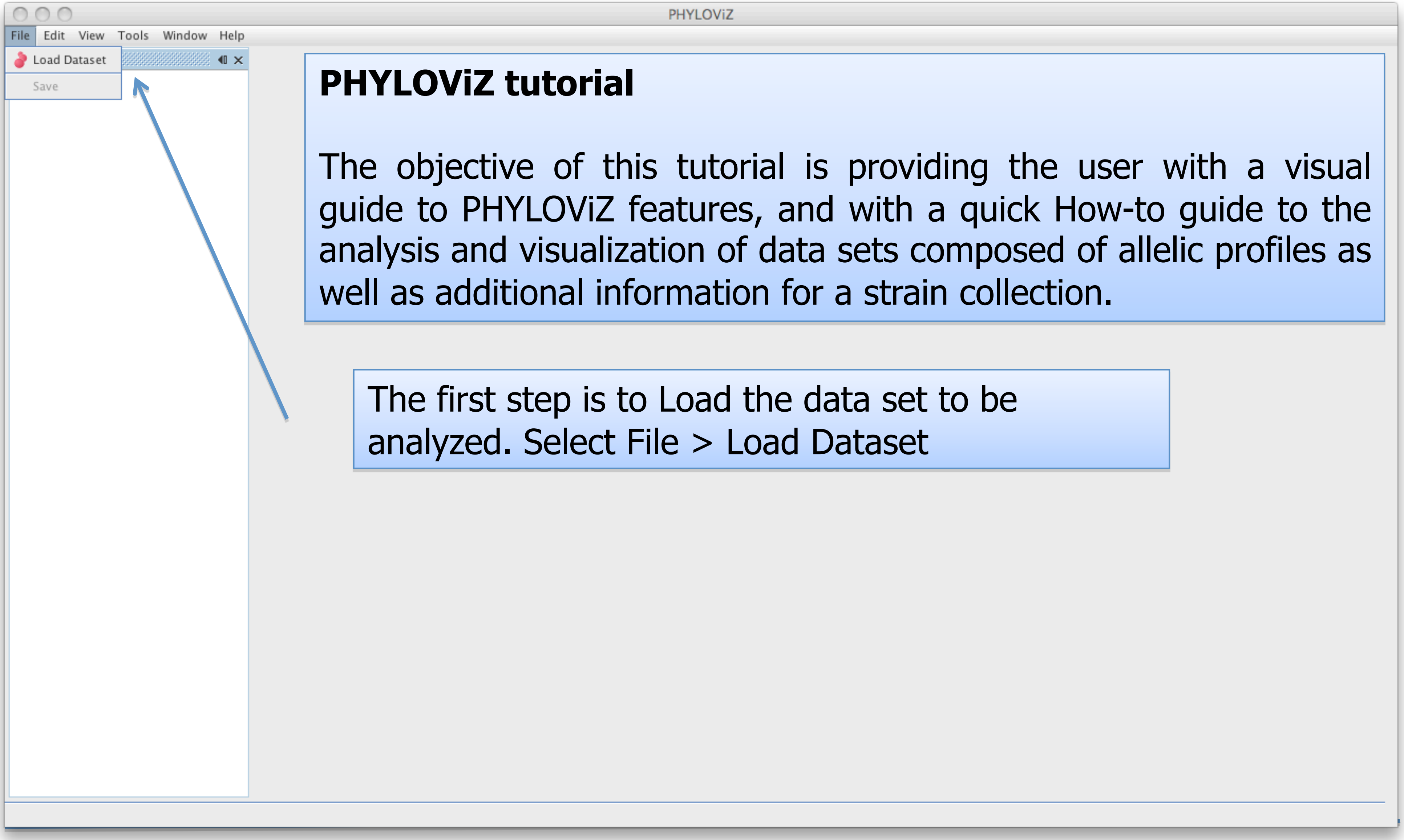
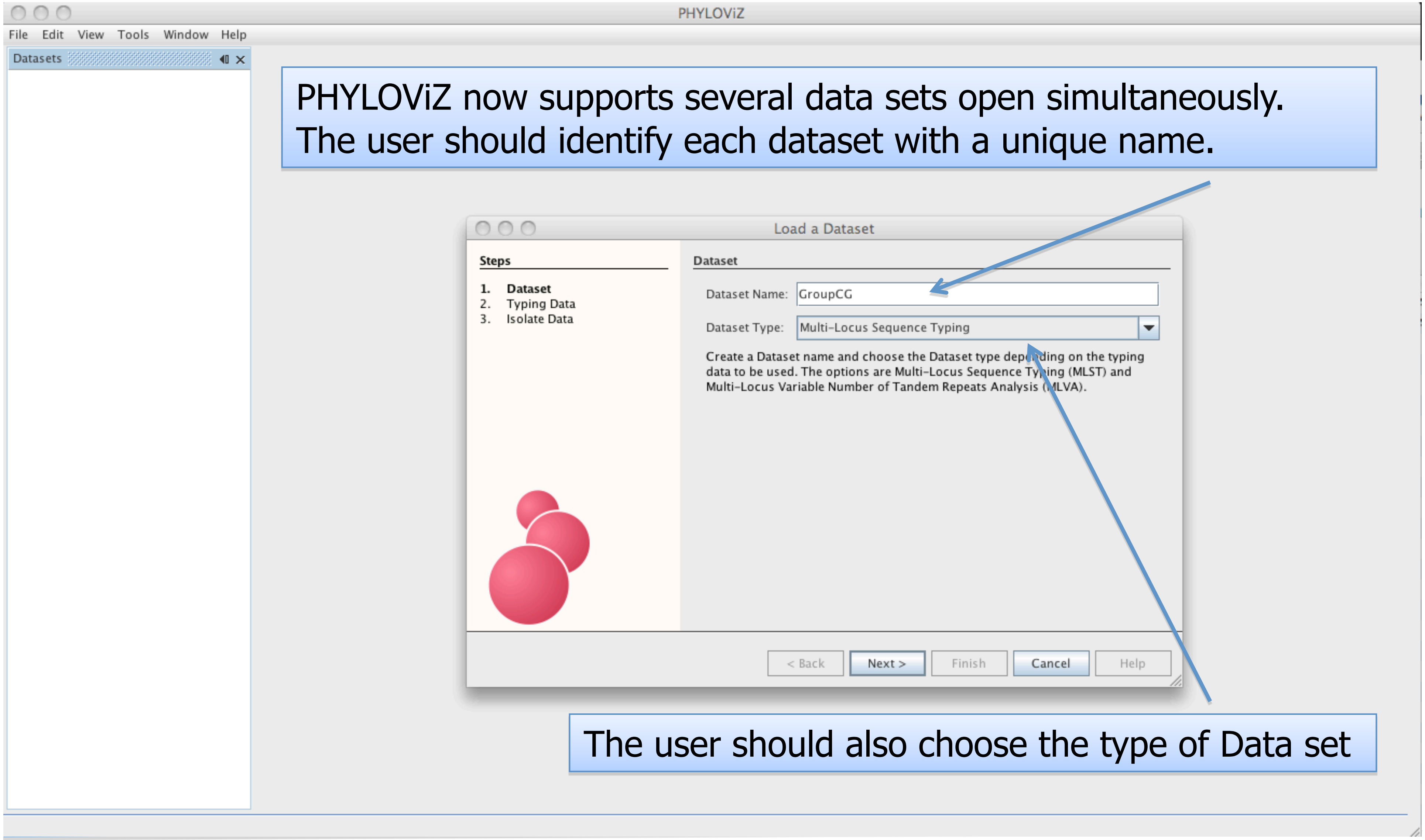


PHYLOViZ tutorial

The objective of this tutorial is providing the user with a visual guide to PHYLOViZ features, and with a quick How-to guide to the analysis and visualization of data sets composed of allelic profiles as well as additional information for a strain collection.

The first step is to Load the data set to be analyzed. Select File > Load Dataset

PHYLOViZ now supports several data sets open simultaneously. The user should identify each dataset with a unique name.



Steps

1. **Dataset**
2. Typing Data
3. Isolate Data

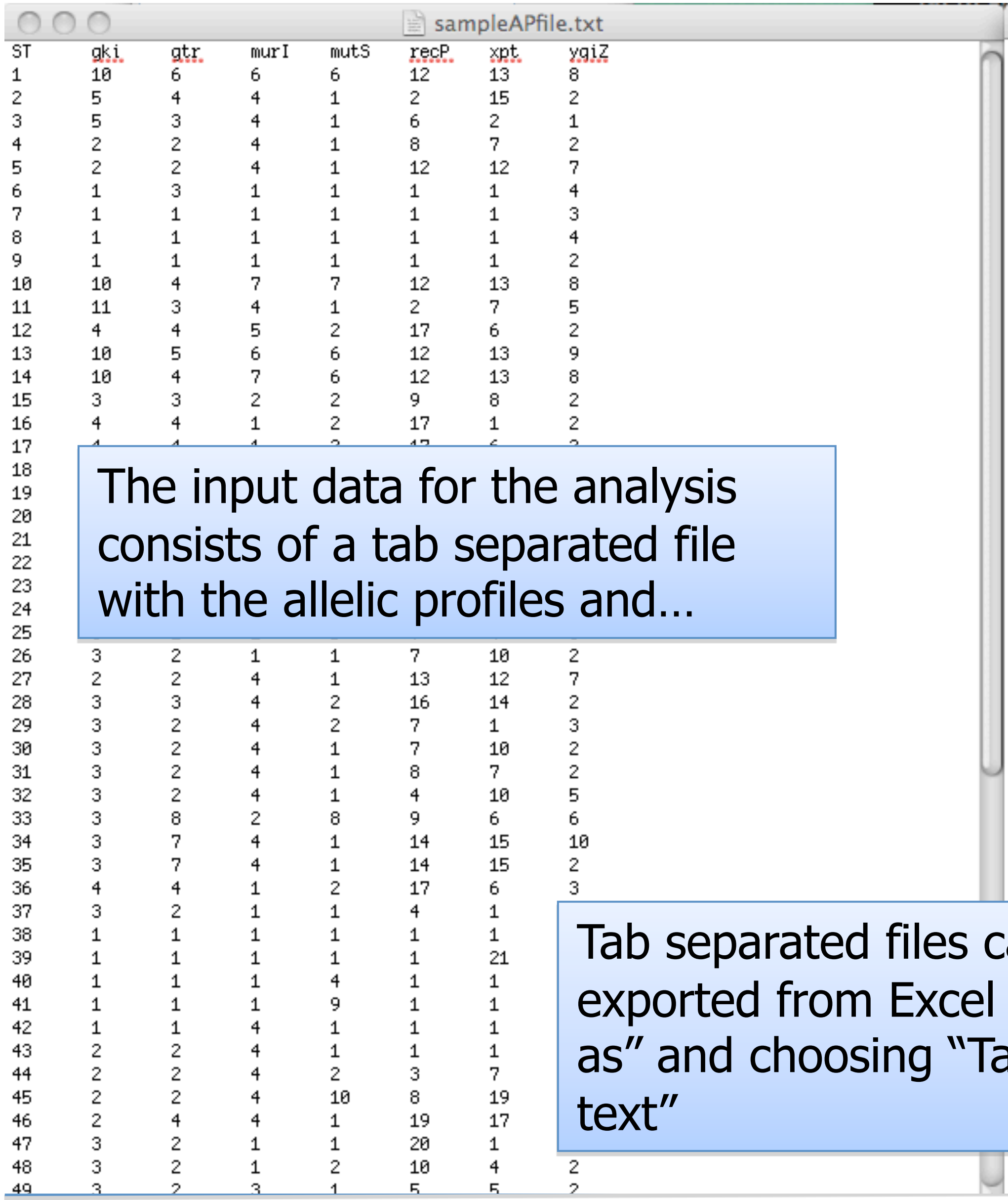
Dataset

Dataset Name: GroupCG

Dataset Type: Multi-Locus Sequence Typing

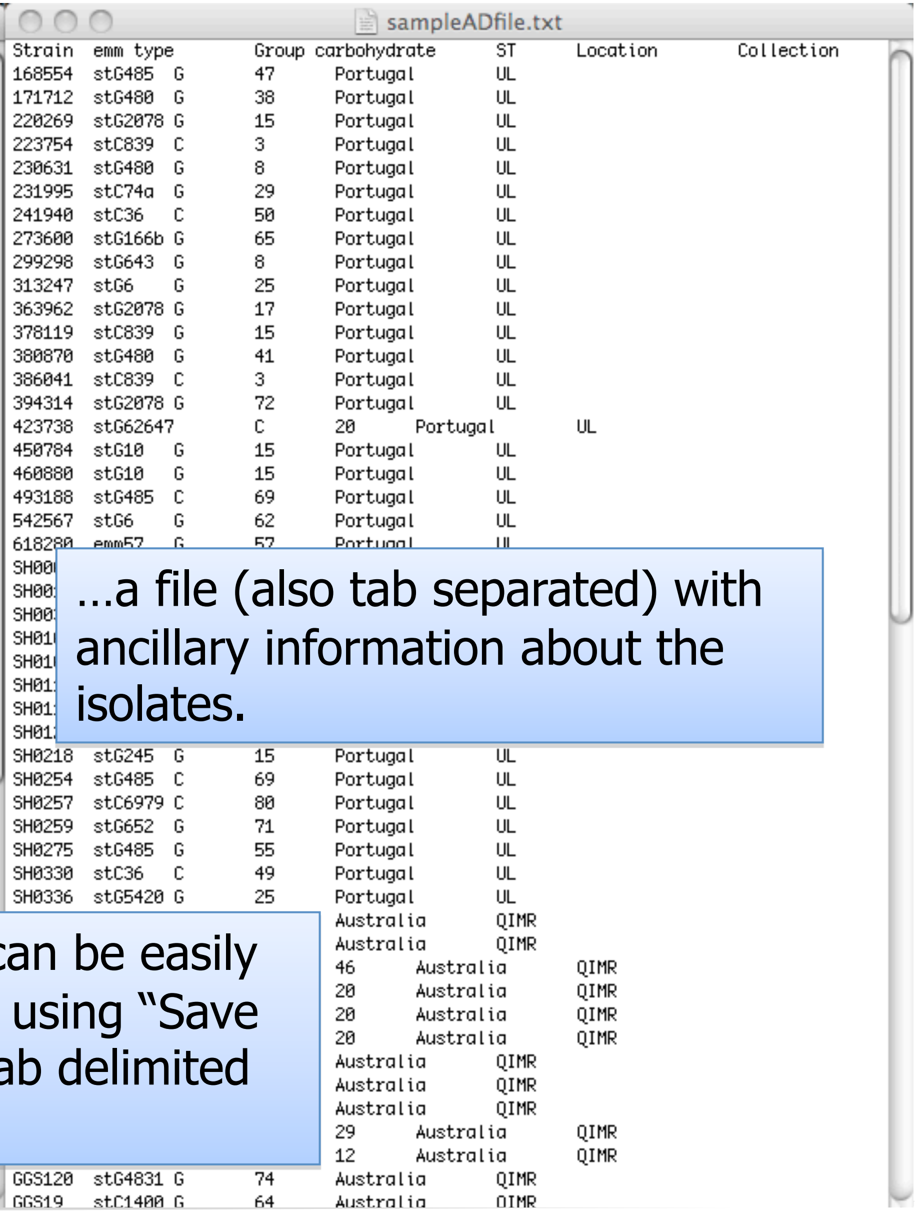
Create a Dataset name and choose the Dataset type depending on the typing data to be used. The options are Multi-Locus Sequence Typing (MLST) and Multi-Locus Variable Number of Tandem Repeats Analysis (MLVA).

