$ $1, 0.85%$ $= 1, 0.8%$ $= 1, 0.8%$
legend) is reserved for ombinations not found by selection. If the figure legend it is als for each combination, the numbers of the isolates f	r all the y the activ so display e absolute found and	ve ved e								ST: ST: ST: ST: ST: ST: ST: ST: ST: ST:	3 Collection: 4 Collection: 17 Collection 25 Collection 69 Collection: 8 Collection: 8 Collection: 15 Collection 15 Collection 20 Collection 29 Collection 38 Collection 38 Collection	NYMC (n = 3 1:UL (n = 3 1:UL (n = - 1:UIM (n : 1:NYMC (n = - 1:UL (n = - NYMC (n = - QIMR (n = - 2)UL (n = - 1:UL (n = - 1:UL (n = - 1:UL (n = - 1:VIMC (n = - 1:NYMC (n = - 1:NYMC (n = -))	= 3, 2.5 $, 2.56%$ $2, 1.71%$ $= 2, 1.71%$ $2, 1.71%$ $2, 1.71%$ $2, 1.71%$ $1, 0.85%$ $= 1, 0.8%$ $1, 0.85%$ $= 1, 0.8%$
legend) is reserved for ombinations not found by selection. the figure legend it is als for each combination, the	r all the y the activ so display e absolute found and	ve ved e								517: 577: 577: 577: 577: 577: 577: 577:	3 Collection: 4 Collection: 17 Collection 25 Collection 3 Collection: 8 Collection: 8 Collection: 15 Collection 20 Collection 20 Collection 3 Collection: 38 Collection	NYMC (n = 3 1:UL (n = 3 1:UL (n = 1 1:UL (n = 1 1:UL (n = 2 1:UL (n = 2 1:UL (n = 2 1:UL (n = 2 1:UL (n = 1 1:UL (n = 1 1:UL (n = 1 1:UL (n = 1 1:UL (n = 1) 1:UL (n = 1 1:UL (n = 1) 1:UL	= 3, 2.5 , 2.56% 2, 1.71% = 2, 1. , 1.71% 2, 1.71% 2, 1.71% 1, 1.71% 1 = 2, 1. , 1.71% 1 = 2, 1. , 1.71% 1 = 1, 0.85% 1, 0.85% = 1, 0.85%= 1, 0.85% = 1, 0.85% = 1, 0.85%= 1, 0.85% = 1, 0.85% = 1, 0.85%= 1, 0.85%= 1, 0.85% = 1, 0.85%= 1,
legend) is reserved for ombinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 30 Collection: 8 Collection: 8 Collection: 15 Collection 20 Collection 20 Collection 3 Collection 3 Collection 3 Collection 4 Collection: 4 Collection:	NYMC(n = 3) 1:UL(n = 3) 1:UL(n = 1) 1:UL(n = 1) 1:UL(n = 1) 1:UL(n = 2) 1:UL(n = 2) 1:UL(n = 2) 0UL(n	= 3, 2.5 $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 2, 1.713$ $= 1, 0.853$
legend) is reserved for combinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f	r all the y the activ so display e absolute found and	ve ved e								517 517 517 517 517 517 517 517 517 517	3 Collection: 4 Collection: 17 Collection 25 Collection 29 Collection 29 Collection: 50 Collection: 8 Collection: 15 Collection 20 Collection 20 Collection 23 Collection: 38 Collection 38 Collection 4 Collection	NYMC (n = 3 :UL (n = 3 :UL (n = 2 :UL (n = 2 :UL (n = 2 :NYMC (n : UL (n = 2 :NYMC (n : UL (n = 2 :NYMC (n : :UL (n = 1 :UL (n = 1 :UL (n = 1 :UL (n = 1 :UL (n = 1) :UL (n =	= 3, 2.5, 2.56% $2, 1.719$ $= 2, 1.719$ $= 2, 1.719$ $= 2, 1.719$ $= 2, 1.719$ $= 2, 1.719$ $= 2, 1.719$ $= 2, 1.71%$ $= 1, 0.0859$ $= 1, 0.859$
legend) is reserved for combinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 25 Collection: 25 Collection: 25 Collection: 3 Collection: 8 Collection: 8 Collection: 20 Collection: 20 Collection: 20 Collection: 38 Collection: 38 Collection: 4 Collection: 5 Collec	NYMC (n = 3 ::UL (n = 3 ::UL (n = 2 ::UL (n = 2) ::UL (n = 2 ::UL (n = 2) ::UL (n = 2 ::UL (n = 2) ::UL (n = 2) :UL (n = 2)	= 3, 2.5 , 2, 56% ( 2, 1, 719 = 2, 1, 719 = 1, 0, 859 = 1, 0, 859