The user should also choose the type of Data set



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

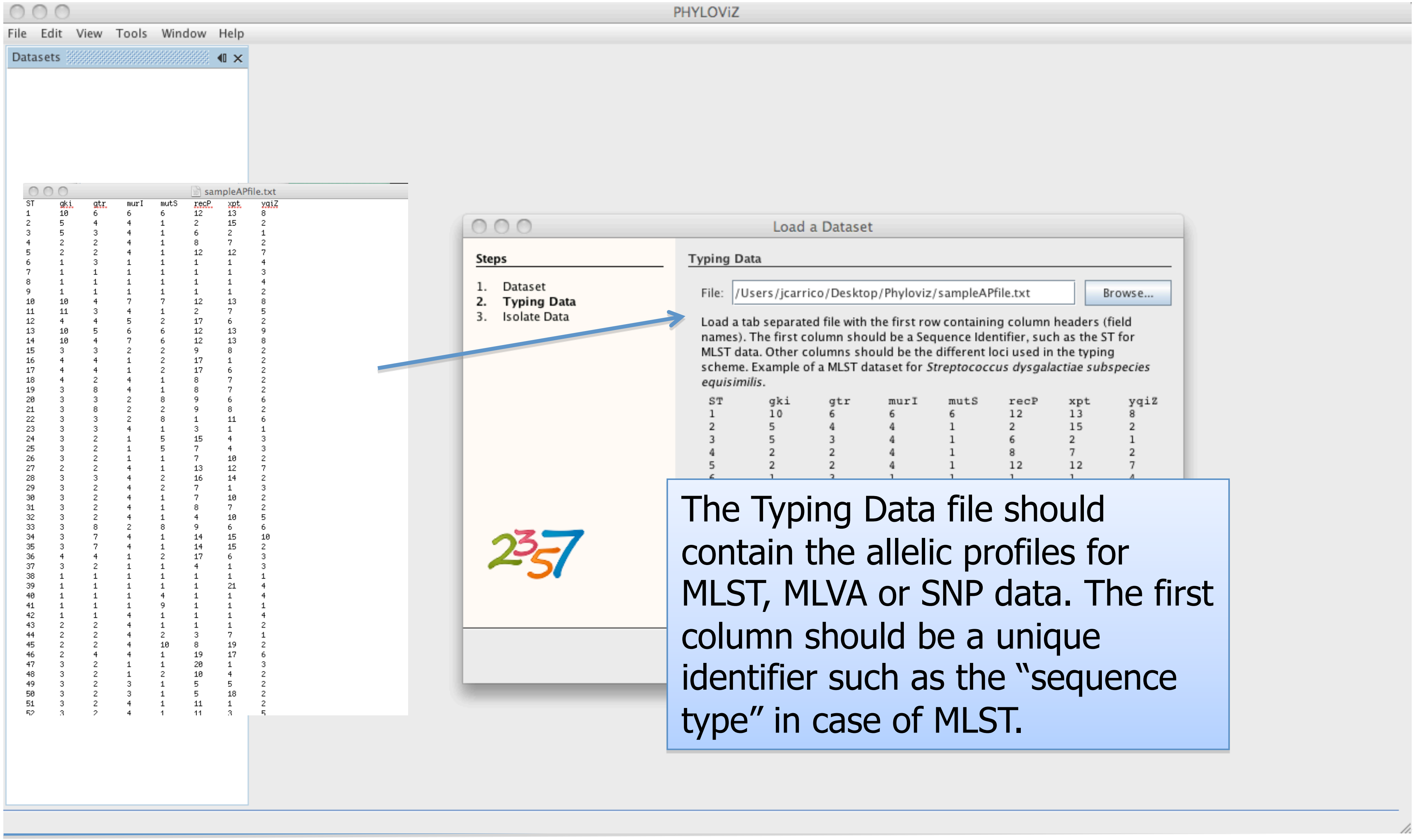
The input data for the analysis consists of a tab separated file with the allelic profiles and…



```
Strain  emm type        Group carbohydrate      ST      Location        Collection
168554  stG485  G       47      Portugal        UL
171712  stG480  G       38      Portugal        UL
220269  stG2078 G       15      Portugal        UL
223754  stC839  C       3       Portugal        UL
230631  stG480  G       8       Portugal        UL
231995  stC74a  G       29      Portugal        UL
241940  stC36   C       50      Portugal        UL
273600  stG166b G       65      Portugal        UL
299298  stG643  G       8       Portugal        UL
313247  stG6    G       25      Portugal        UL
363962  stG2078 G       17      Portugal        UL
378119  stC839  G       15      Portugal        UL
380870  stG480  G       41      Portugal        UL
386041  stC839  C       3       Portugal        UL
394314  stG2078 G       72      Portugal        UL
423738  stG62647        C       20      Portugal        UL
450784  stG10   G       15      Portugal        UL
460880  stG10   G       15      Portugal        UL
493188  stG485  C       69      Portugal        UL
542567  stG6    G       62      Portugal        UL
618280  emm57   G       57      Portugal        UL
SH00
SH00
SH00
SH01
SH01
SH01
SH01
SH01
SH0218  stG245  G       15      Portugal        UL
SH0254  stG485  C       69      Portugal        UL
SH0257  stC6979 C       80      Portugal        UL
SH0259  stG652  G       71      Portugal        UL
SH0275  stG485  G       55      Portugal        UL
SH0330  stC36   C       49      Portugal        UL
SH0336  stG5420 G       25      Portugal        UL
                                Australia       QIMR
                                Australia       QIMR
                                46      Australia       QIMR
                                20      Australia       QIMR
                                20      Australia       QIMR
                                20      Australia       QIMR
                                Australia       QIMR
                                Australia       QIMR
                                Australia       QIMR
                                29      Australia       QIMR
                                12      Australia       QIMR
GGS120  stG4831 G       74      Australia       QIMR
GGS19   stC1400 G       64      Australia       QIMR
```

…a file (also tab separated) with ancillary information about the isolates.

Tab separated files can be easily exported from Excel using "Save as" and choosing "Tab delimited text"



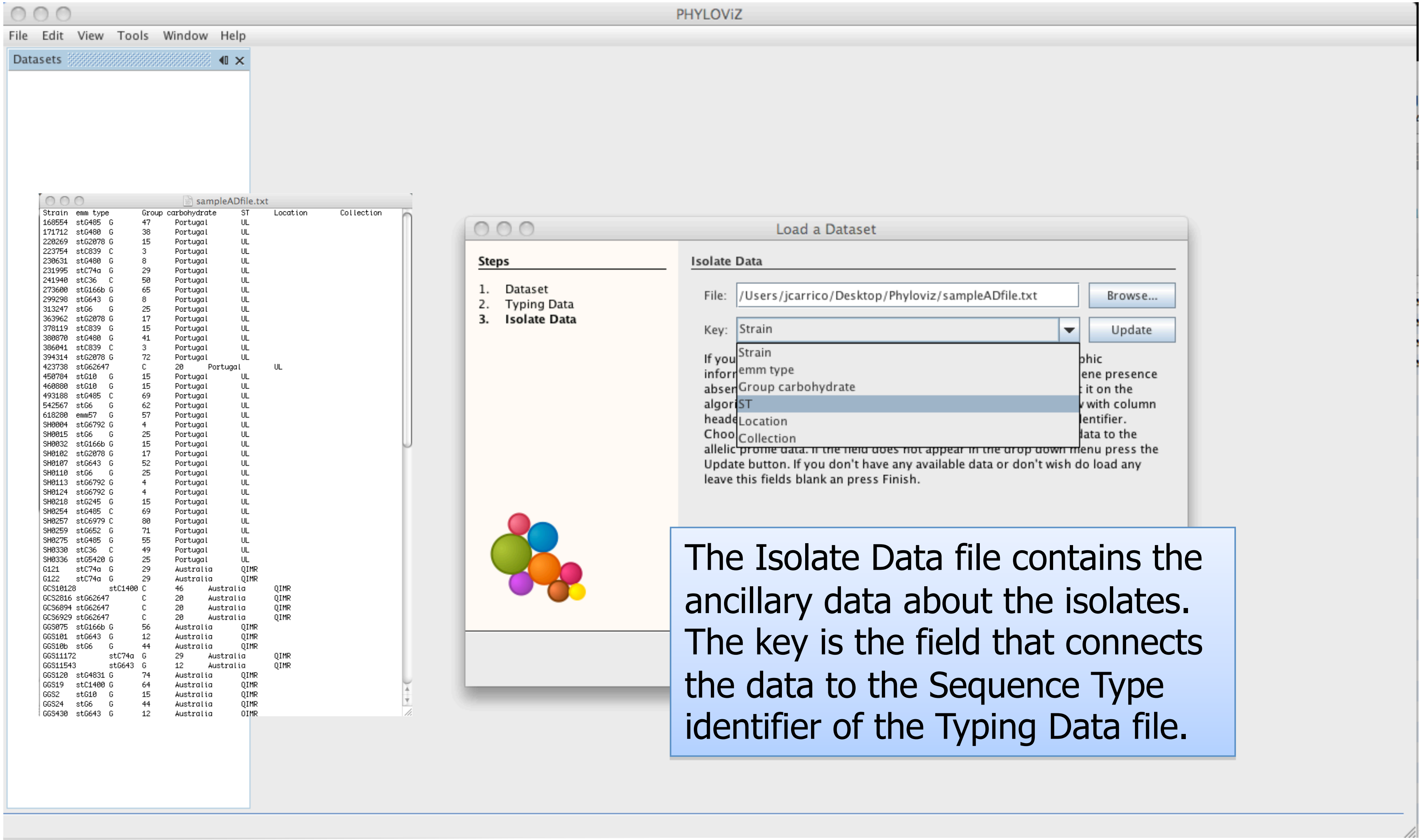
ST	gki	gtr	murI	mutS	recP	xpt	ygiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	13	9
14	10	4	7	6	12	13	8
15	3	3	2	2	9	8	2
16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

File: /Users/jcarrico/Desktop/Phyloviz/sampleAPfile.txt

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 *Streptococcus dysgalactiae subspecies equisimilis*.

ST	gki	gtr	murI	mutS	recP	xpt	ygiZ
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

The Typing Data file should contain the allelic profiles for MLST, MLVA or SNP data. The first column should be a unique identifier such as the "sequence type" in case of MLST.



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

1. Dataset
2. Typing Data
3. **Isolate Data**

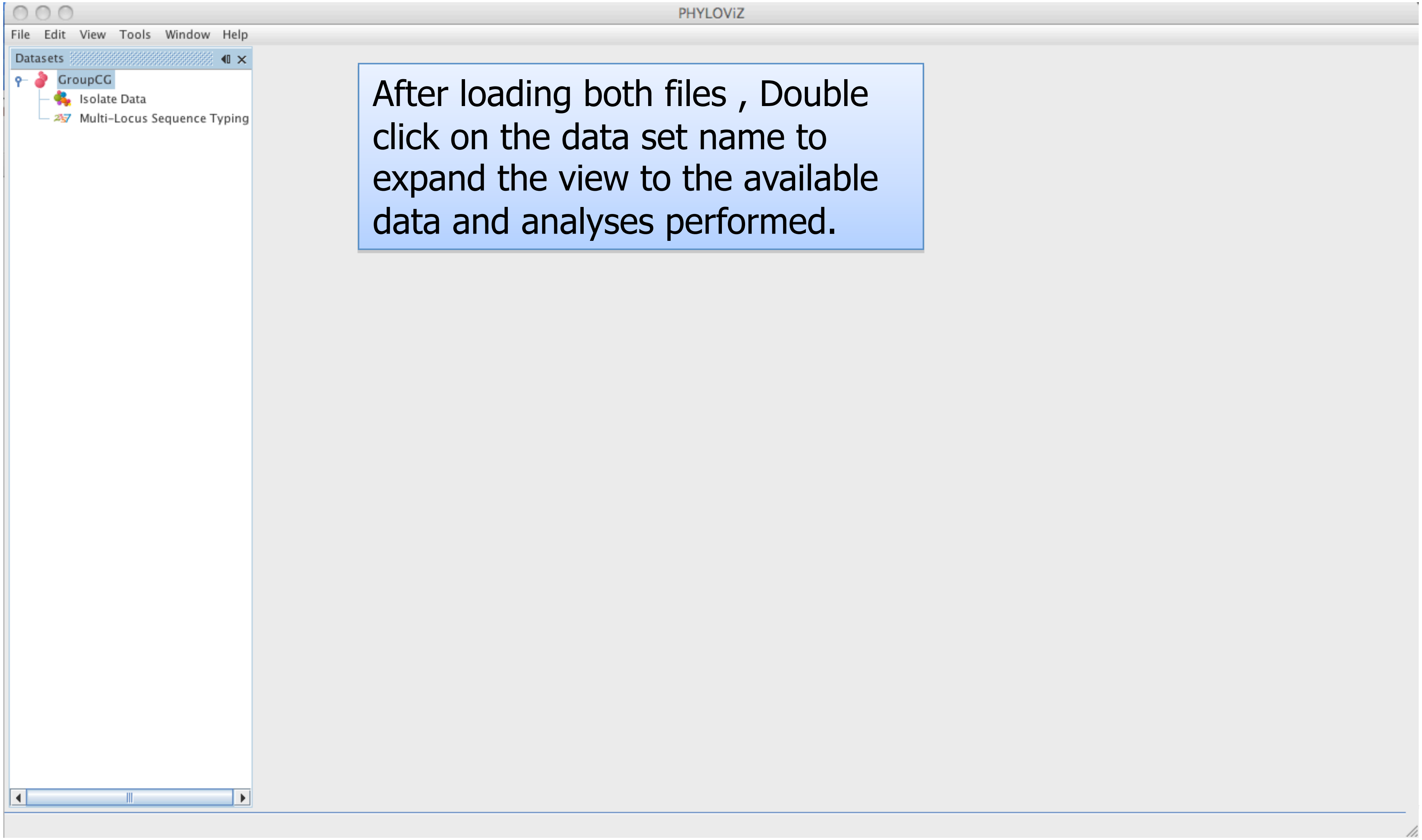
Isolate Data

File: /Users/jcarrico/Desktop/Phyloviz/sampleADfile.txt

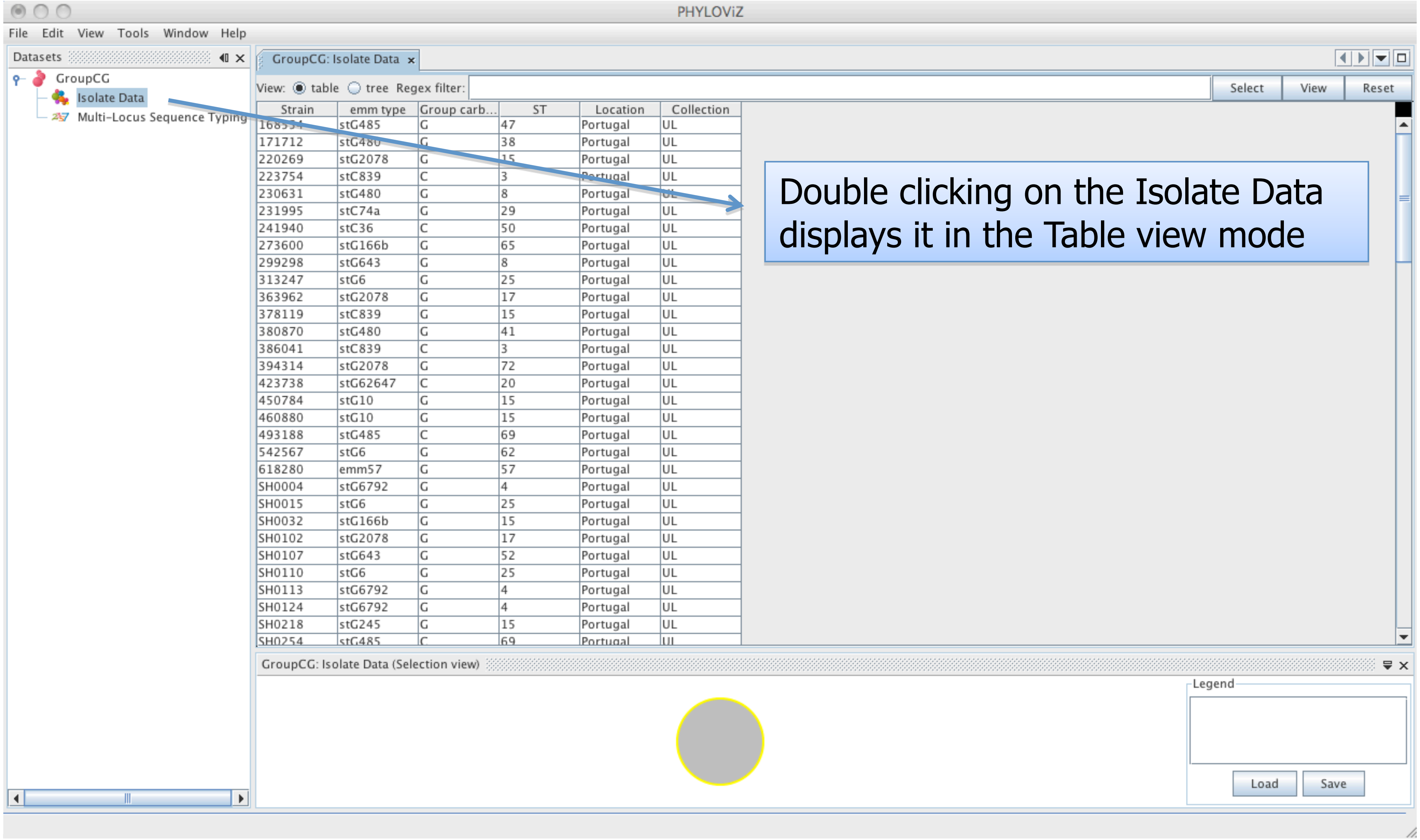
Key: Strain

If you … phic
infor… ene presence
absen… it on the
algori… with column
heade… entifier.
Choo… 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.

The Isolate Data file contains the ancillary data about the isolates. The key is the field that connects the data to the Sequence Type identifier of the Typing Data file.

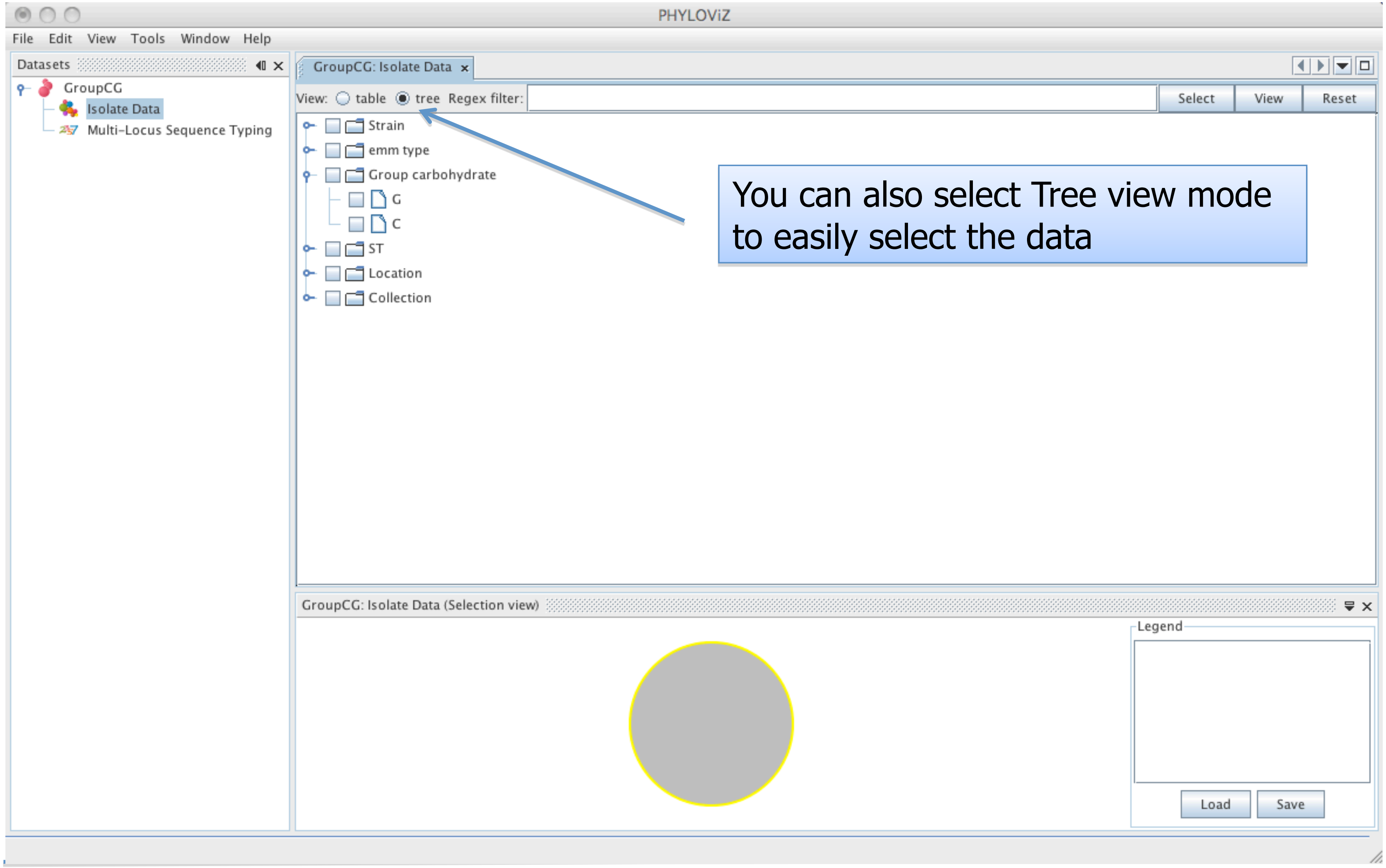


After loading both files , Double click on the data set name to expand the view to the available data and analyses performed.



Strain	emm type	Group carb...	ST	Location	Collection
168354	stG485	G	47	Portugal	UL
171712	stG480	G	38	Portugal	UL
220269	stG2078	G	15	Portugal	UL
223754	stC839	C	3	Portugal	UL
230631	stG480	G	8	Portugal	UL
231995	stC74a	G	29	Portugal	UL
241940	stC36	C	50	Portugal	UL
273600	stG166b	G	65	Portugal	UL
299298	stG643	G	8	Portugal	UL
313247	stG6	G	25	Portugal	UL
363962	stG2078	G	17	Portugal	UL
378119	stC839	G	15	Portugal	UL
380870	stG480	G	41	Portugal	UL
386041	stC839	C	3	Portugal	UL
394314	stG2078	G	72	Portugal	UL
423738	stG62647	C	20	Portugal	UL
450784	stG10	G	15	Portugal	UL
460880	stG10	G	15	Portugal	UL
493188	stG485	C	69	Portugal	UL
542567	stG6	G	62	Portugal	UL
618280	emm57	G	57	Portugal	UL
SH0004	stG6792	G	4	Portugal	UL
SH0015	stG6	G	25	Portugal	UL
SH0032	stG166b	G	15	Portugal	UL
SH0102	stG2078	G	17	Portugal	UL
SH0107	stG643	G	52	Portugal	UL
SH0110	stG6	G	25	Portugal	UL
SH0113	stG6792	G	4	Portugal	UL
SH0124	stG6792	G	4	Portugal	UL
SH0218	stG245	G	15	Portugal	UL
SH0254	stG485	C	69	Portugal	UL

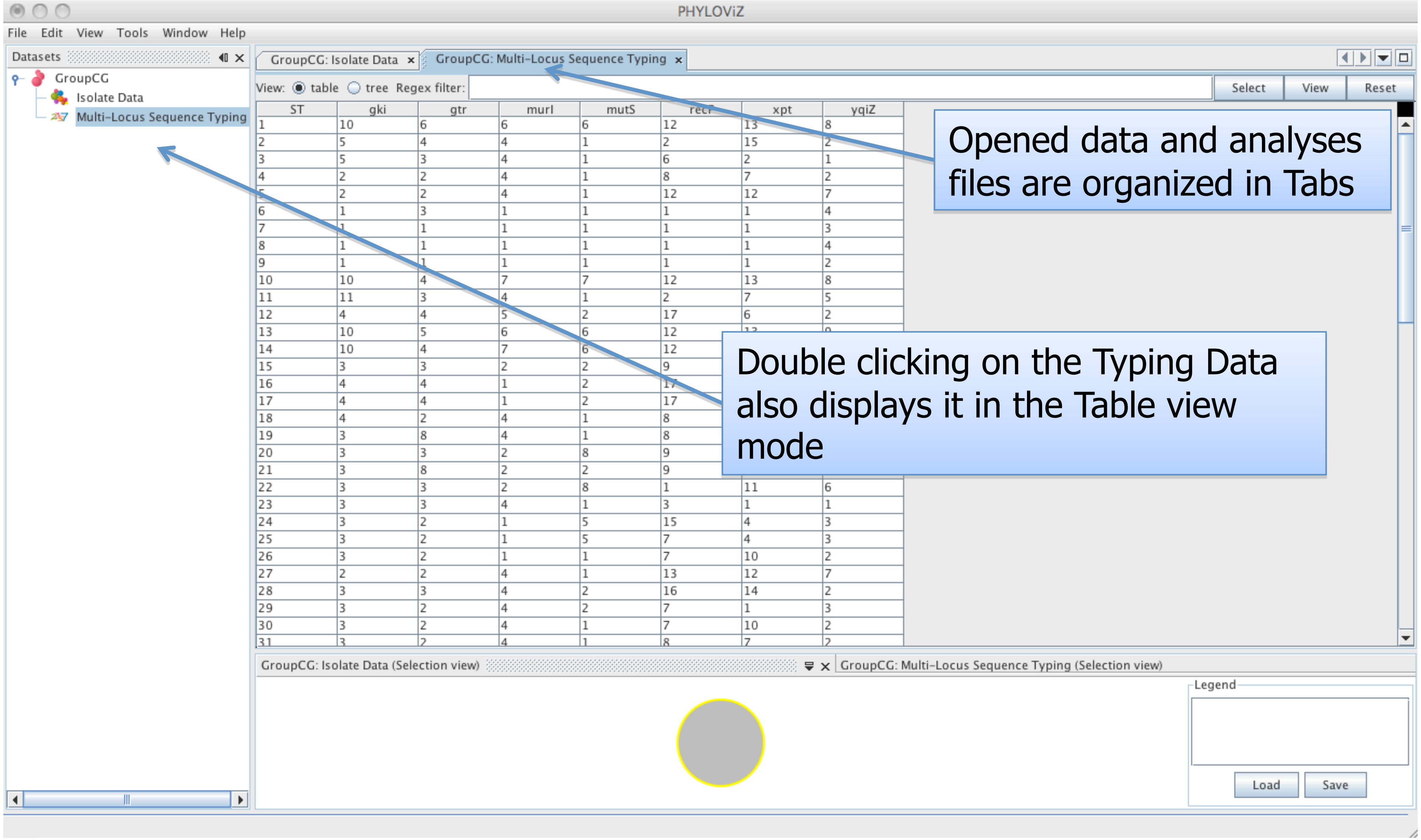
Double clicking on the Isolate Data displays it in the Table view mode



View: ○ table ● tree Regex filter:

- ☐ Strain
- ☐ emm type
- ☐ Group carbohydrate
  - ☐ G
  - ☐ C
- ☐ ST
- ☐ Location
- ☐ Collection

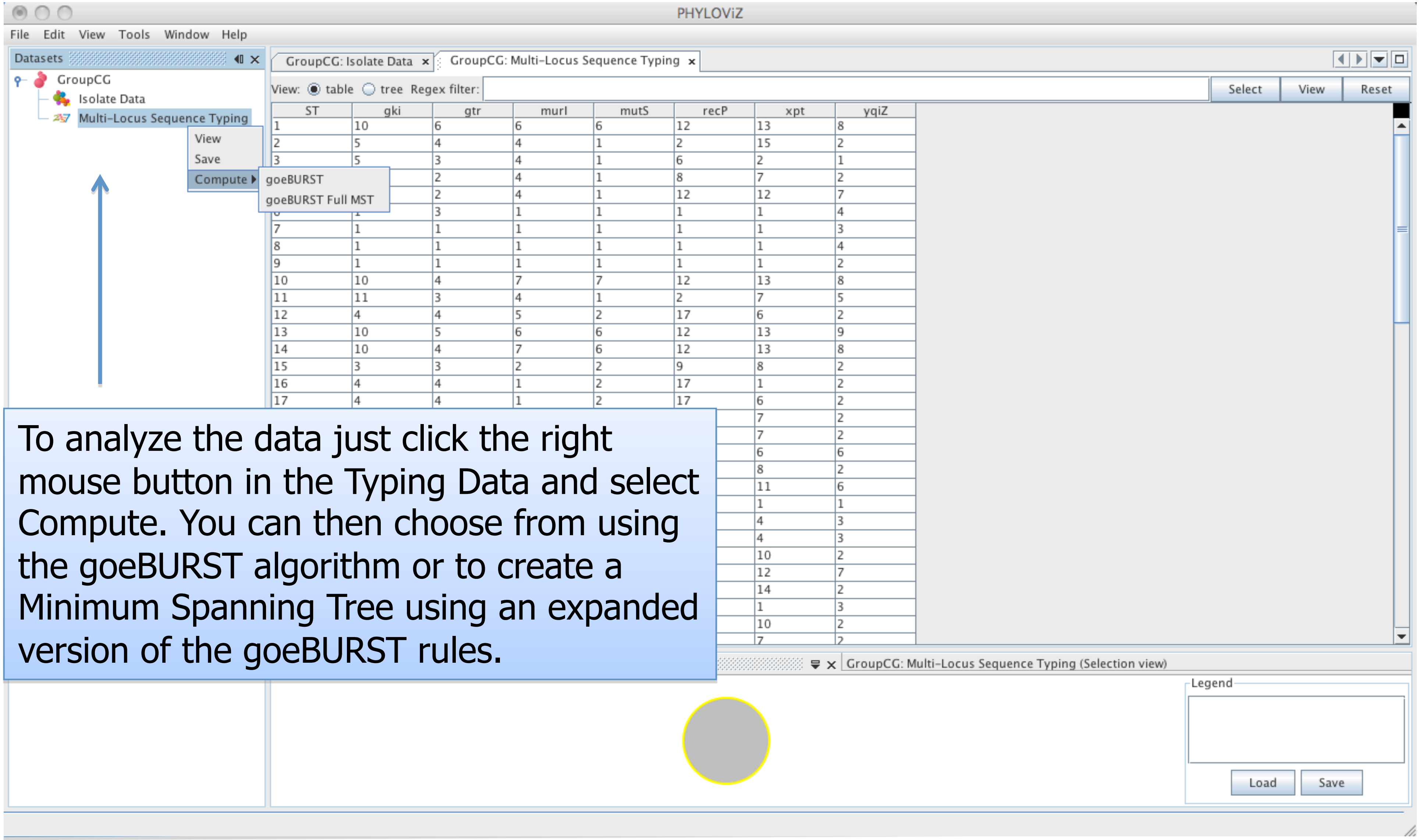
You can also select Tree view mode to easily select the data



ST	gki	gtr	murI	mutS	recF	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	[illegible]	[illegible]
14	10	4	7	6	12	[illegible]	[illegible]
15	3	3	2	2	9	[illegible]	[illegible]
16	4	4	1	2	17	[illegible]	[illegible]
17	4	4	1	2	17	[illegible]	[illegible]
18	4	2	4	1	8	[illegible]	[illegible]
19	3	8	4	1	8	[illegible]	[illegible]
20	3	3	2	8	9	[illegible]	[illegible]
21	3	8	2	2	9	[illegible]	[illegible]
22	3	3	2	8	1	11	6
23	3	3	4	1	3	1	1
24	3	2	1	5	15	4	3
25	3	2	1	5	7	4	3
26	3	2	1	1	7	10	2
27	2	2	4	1	13	12	7
28	3	3	4	2	16	14	2
29	3	2	4	2	7	1	3
30	3	2	4	1	7	10	2
31	3	2	4	1	8	7	2

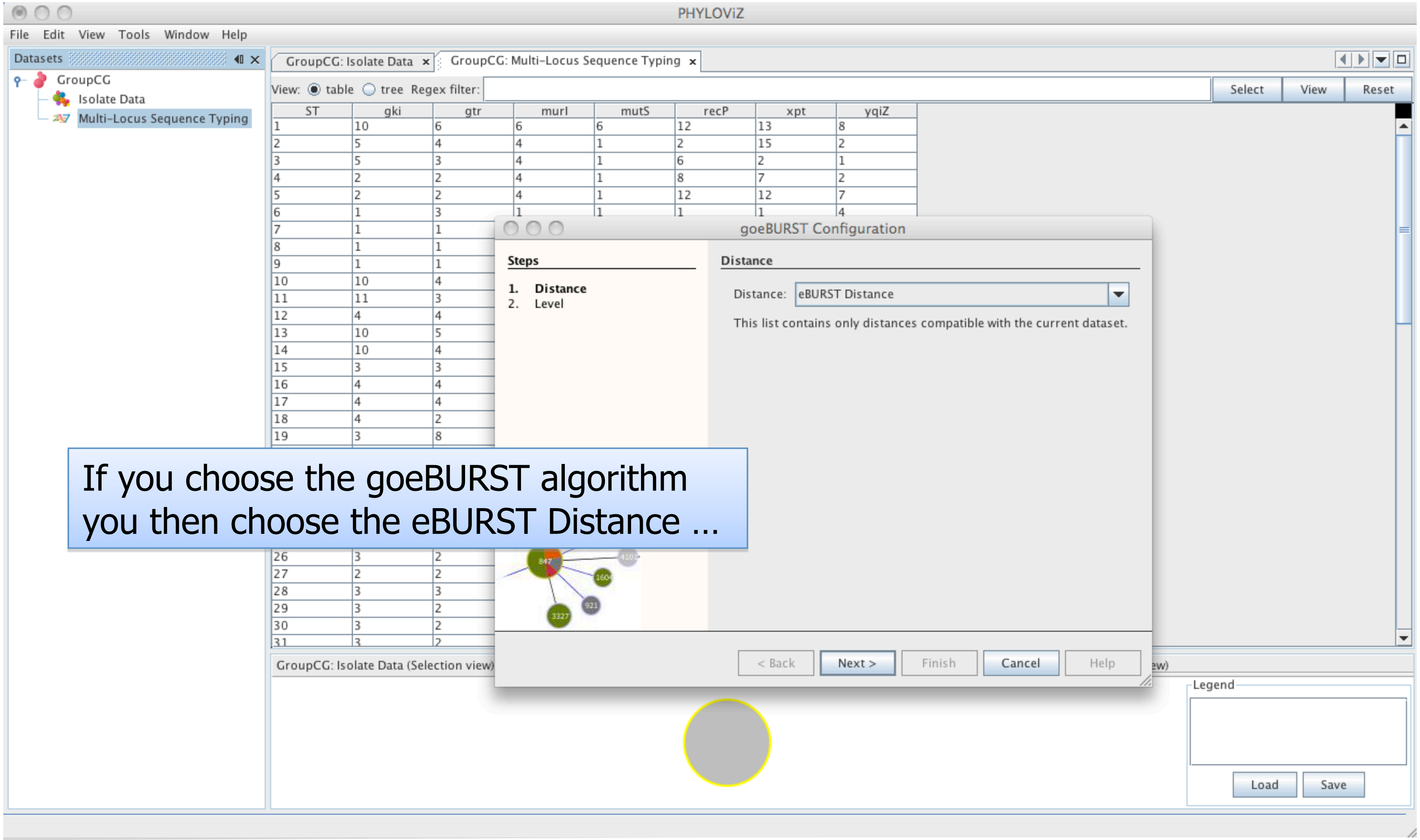
Opened data and analyses files are organized in Tabs

Double clicking on the Typing Data also displays it in the Table view mode



ST	gki	gtr	murI	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
		2	4	1	8	7	2
		2	4	1	12	12	7
		3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	13	9
14	10	4	7	6	12	13	8
15	3	3	2	2	9	8	2
16	4	4	1	2	17	1	2
17	4	4	1	2	17	6	2
						7	2
						7	2
						6	6
						8	2
						11	6
						1	1
						4	3
						4	3
						10	2
						12	7
						14	2
						1	3
						10	2
						7	2

To analyze the data just click the right mouse button in the Typing Data and select Compute. You can then choose from using the goeBURST algorithm or to create a Minimum Spanning Tree using an expanded version of the goeBURST rules.