legend) is reserved for ombinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 25 Collection 25 Collection 29 Collection 30 Collection: 40 Collection: 40 Collection: 40 Collection: 30 Collection: 30 Collection: 30 Collection: 30 Collection: 30 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 50 Collecti	NYMC (n = 3 ::UL (n = 3 ::UL n = 1 ::UL n = 1 ::UL n = 2 ::UL (n = 1 ::UL (n = 2 ::UL (n = 1 ::UL (n = 1) ::UL (n = 1) :UL (n = 1) ::UL (n =	= 3, 2.5, 2, 2.56% $2, 1.715$ $= 2, 1.715$ $= 2, 1.715$ $= 2, 1.715$ $= 2, 1.775$ $= 2, 1.775$ $= 2, 1.775$ $1, 0.855$
tombinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 25 Collection: 25 Collection: 25 Collection: 3 Collection: 8 Collection: 8 Collection: 20 Collection: 20 Collection: 20 Collection: 38 Collection: 38 Collection: 4 Collection: 5 Collec	NYMC (n = 3 ::UL (n = 3 ::UL n = 1 ::UL n = 1 ::UL n = 2 ::UL (n = 1 ::UL (n = 2 ::UL (n = 1 ::UL (n = 1) ::UL (n = 1) :UL (n = 1) ::UL (n =	= 3, 2.5 , 2, 56% () 2, 1, 719 = 2, 1, 719 = 1, 0, 859 = 1, 0, 1,
legend) is reserved for combinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 25 Collection 29 Collection: 29 Collection: 3 Collection: 8 Collection: 8 Collection: 10 Collection: 20 Collection: 20 Collection: 20 Collection: 33 Collection: 33 Collection: 34 Collection: 35 Collection: 45 Collection: 45 Collection: 45 Collection: 45 Collection: 45 Collection: 45 Collection: 45 Collection: 45 Collection: 55 Collection: 52 Collection: 52 Collection: 52 Collection: 52 Collection: 52 Collection: 52 Collection: 52 Collection: 53 Collection: 53 Collection: 54 Collection: 55 Collectio	NYMC (n = 3 ::UL (n = 2 ::UL (n = 1) ::UL (n = 2 ::UL (n = 1) ::UL (n = 2 ::UL (n = 1) ::UL (n = 1) :	= 3, 2.5 $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 2, 1.7, 19$ $= 1, 0.859$ $= 1, 0.859$ $= 1, 0.859$ $= 1, 0.859$ $= 1, 0.859$ $1, 0$
legend) is reserved for combinations not found by selection. In the figure legend it is als for each combination, the numbers of the isolates f percentage of the datas	r all the y the activ so display e absolute found and	ve ved e								517: 517: 517: 517: 517: 517: 517: 517:	3 Collection: 4 Collection: 25 Collection 25 Collection 29 Collection 30 Collection: 40 Collection: 40 Collection: 40 Collection: 30 Collection: 30 Collection: 30 Collection: 30 Collection: 30 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 40 Collection: 50 Collecti	NYMC (n = 3 ::UL (n = 3 ::UL n = 1 ::UL n = 1 ::UL n = 2 ::UL (n = 1 ::UL (n = 2 ::UL (n = 1 ::UL (n = 1) ::UL (n = 1) :UL (n = 1) ::UL (n =	$\begin{array}{l} = 3, 2, 5, 5\\ 2, 1, 719\\ 2, 2, 1, 719\\ 2, 2, 1, 719\\ 2, 2, 1, 719\\ 2, 1, 719\\ 2, 1, 719\\ 2, 1, 719\\ 2, 1, 719\\ 2, 1, 719\\ 2, 1, 719\\ 1, 1, 710\\ 1, 1, 1, 710\\ 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, $