View: ● table ○ tree Regex filter:

ST	gki	gtr	murI	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1					
8	1	1					
9	1	1					
10	10	4					
11	11	3					
12	4	4					
13	10	5					
14	10	4					
15	3	3					
16	4	4					
17	4	4					
18	4	2					
19	3	8					
26	3	2					
27	2	2					
28	3	3					
29	3	2					
30	3	2					
31	3	2					

goeBURST Configuration

Steps

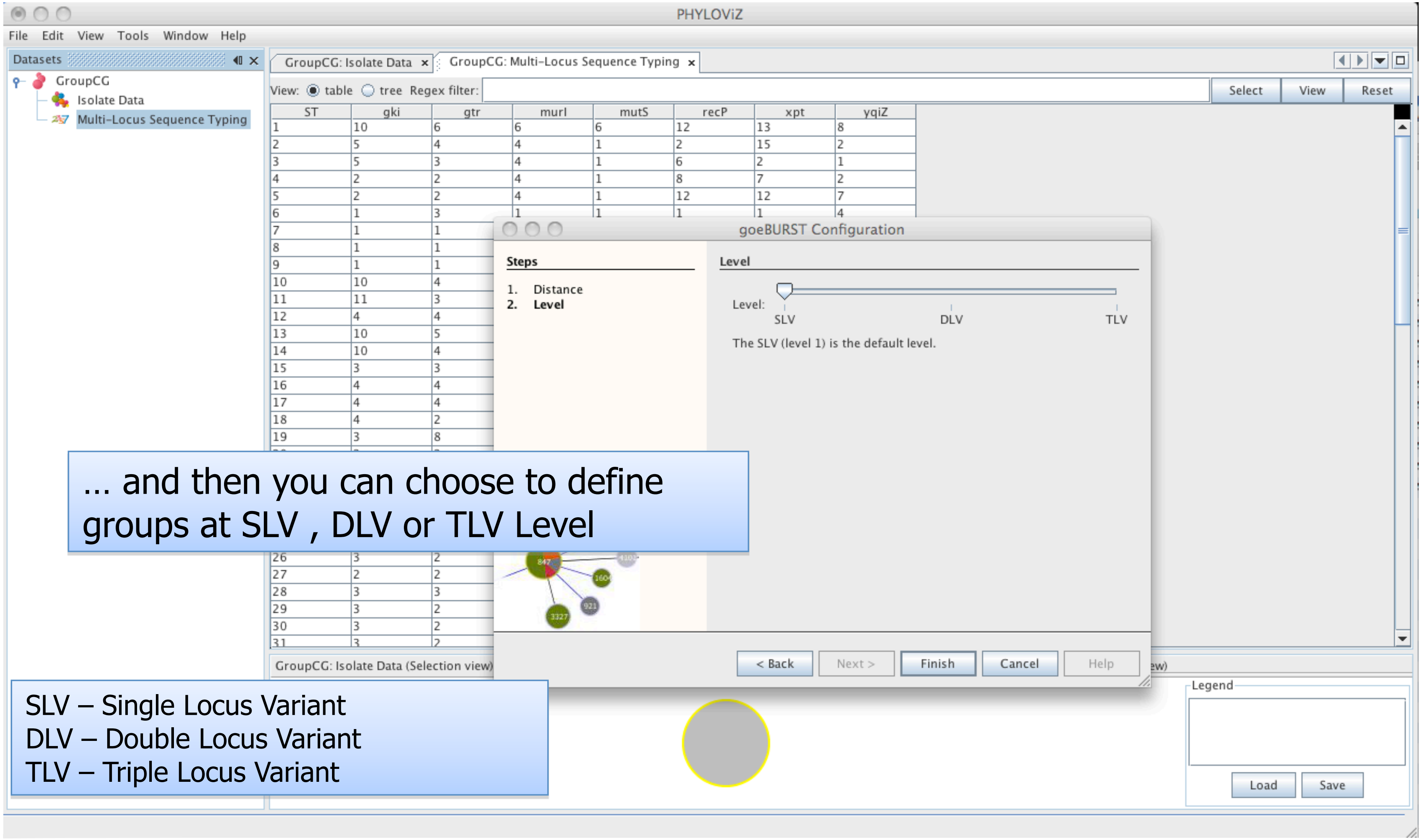
1. **Distance**
2. Level

Distance

Distance: eBURST Distance

This list contains only distances compatible with the current dataset.

If you choose the goeBURST algorithm you then choose the eBURST Distance …



ST	gki	gtr	murl	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1					
8	1	1					
9	1	1					
10	10	4					
11	11	3					
12	4	4					
13	10	5					
14	10	4					
15	3	3					
16	4	4					
17	4	4					
18	4	2					
19	3	8					
26	3	2					
27	2	2					
28	3	3					
29	3	2					
30	3	2					
31	3	2					

Steps

1. Distance
2. **Level**

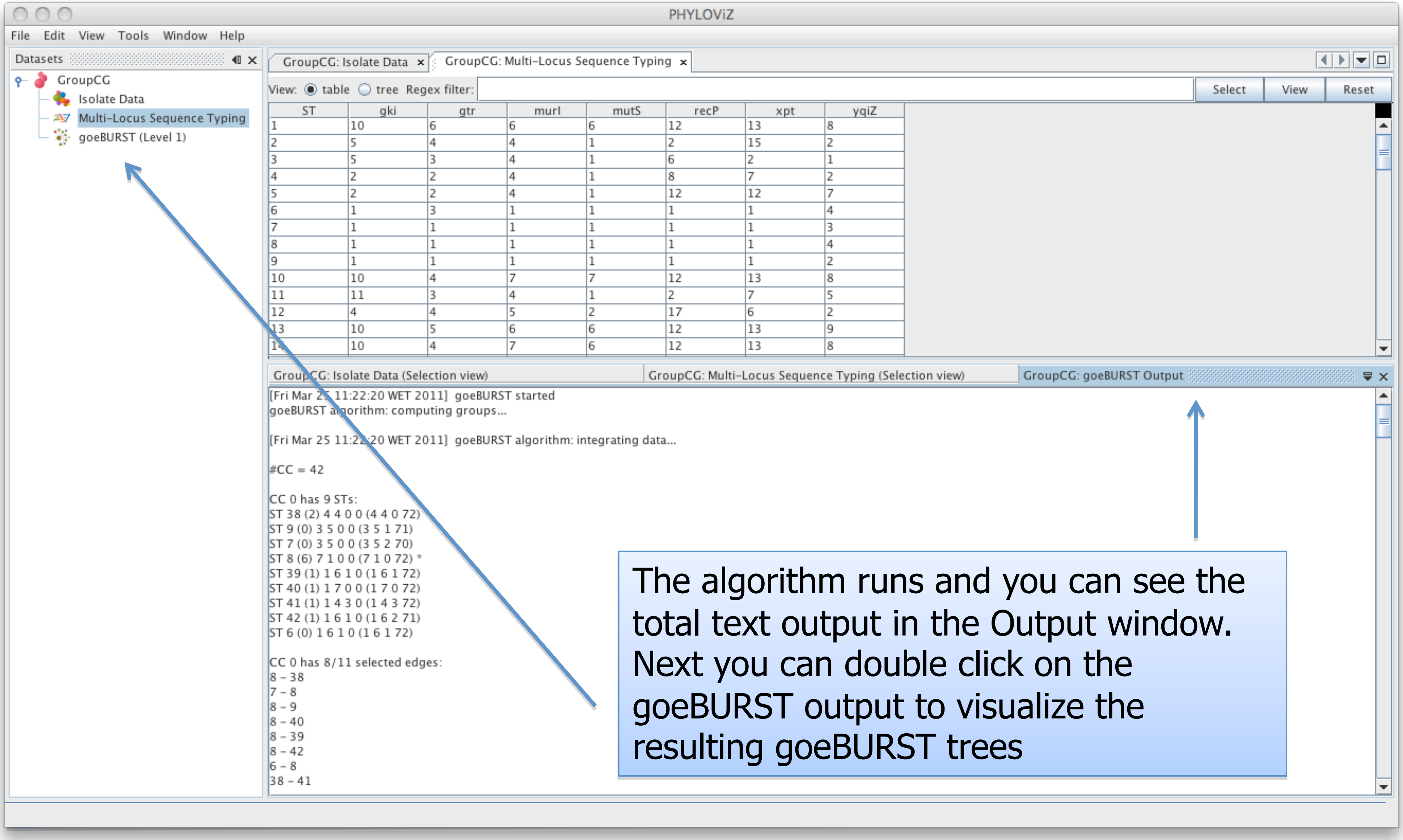
Level

Level: SLV DLV TLV

The SLV (level 1) is the default level.

… and then you can choose to define groups at SLV , DLV or TLV Level

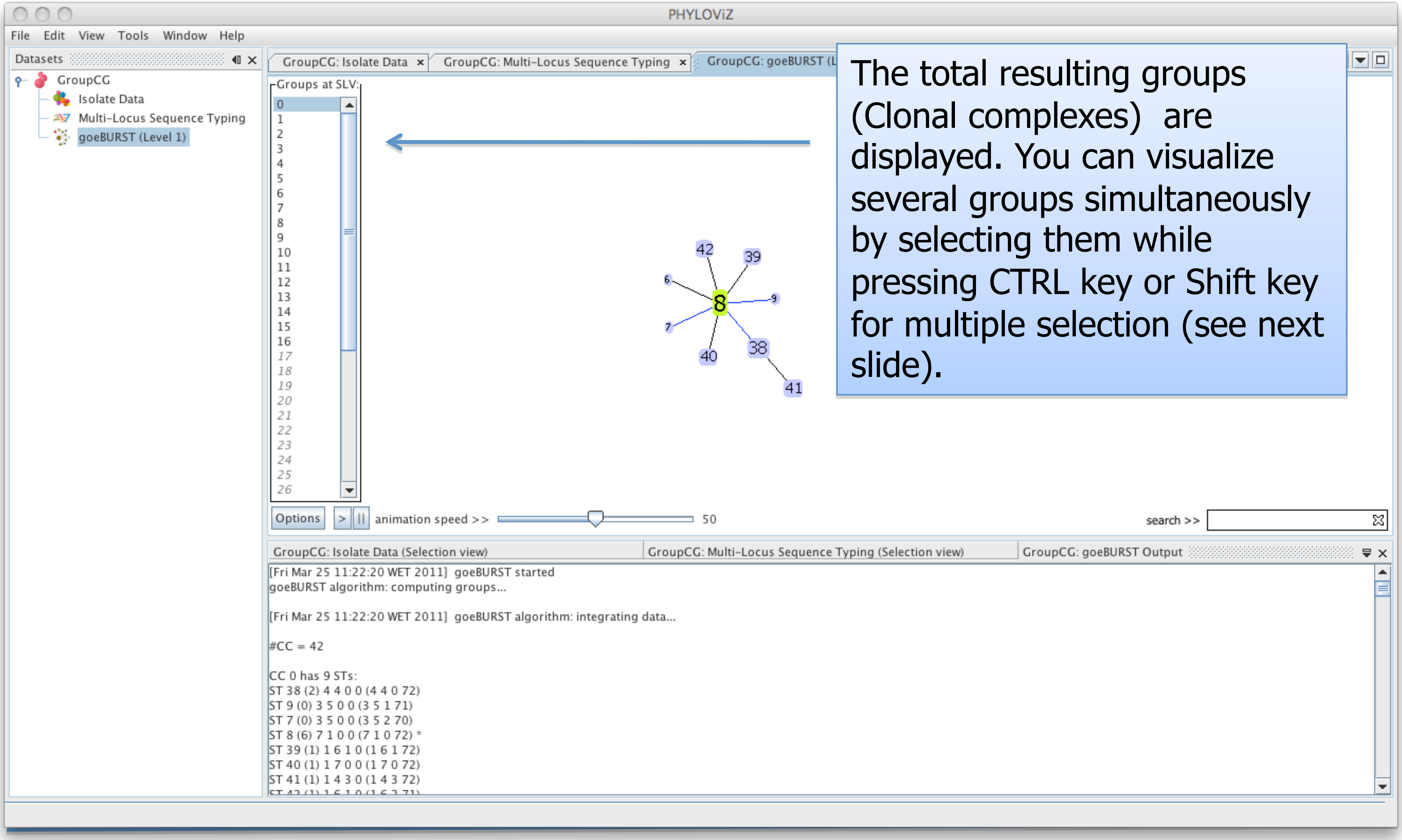
SLV – Single Locus Variant
DLV – Double Locus Variant
TLV – Triple Locus Variant



ST	gki	gtr	murI	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	13	9
14	10	4	7	6	12	13	8

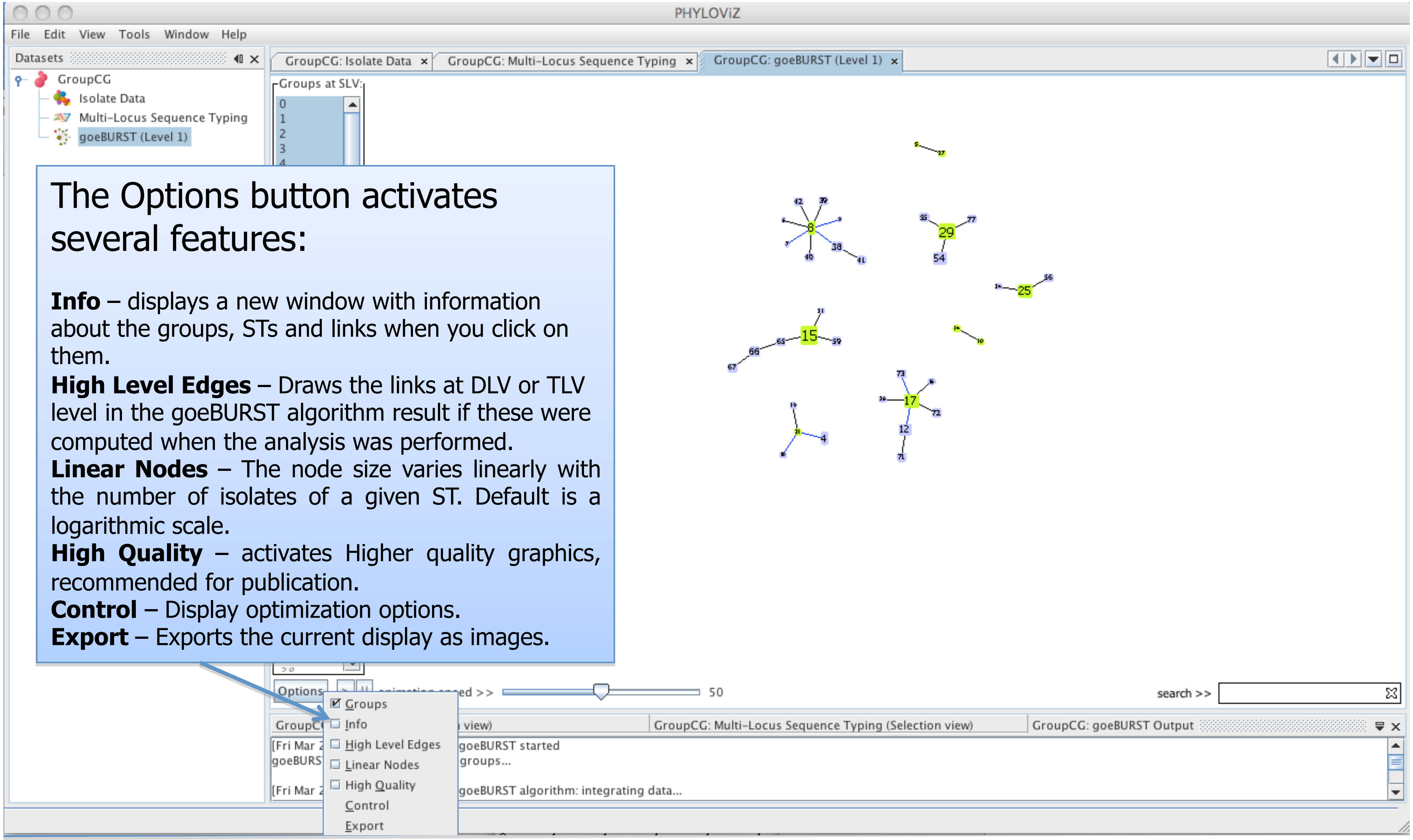
```
[Fri Mar 25 11:22:20 WET 2011]  goeBURST started
goeBURST algorithm: computing groups...

[Fri Mar 25 11:22:20 WET 2011]  goeBURST algorithm: integrating data...

#CC = 42

CC 0 has 9 STs:
ST 38 (2) 4 4 0 0 (4 4 0 72)
ST 9 (0) 3 5 0 0 (3 5 1 71)
ST 7 (0) 3 5 0 0 (3 5 2 70)
ST 8 (6) 7 1 0 0 (7 1 0 72) *
ST 39 (1) 1 6 1 0 (1 6 1 72)
ST 40 (1) 1 7 0 0 (1 7 0 72)
ST 41 (1) 1 4 3 0 (1 4 3 72)
ST 42 (1) 1 6 1 0 (1 6 2 71)
ST 6 (0) 1 6 1 0 (1 6 1 72)

CC 0 has 8/11 selected edges:
8 - 38
7 - 8
8 - 9
8 - 40
8 - 39
8 - 42
6 - 8
38 - 41
```

The algorithm runs and you can see the total text output in the Output window. Next you can double click on the goeBURST output to visualize the resulting goeBURST trees



The total resulting groups (Clonal complexes) are displayed. You can visualize several groups simultaneously by selecting them while pressing CTRL key or Shift key for multiple selection (see next slide).

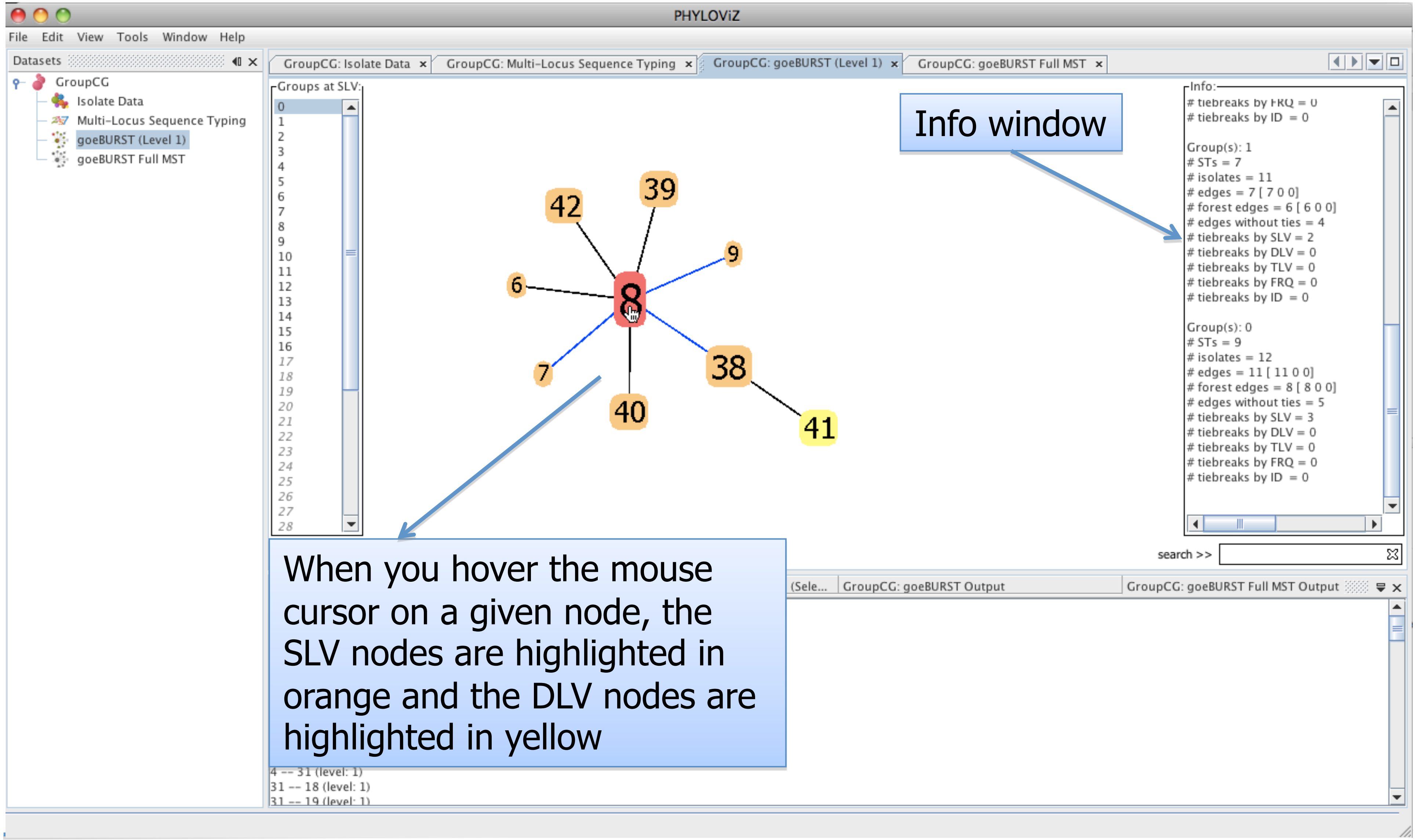
```
[Fri Mar 25 11:22:20 WET 2011]  goeBURST started
goeBURST algorithm: computing groups...

[Fri Mar 25 11:22:20 WET 2011]  goeBURST algorithm: integrating data...

#CC = 42

CC 0 has 9 STs:
ST 38 (2) 4 4 0 0 (4 4 0 72)
ST 9 (0) 3 5 0 0 (3 5 1 71)
ST 7 (0) 3 5 0 0 (3 5 2 70)
ST 8 (6) 7 1 0 0 (7 1 0 72) *
ST 39 (1) 1 6 1 0 (1 6 1 72)
ST 40 (1) 1 7 0 0 (1 7 0 72)
ST 41 (1) 1 4 3 0 (1 4 3 72)
```



```
[Fri Mar 2 goeBURST started
goeBURS groups...
[Fri Mar 2 goeBURST algorithm: integrating data...
```

The Options button activates several features:

Info – displays a new window with information about the groups, STs and links when you click on them.
High Level Edges – Draws the links at DLV or TLV level in the goeBURST algorithm result if these were computed when the analysis was performed.
Linear Nodes – The node size varies linearly with the number of isolates of a given ST. Default is a logarithmic scale.
High Quality – activates Higher quality graphics, recommended for publication.
Control – Display optimization options.
Export – Exports the current display as images.

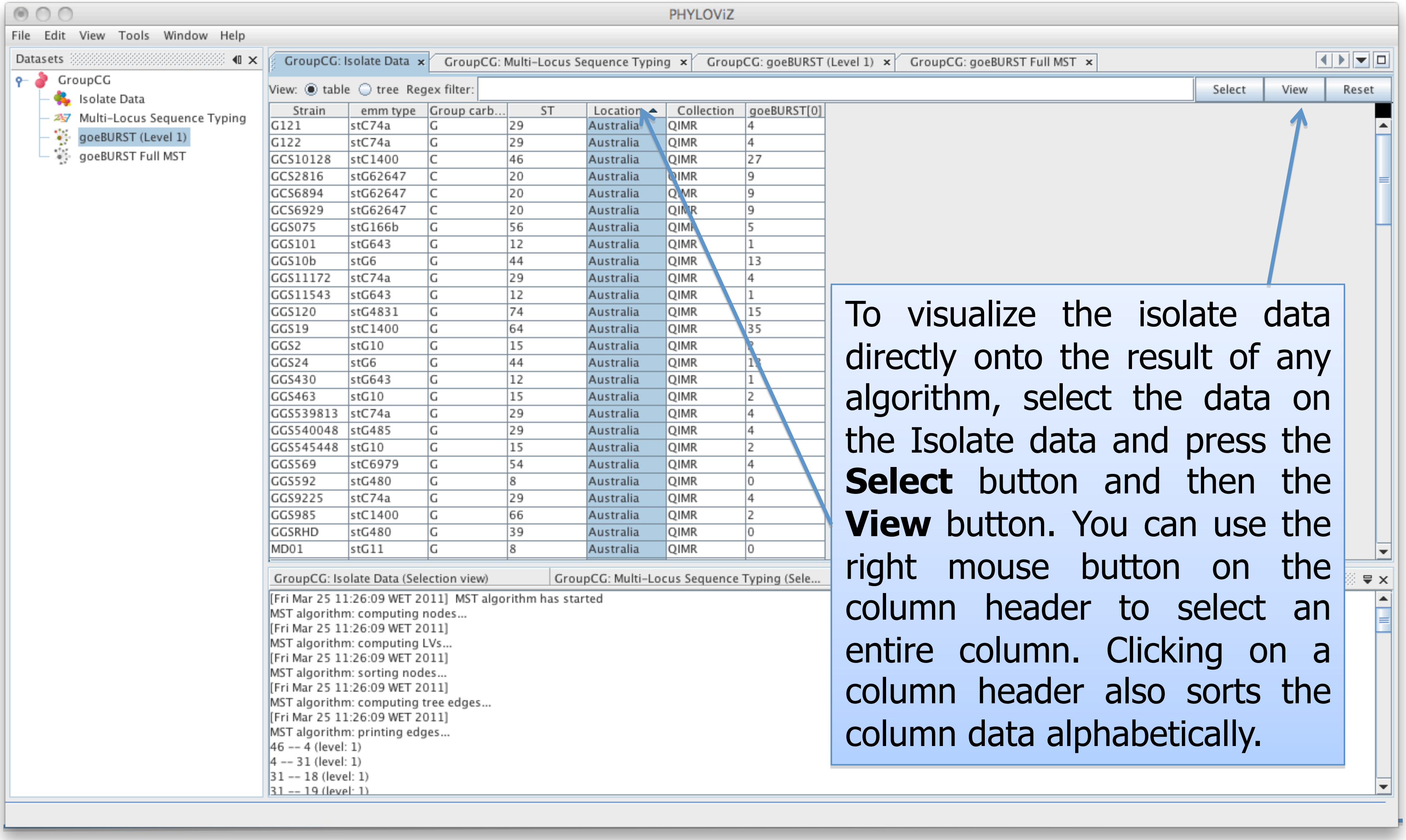


Info window

```
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 1
# STs = 7
# isolates = 11
# edges = 7 [ 7 0 0]
# forest edges = 6 [ 6 0 0]
# edges without ties = 4
# tiebreaks by SLV = 2
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 0
# STs = 9
# isolates = 12
# edges = 11 [ 11 0 0]
# forest edges = 8 [ 8 0 0]
# edges without ties = 5
# tiebreaks by SLV = 3
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0
```

When you hover the mouse cursor on a given node, the SLV nodes are highlighted in orange and the DLV nodes are highlighted in yellow

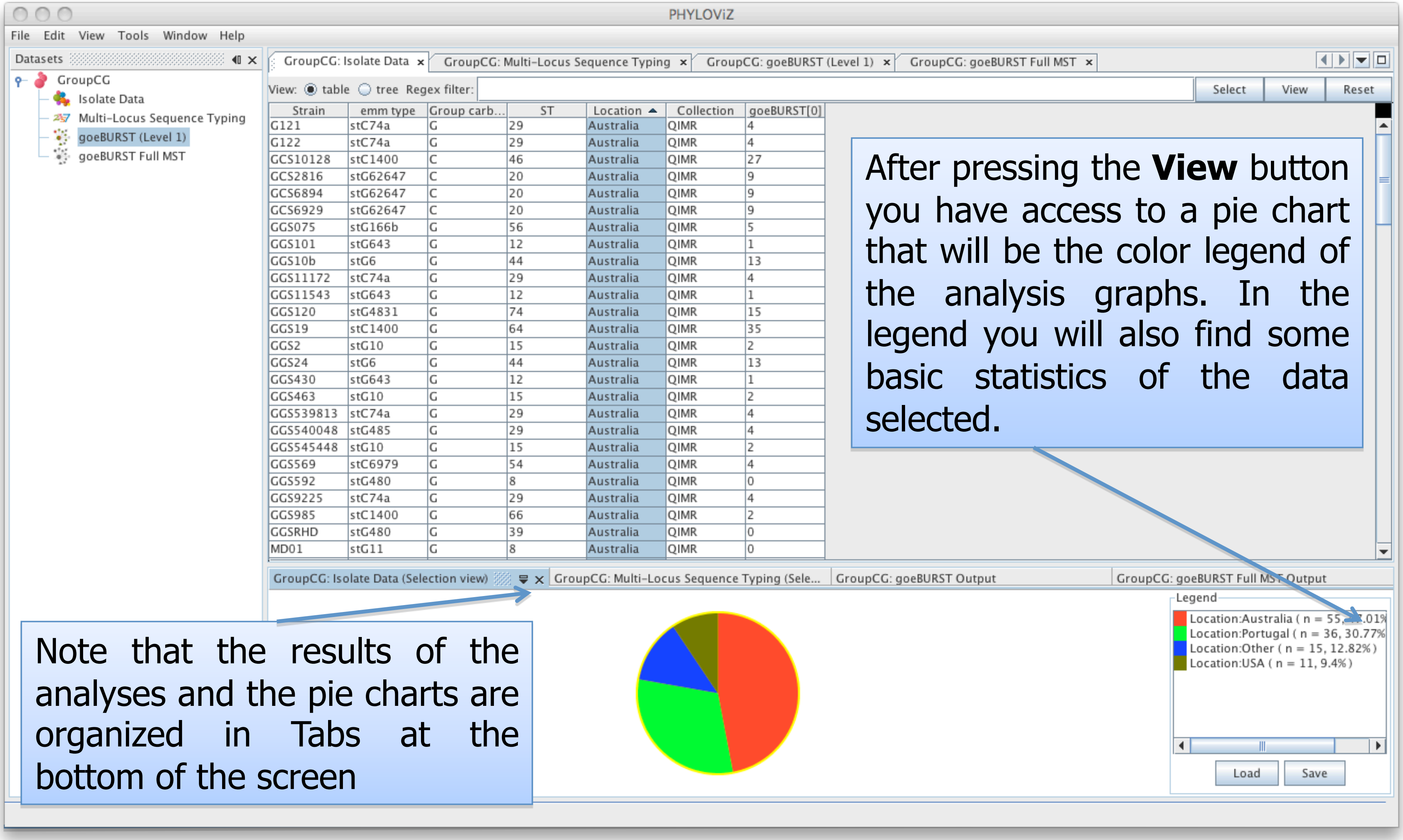
```
4 -- 31 (level: 1)
31 -- 18 (level: 1)
31 -- 19 (level: 1)
```



Strain	emm type	Group carb...	ST	Location	Collection	goeBURST[0]
G121	stC74a	G	29	Australia	QIMR	4
G122	stC74a	G	29	Australia	QIMR	4
GCS10128	stC1400	C	46	Australia	QIMR	27
GCS2816	stG62647	C	20	Australia	QIMR	9
GCS6894	stG62647	C	20	Australia	QIMR	9
GCS6929	stG62647	C	20	Australia	QIMR	9
GGS075	stG166b	G	56	Australia	QIMR	5
GGS101	stG643	G	12	Australia	QIMR	1
GGS10b	stG6	G	44	Australia	QIMR	13
GGS11172	stC74a	G	29	Australia	QIMR	4
GGS11543	stG643	G	12	Australia	QIMR	1
GGS120	stG4831	G	74	Australia	QIMR	15
GGS19	stC1400	G	64	Australia	QIMR	35
GGS2	stG10	G	15	Australia	QIMR	2
GGS24	stG6	G	44	Australia	QIMR	13
GGS430	stG643	G	12	Australia	QIMR	1
GGS463	stG10	G	15	Australia	QIMR	2
GGS539813	stC74a	G	29	Australia	QIMR	4
GGS540048	stG485	G	29	Australia	QIMR	4
GGS545448	stG10	G	15	Australia	QIMR	2
GGS569	stC6979	G	54	Australia	QIMR	4
GGS592	stG480	G	8	Australia	QIMR	0
GGS9225	stC74a	G	29	Australia	QIMR	4
GGS985	stC1400	G	66	Australia	QIMR	2
GGSRHD	stG480	G	39	Australia	QIMR	0
MD01	stG11	G	8	Australia	QIMR	0

```
[Fri Mar 25 11:26:09 WET 2011]  MST algorithm has started
MST algorithm: computing nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing LVs...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: sorting nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing tree edges...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: printing edges...
46 -- 4 (level: 1)
4 -- 31 (level: 1)
31 -- 18 (level: 1)
31 -- 19 (level: 1)
```

To visualize the isolate data directly onto the result of any algorithm, select the data on the Isolate data and press the **Select** button and then the **View** button. You can use the right mouse button on the column header to select an entire column. Clicking on a column header also sorts the column data alphabetically.