• Results on Display Tab

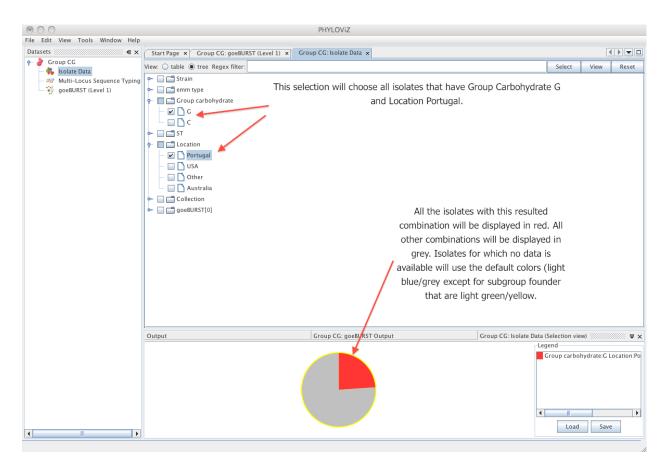


## 5.5 Queries using the tree view

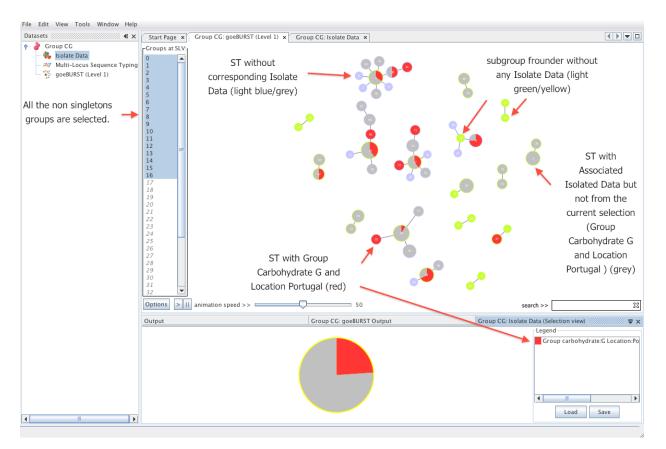
The *Tree* view offers a faster way to create simple queries. The user can also use the regex filter to search the dataset but all the possibilities for each dataset column are automatically indexed in a tree like manner. By pressing the *Select* button and switching to *Table* view the user can see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

#### 5.5.1 Query examples

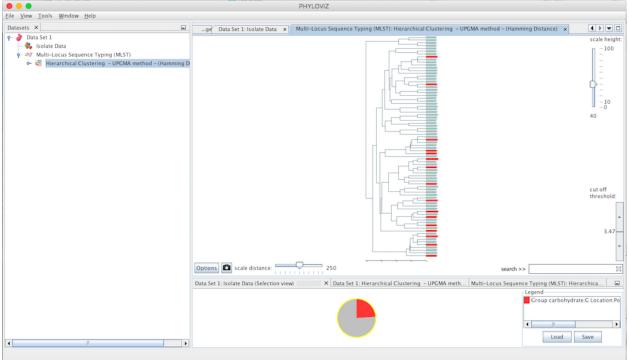
• Tree view with selections



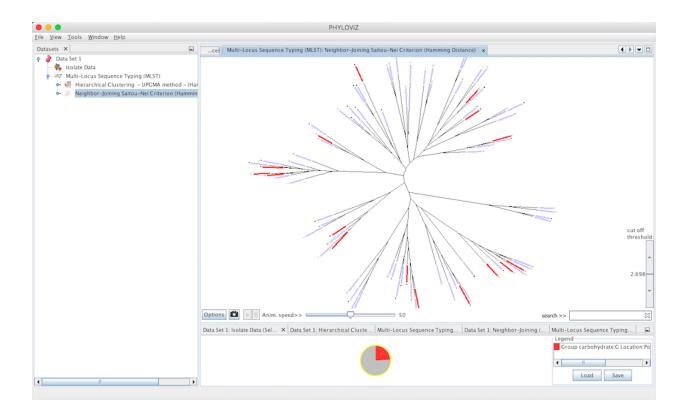
• Queries on the results produced by the goeBURST and goeBURST Full MST algorithm