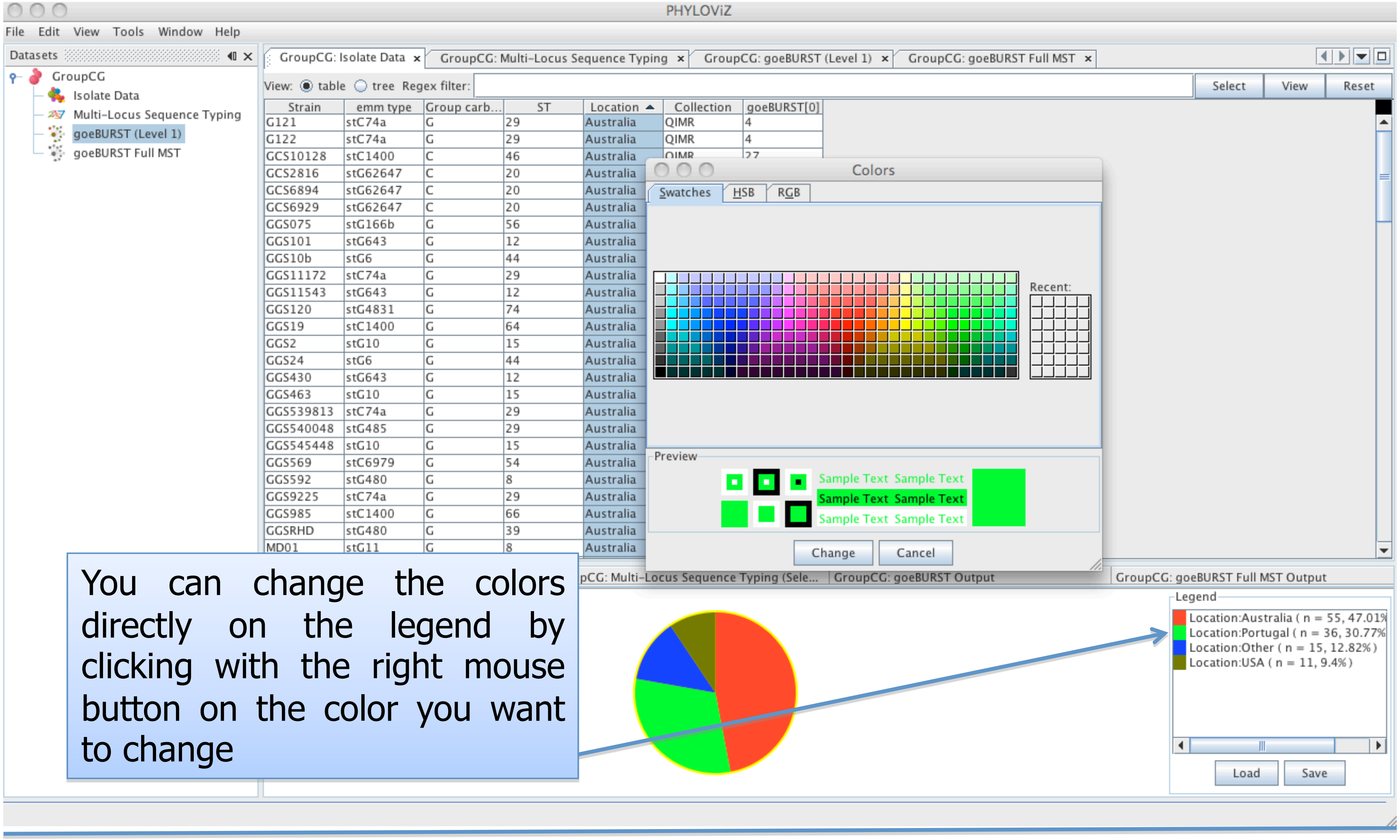


Strain	emm type	Group carb...	ST	Location ▲	Collection	goeBURST[0]
G121	stC74a	G	29	Australia	QIMR	4
G122	stC74a	G	29	Australia	QIMR	4
GCS10128	stC1400	C	46	Australia	QIMR	27
GCS2816	stG62647	C	20	Australia	QIMR	9
GCS6894	stG62647	C	20	Australia	QIMR	9
GCS6929	stG62647	C	20	Australia	QIMR	9
GGS075	stG166b	G	56	Australia	QIMR	5
GGS101	stG643	G	12	Australia	QIMR	1
GGS10b	stG6	G	44	Australia	QIMR	13
GGS11172	stC74a	G	29	Australia	QIMR	4
GGS11543	stG643	G	12	Australia	QIMR	1
GGS120	stG4831	G	74	Australia	QIMR	15
GGS19	stC1400	G	64	Australia	QIMR	35
GGS2	stG10	G	15	Australia	QIMR	2
GGS24	stG6	G	44	Australia	QIMR	13
GGS430	stG643	G	12	Australia	QIMR	1
GGS463	stG10	G	15	Australia	QIMR	2
GGS539813	stC74a	G	29	Australia	QIMR	4
GGS540048	stG485	G	29	Australia	QIMR	4
GGS545448	stG10	G	15	Australia	QIMR	2
GGS569	stC6979	G	54	Australia	QIMR	4
GGS592	stG480	G	8	Australia	QIMR	0
GGS9225	stC74a	G	29	Australia	QIMR	4
GGS985	stC1400	G	66	Australia	QIMR	2
GGSRHD	stG480	G	39	Australia	QIMR	0
MD01	stG11	G	8	Australia	QIMR	0

🟥 Location:Australia (n = 55, [illegible].01%
🟩 Location:Portugal (n = 36, 30.77%
🟦 Location:Other (n = 15, 12.82%)
■ Location:USA (n = 11, 9.4%)

After pressing the **View** button you have access to a pie chart that will be the color legend of the analysis graphs. In the legend you will also find some basic statistics of the data selected.

Note that the results of the analyses and the pie charts are organized in Tabs at the bottom of the screen



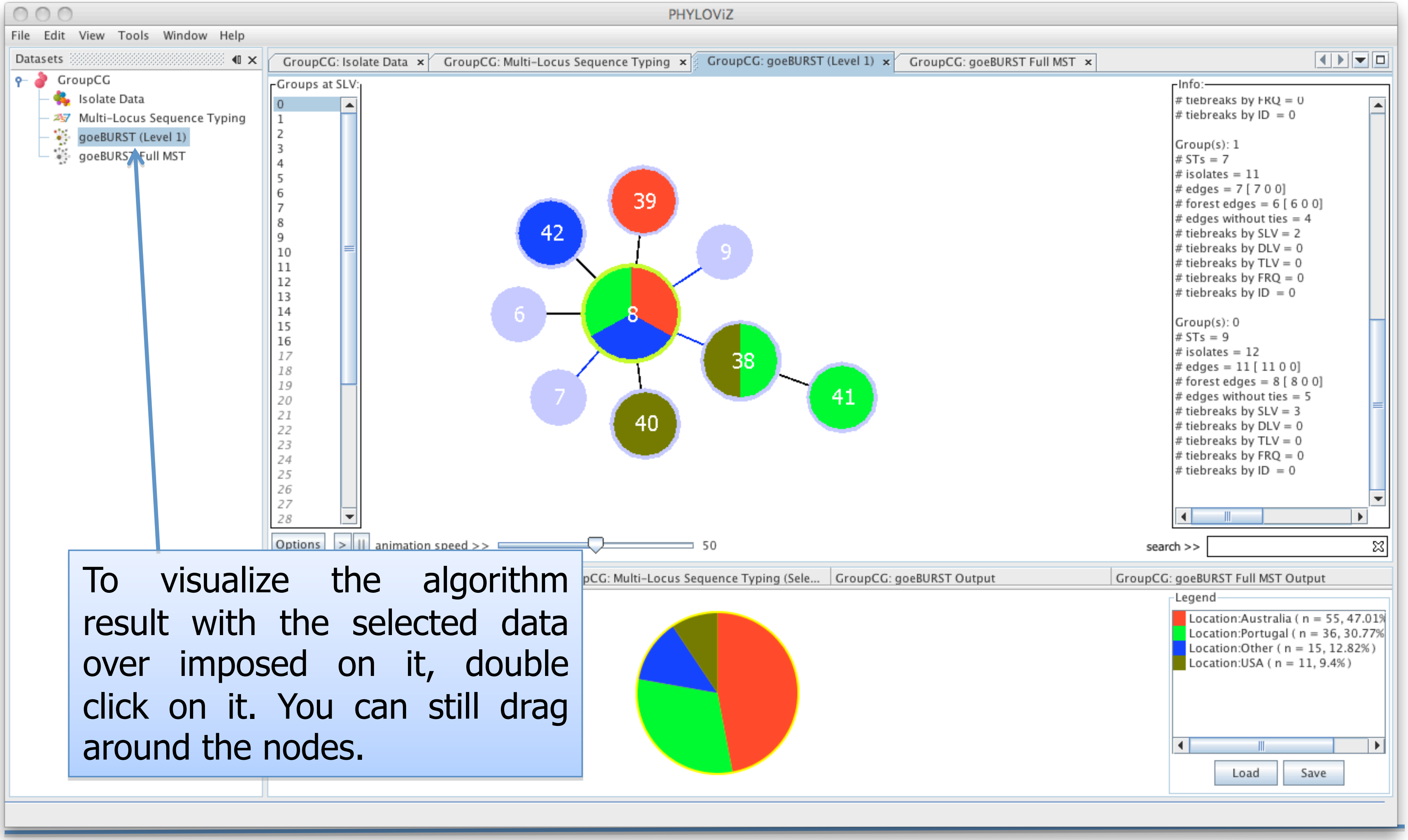
View: ● table ○ tree Regex filter:

Strain	emm type	Group carb...	ST	Location ▲	Collection	goeBURST[0]
G121	stC74a	G	29	Australia	QIMR	4
G122	stC74a	G	29	Australia	QIMR	4
GCS10128	stC1400	C	46	Australia	QIMR	27
GCS2816	stG62647	C	20	Australia		
GCS6894	stG62647	C	20	Australia		
GCS6929	stG62647	C	20	Australia		
GGS075	stG166b	G	56	Australia		
GGS101	stG643	G	12	Australia		
GGS10b	stG6	G	44	Australia		
GGS11172	stC74a	G	29	Australia		
GGS11543	stG643	G	12	Australia		
GGS120	stG4831	G	74	Australia		
GGS19	stC1400	G	64	Australia		
GGS2	stG10	G	15	Australia		
GGS24	stG6	G	44	Australia		
GGS430	stG643	G	12	Australia		
GGS463	stG10	G	15	Australia		
GGS539813	stC74a	G	29	Australia		
GGS540048	stG485	G	29	Australia		
GGS545448	stG10	G	15	Australia		
GGS569	stC6979	G	54	Australia		
GGS592	stG480	G	8	Australia		
GGS9225	stC74a	G	29	Australia		
GGS985	stC1400	G	66	Australia		
GGSRHD	stG480	G	39	Australia		
MD01	stG11	G	8	Australia		

Legend

🟥 Location:Australia (n = 55, 47.01%
🟩 Location:Portugal (n = 36, 30.77%
🟦 Location:Other (n = 15, 12.82%)
🟫 Location:USA (n = 11, 9.4%)

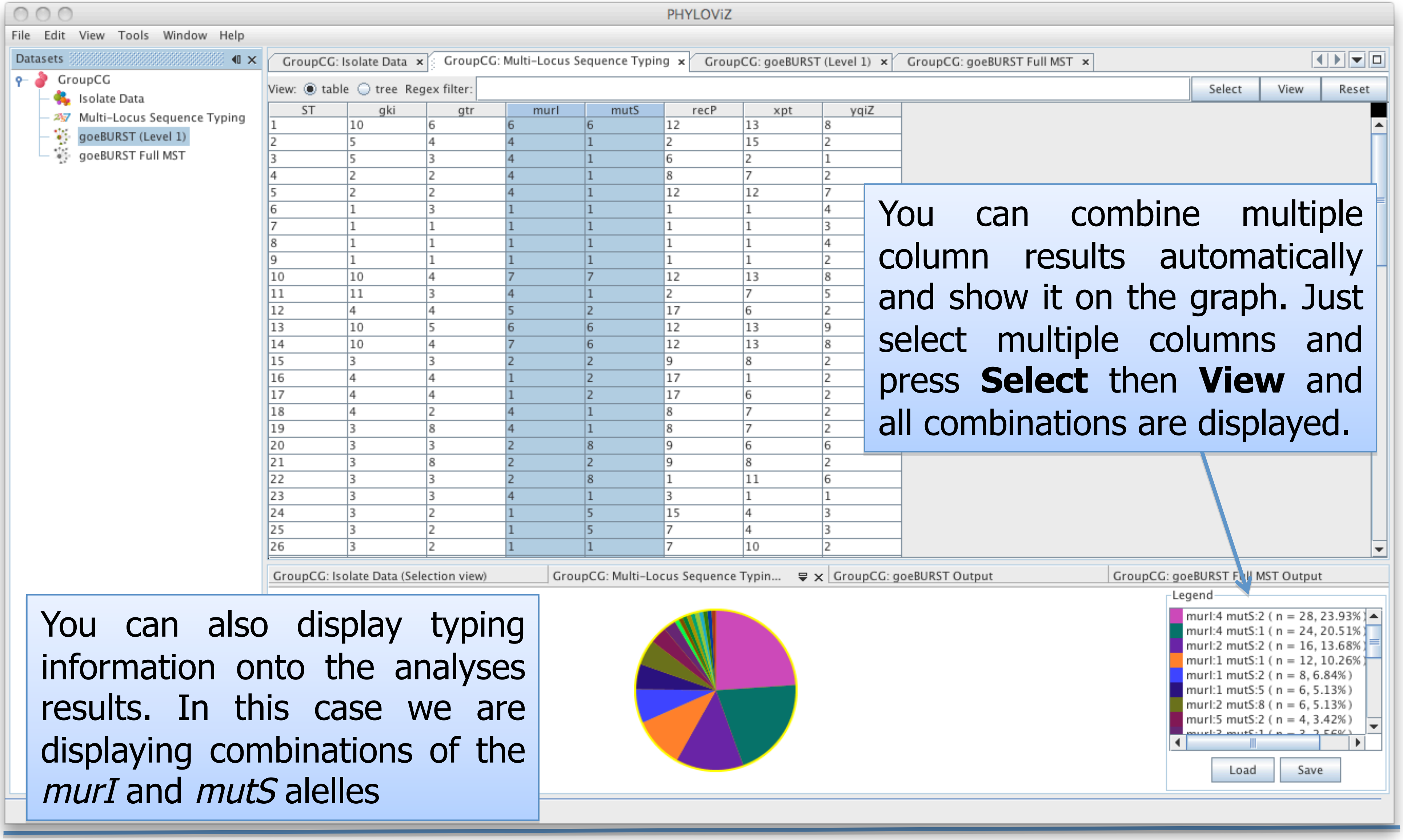
You can change the colors directly on the legend by clicking with the right mouse button on the color you want to change



```
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 1
# STs = 7
# isolates = 11
# edges = 7 [ 7 0 0]
# forest edges = 6 [ 6 0 0]
# edges without ties = 4
# tiebreaks by SLV = 2
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 0
# STs = 9
# isolates = 12
# edges = 11 [ 11 0 0]
# forest edges = 8 [ 8 0 0]
# edges without ties = 5
# tiebreaks by SLV = 3
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0
```

🟥 Location:Australia (n = 55, 47.01%
🟩 Location:Portugal (n = 36, 30.77%
🟦 Location:Other (n = 15, 12.82%)
🟫 Location:USA (n = 11, 9.4%)

To visualize the algorithm result with the selected data over imposed on it, double click on it. You can still drag around the nodes.



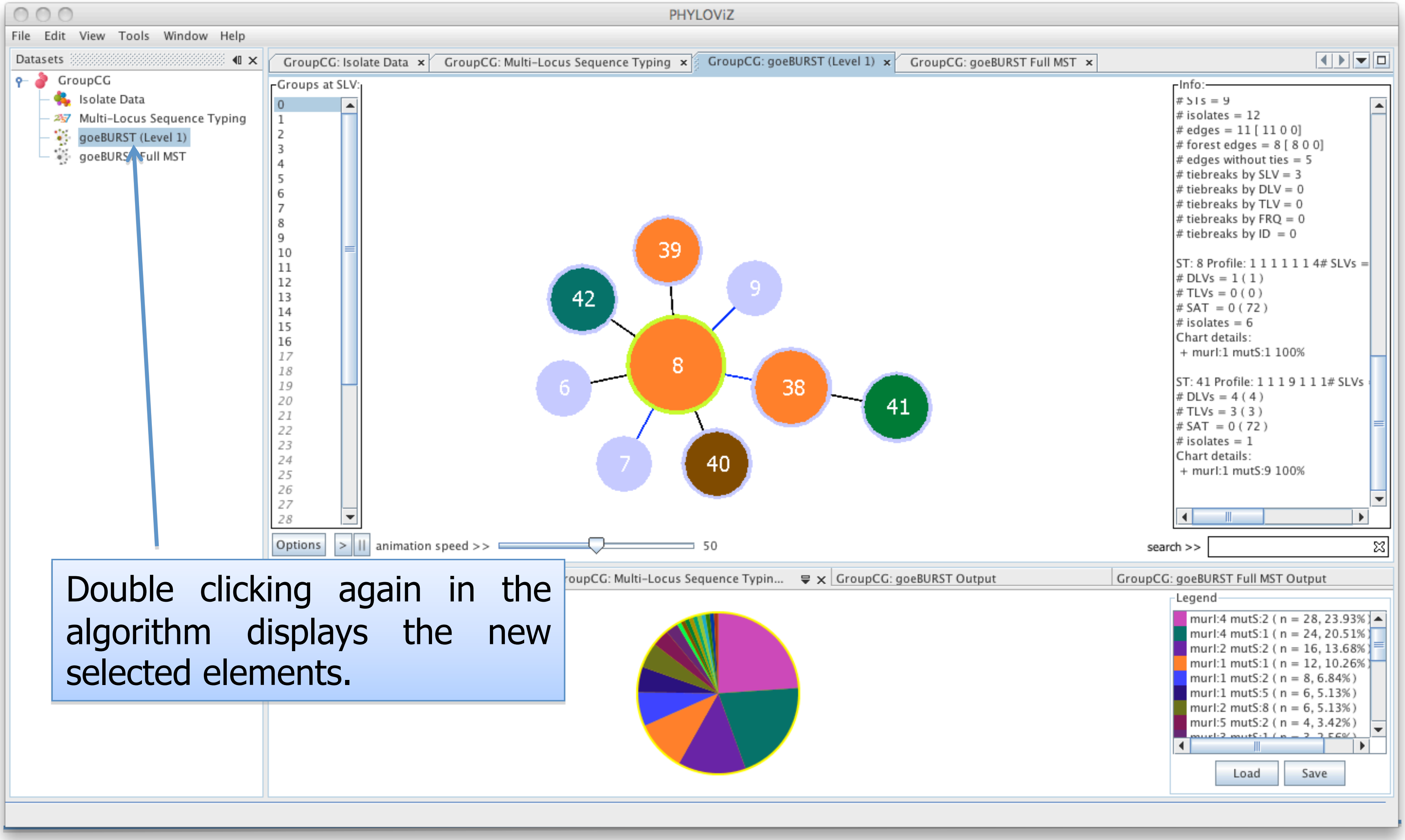
ST	gki	gtr	murI	mutS	recP	xpt	yqiZ
1	10	6	6	6	12	13	8
2	5	4	4	1	2	15	2
3	5	3	4	1	6	2	1
4	2	2	4	1	8	7	2
5	2	2	4	1	12	12	7
6	1	3	1	1	1	1	4
7	1	1	1	1	1	1	3
8	1	1	1	1	1	1	4
9	1	1	1	1	1	1	2
10	10	4	7	7	12	13	8
11	11	3	4	1	2	7	5
12	4	4	5	2	17	6	2
13	10	5	6	6	12	13	9
14	10	4	7	6	12	13	8
15	3	3	2	2	9	8	2
16	4	4	1	2	17	1	2
17	4	4	1	2	17	6	2
18	4	2	4	1	8	7	2
19	3	8	4	1	8	7	2
20	3	3	2	8	9	6	6
21	3	8	2	2	9	8	2
22	3	3	2	8	1	11	6
23	3	3	4	1	3	1	1
24	3	2	1	5	15	4	3
25	3	2	1	5	7	4	3
26	3	2	1	1	7	10	2