#### • Queries on the results produced by the Hierarchical Clustering algorithm

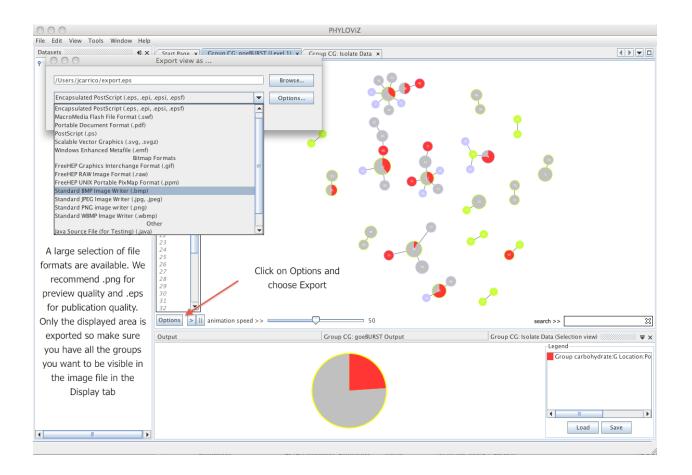


• Queries on the results produced by the Neighbor Joinning algorithm



# 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.

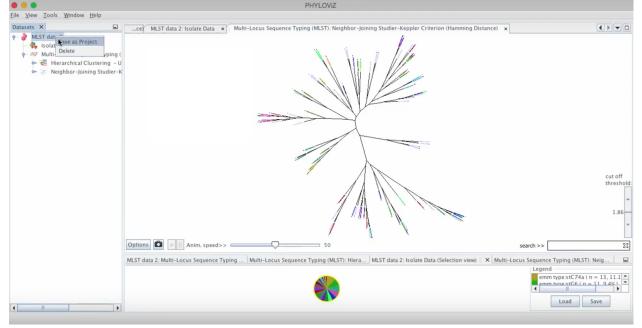


### **Project management**

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.

000	PHYLOVIZ	
<u>File View Tools Window H</u> elp		
Datasets X III MLST data 2 MLST data 2 Stolate Data P W Multi-Locus Sequence Typing ( C Hierarchical Clustering - U C X Neighbor-Joining Studier-K		x
	Folder game:       /Users/PhylovizProjects         Files of Type:       All Files         Save       Cancel         Qptions       Image: Cancel         MLST data 2:       Multi-Locus Sequence Typing         MLST data 2:       Multi-Locus Sequence Typing         MLST data 2:       Multi-Locus Sequence Typing	cut off threshold .1.86 w search >> Search Search Sear
		Load Save

# 6.2 Loading

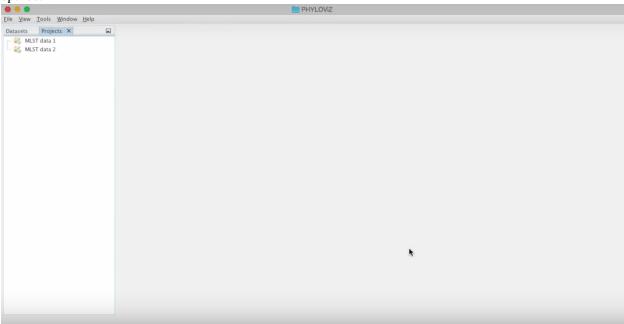
Go to File menu and choose Open Project.

Eile View Tools Window Help	
∂ Load Dataset	
Load Dataset from MLST DBs	
Open Project X- Open Recent Project	+①-〇
Open Recent Project	•
<u>C</u> lose Project	
Close Other Projects	
Close All Projects	
Save	

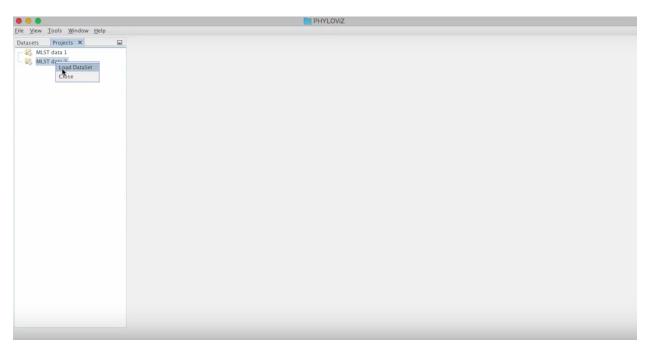
The next step is to find the project that you would like to load. After finding it click on Open Project.

0.0		PHYLOVIZ
ile <u>V</u> iew <u>T</u> ools <u>W</u> indow		
atasets X		
	• • •	Open Project
	Look In: PhylovizProjects	
	🔣 MLST data 2	Project Name: MLST data 2
		Open <u>R</u> equired Projects:
	Folder name: //Users/PhylovizProjects/MLST data	2
	Files of Type: Project Folder	
		Open Project Cancel

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.