Legend

- murI:4 mutS:2 (n = 28, 23.93%)
- murI:4 mutS:1 (n = 24, 20.51%)
- murI:2 mutS:2 (n = 16, 13.68%)
- murI:1 mutS:1 (n = 12, 10.26%)
- murI:1 mutS:2 (n = 8, 6.84%)
- murI:1 mutS:5 (n = 6, 5.13%)
- murI:2 mutS:8 (n = 6, 5.13%)
- murI:5 mutS:2 (n = 4, 3.42%)

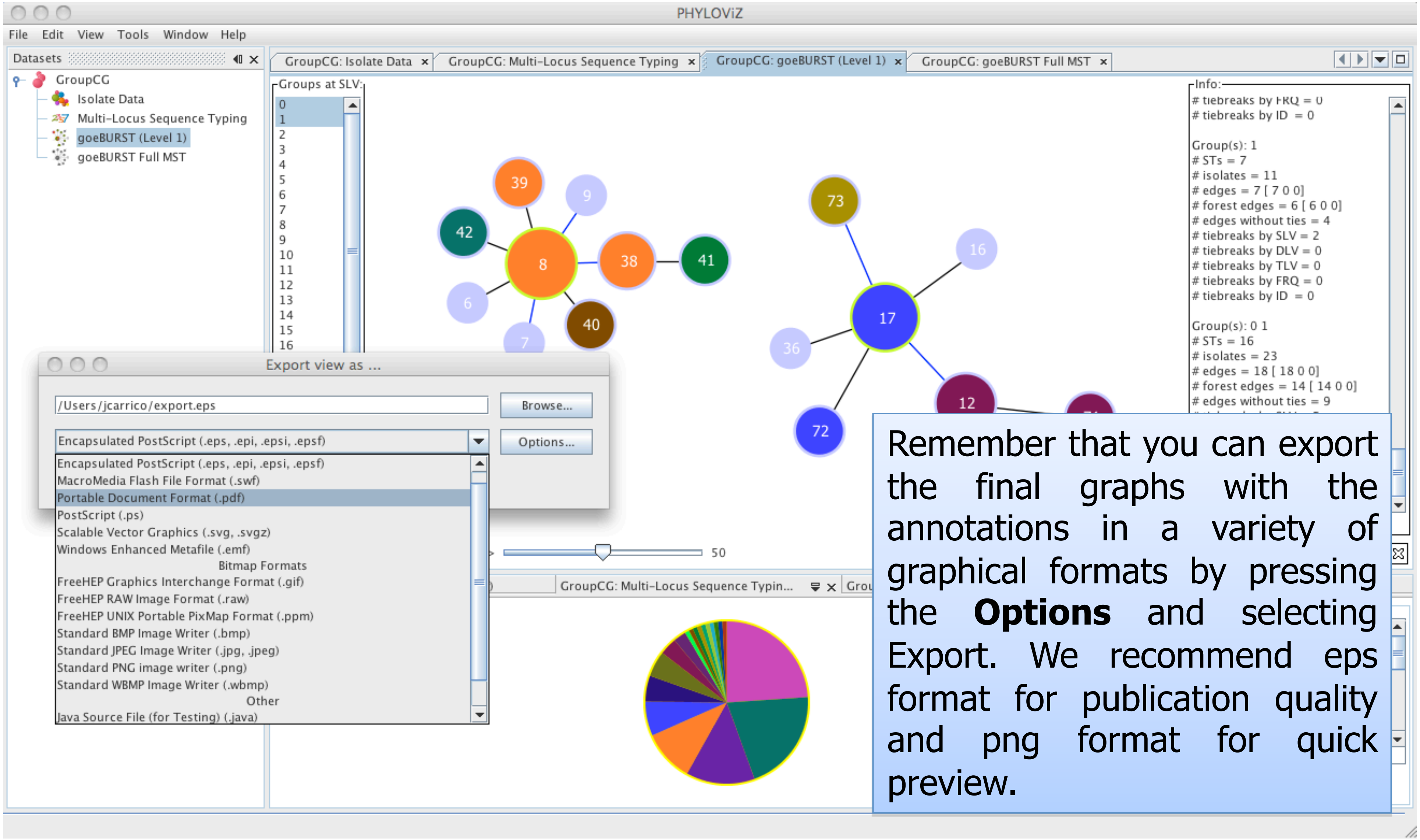
You can combine multiple column results automatically and show it on the graph. Just select multiple columns and press **Select** then **View** and all combinations are displayed.

You can also display typing information onto the analyses results. In this case we are displaying combinations of the *murI* and *mutS* alelles



```
# STs = 9
# isolates = 12
# edges = 11 [ 11 0 0]
# forest edges = 8 [ 8 0 0]
# edges without ties = 5
# tiebreaks by SLV = 3
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

ST: 8 Profile: 1 1 1 1 1 1 4# SLVs =
# DLVs = 1 ( 1 )
# TLVs = 0 ( 0 )
# SAT  = 0 ( 72 )
# isolates = 6
Chart details:
 + murI:1 mutS:1 100%

ST: 41 Profile: 1 1 1 9 1 1 1# SLVs
# DLVs = 4 ( 4 )
# TLVs = 3 ( 3 )
# SAT  = 0 ( 72 )
# isolates = 1
Chart details:
 + murI:1 mutS:9 100%
```

Double clicking again in the algorithm displays the new selected elements.



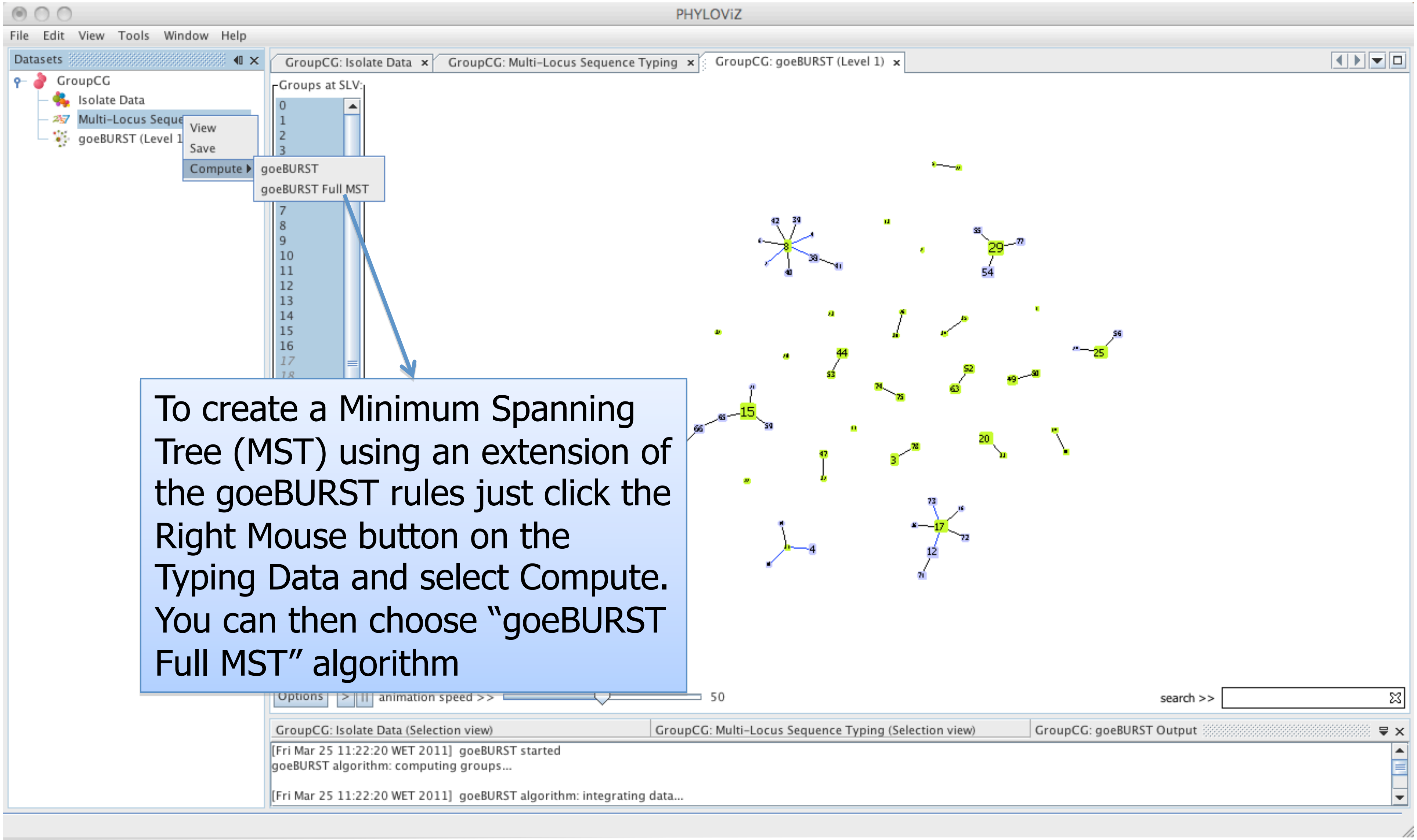
/Users/jcarrico/export.eps

Encapsulated PostScript (.eps, .epi, .epsi, .epsf)

Info:

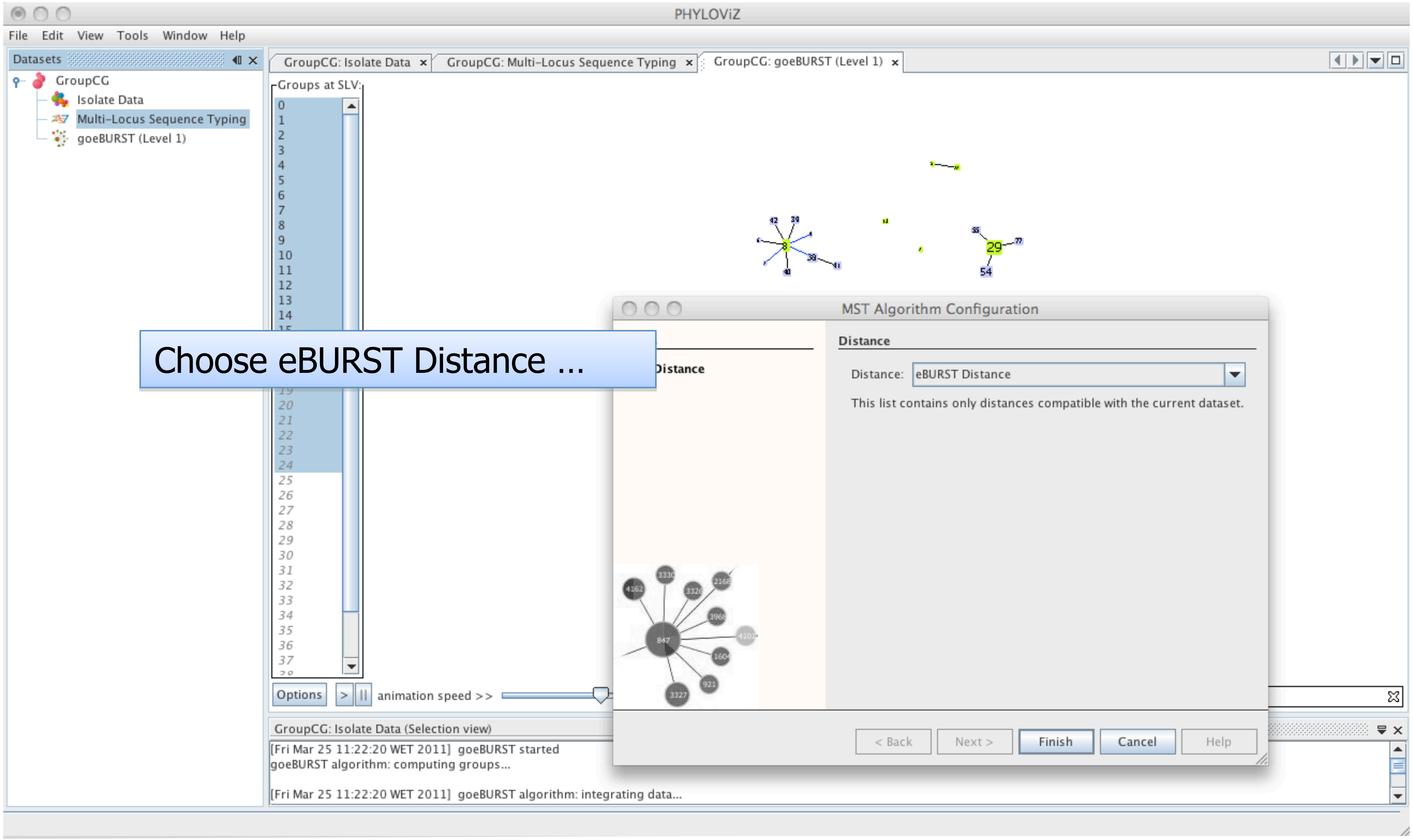
```
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 1
# STs = 7
# isolates = 11
# edges = 7 [ 7 0 0]
# forest edges = 6 [ 6 0 0]
# edges without ties = 4
# tiebreaks by SLV = 2
# tiebreaks by DLV = 0
# tiebreaks by TLV = 0
# tiebreaks by FRQ = 0
# tiebreaks by ID  = 0

Group(s): 0 1
# STs = 16
# isolates = 23
# edges = 18 [ 18 0 0]
# forest edges = 14 [ 14 0 0]
# edges without ties = 9
```

Remember that you can export the final graphs with the annotations in a variety of graphical formats by pressing the **Options** and selecting Export. We recommend eps format for publication quality and png format for quick preview.



```
[Fri Mar 25 11:22:20 WET 2011]  goeBURST started
goeBURST algorithm: computing groups...

[Fri Mar 25 11:22:20 WET 2011]  goeBURST algorithm: integrating data...
```

To create a Minimum Spanning Tree (MST) using an extension of the goeBURST rules just click the Right Mouse button on the Typing Data and select Compute. You can then choose "goeBURST Full MST" algorithm



Choose eBURST Distance …

MST Algorithm Configuration

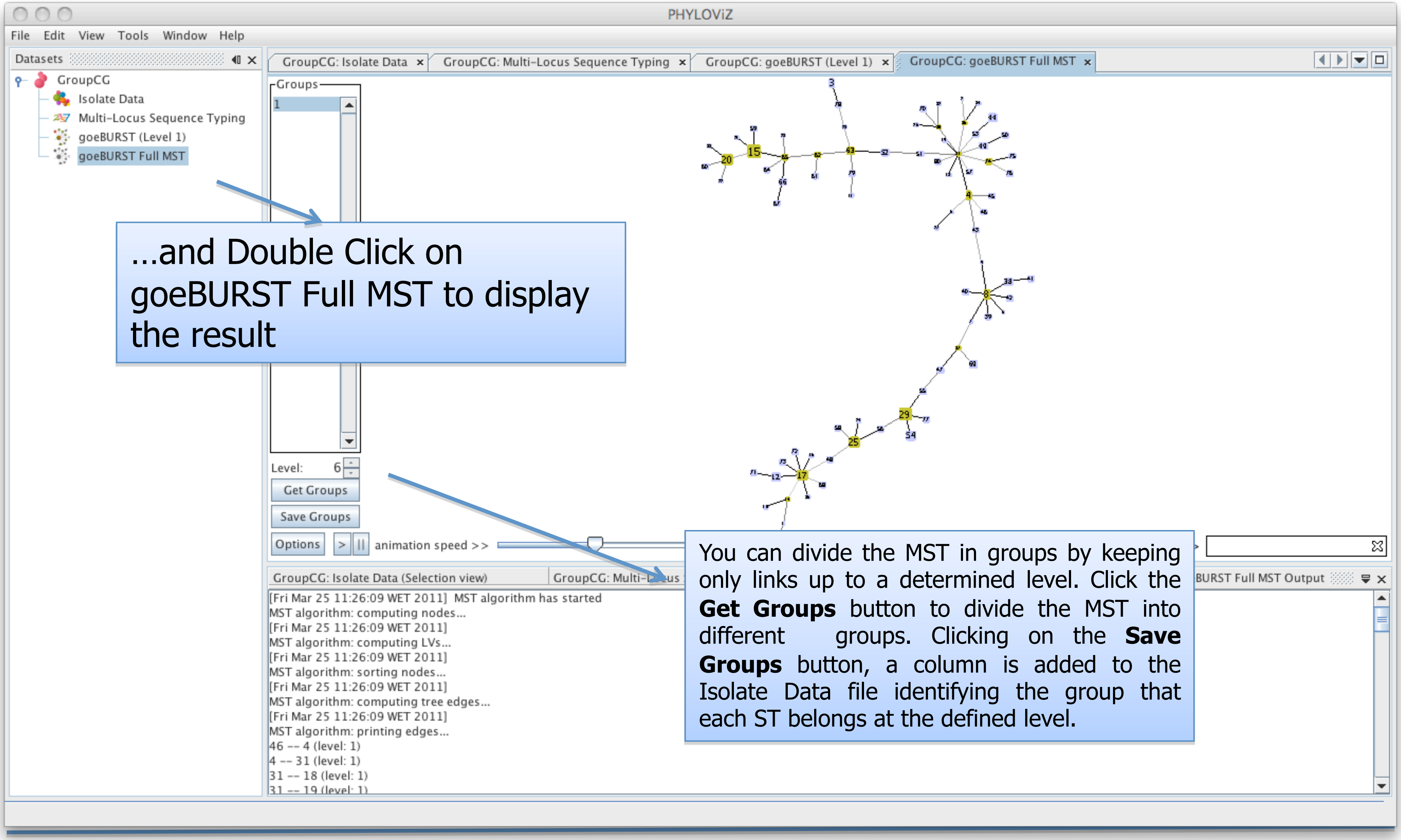
Distance

Distance: eBURST Distance

This list contains only distances compatible with the current dataset.

GroupCG: Isolate Data (Selection view)

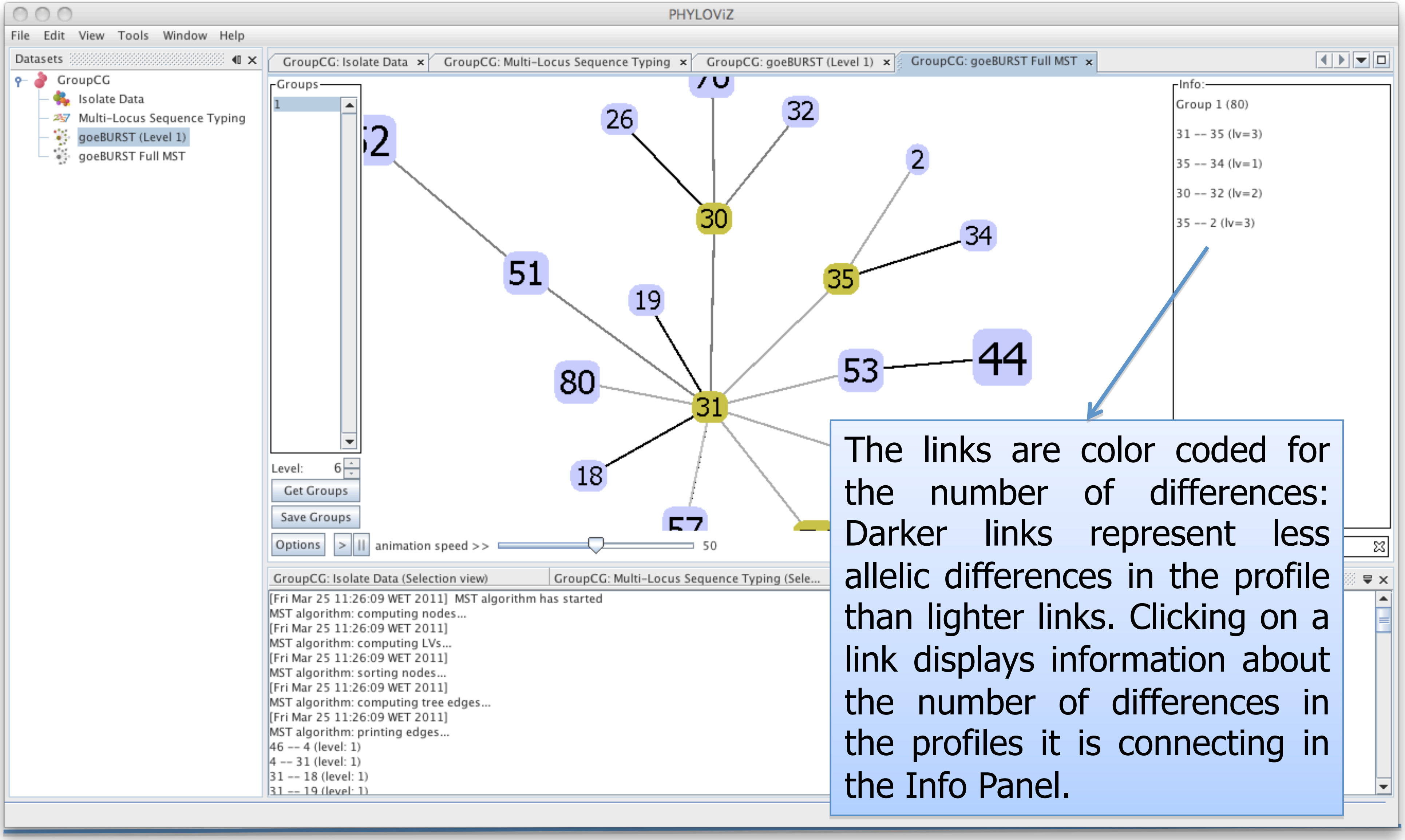
```
[Fri Mar 25 11:22:20 WET 2011]  goeBURST started
goeBURST algorithm: computing groups...

[Fri Mar 25 11:22:20 WET 2011]  goeBURST algorithm: integrating data...
```



```
[Fri Mar 25 11:26:09 WET 2011]  MST algorithm has started
MST algorithm: computing nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing LVs...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: sorting nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing tree edges...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: printing edges...
46 -- 4 (level: 1)
4 -- 31 (level: 1)
31 -- 18 (level: 1)
31 -- 19 (level: 1)
```

…and Double Click on goeBURST Full MST to display the result

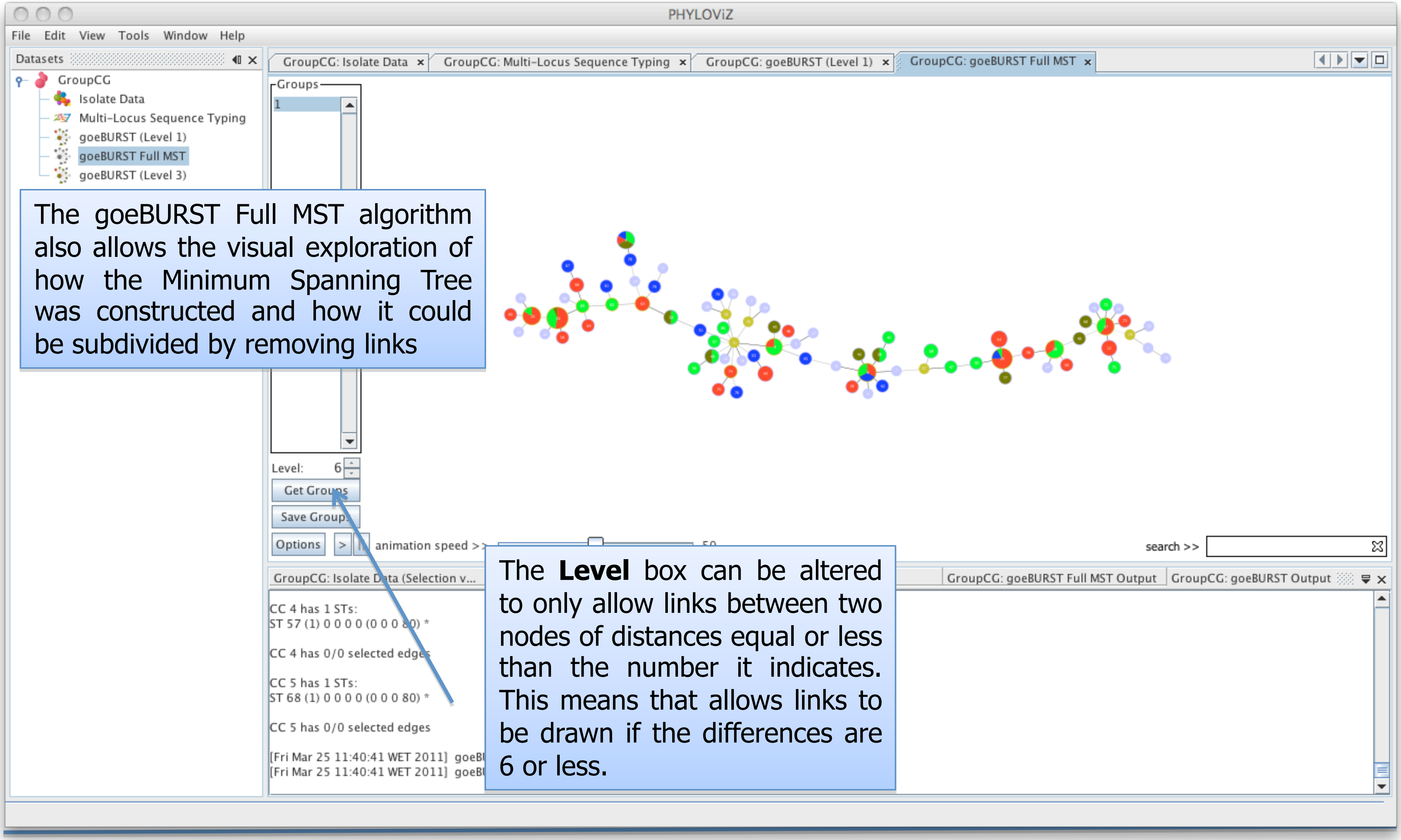
You can divide the MST in groups by keeping only links up to a determined level. Click the **Get Groups** button to divide the MST into different groups. Clicking on the **Save Groups** button, a column is added to the Isolate Data file identifying the group that each ST belongs at the defined level.



```
Group 1 (80)
31 -- 35 (lv=3)
35 -- 34 (lv=1)
30 -- 32 (lv=2)
35 -- 2 (lv=3)
```

```
[Fri Mar 25 11:26:09 WET 2011]  MST algorithm has started
MST algorithm: computing nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing LVs...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: sorting nodes...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: computing tree edges...
[Fri Mar 25 11:26:09 WET 2011]
MST algorithm: printing edges...
46 -- 4 (level: 1)
4 -- 31 (level: 1)
31 -- 18 (level: 1)
31 -- 19 (level: 1)
```

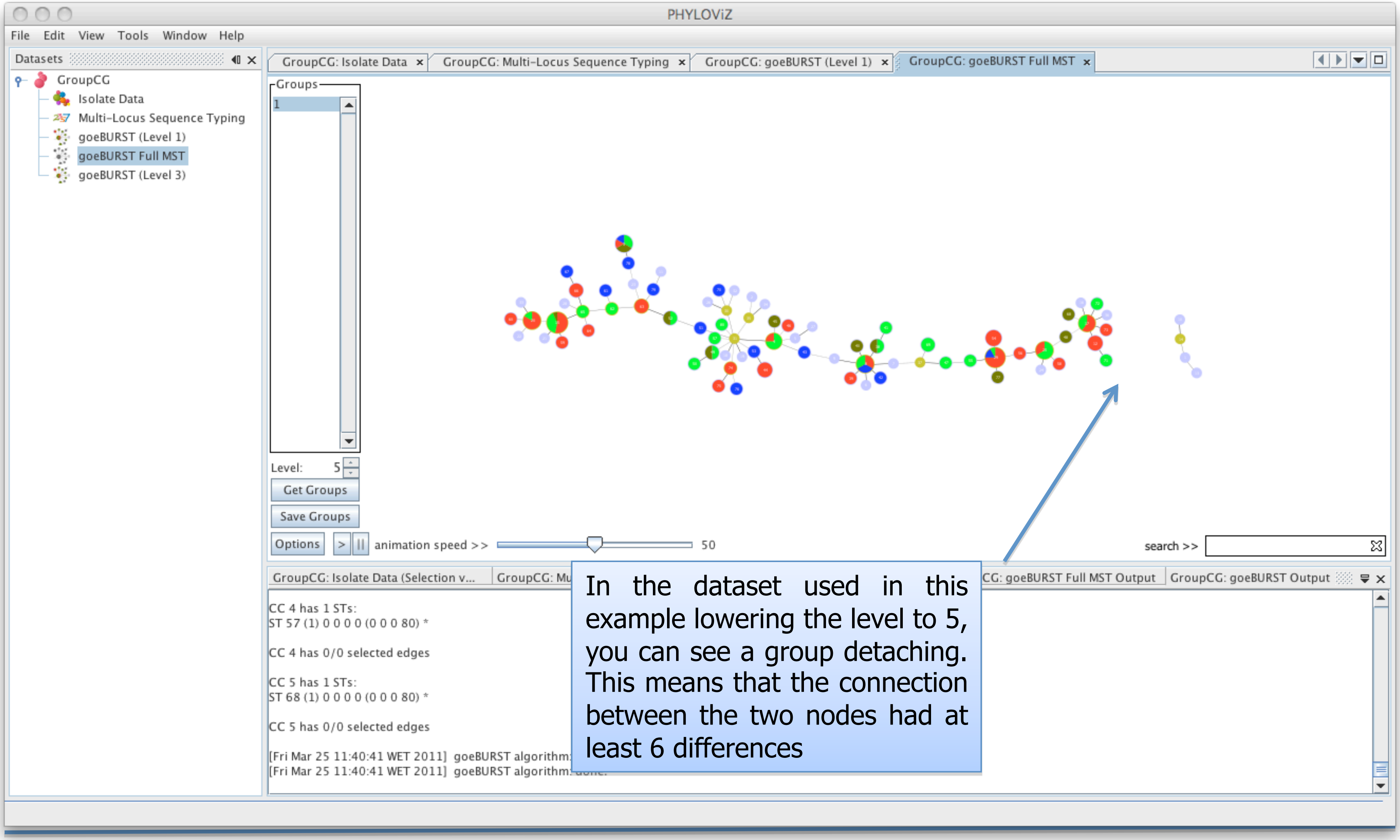
The links are color coded for the number of differences: Darker links represent less allelic differences in the profile than lighter links. Clicking on a link displays information about the number of differences in the profiles it is connecting in the Info Panel.



```
CC 4 has 1 STs:
ST 57 (1) 0 0 0 0 (0 0 0 80) *

CC 4 has 0/0 selected edges

CC 5 has 1 STs:
ST 68 (1) 0 0 0 0 (0 0 0 80) *

CC 5 has 0/0 selected edges

[Fri Mar 25 11:40:41 WET 2011]  goeBU
[Fri Mar 25 11:40:41 WET 2011]  goeBU
```

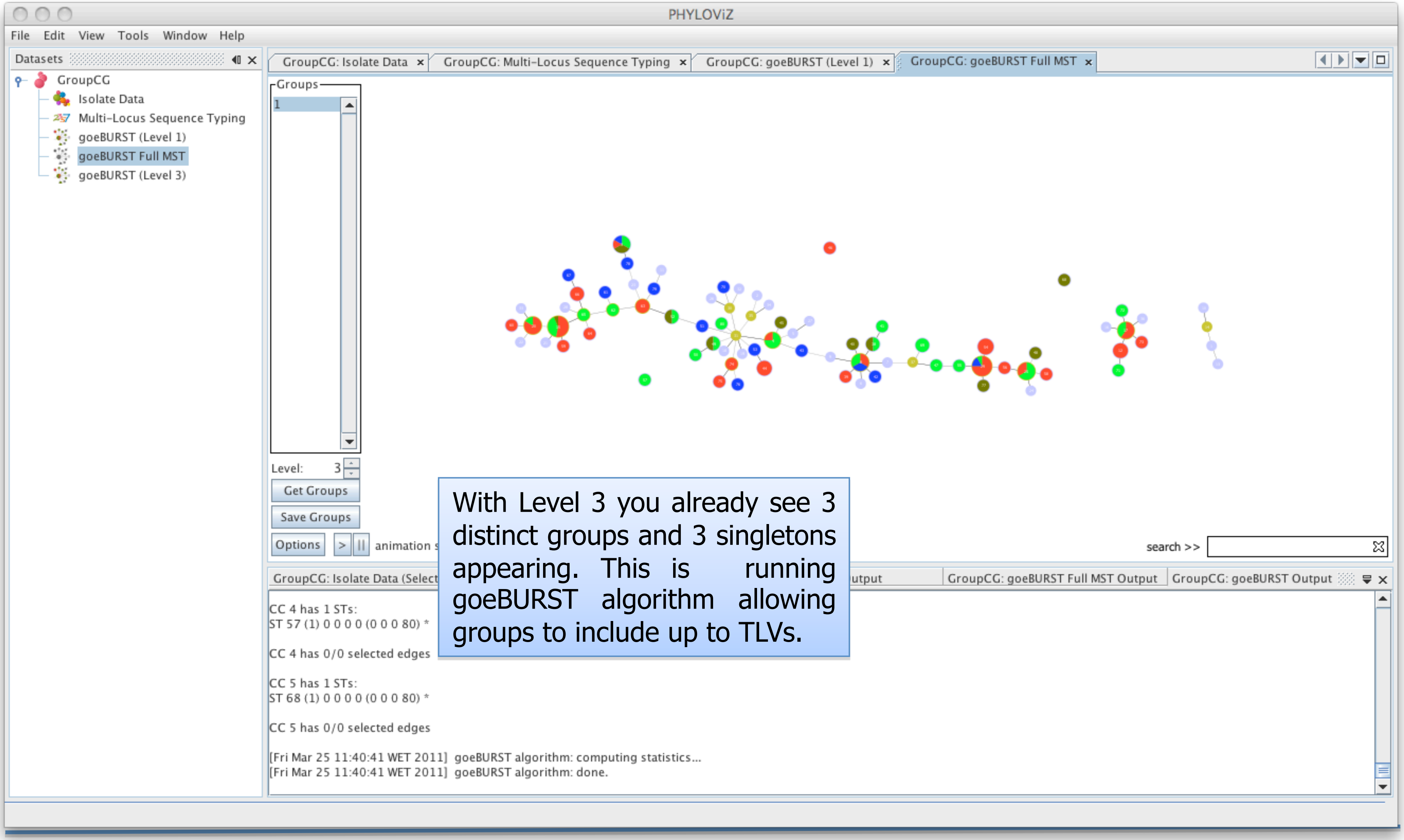
The goeBURST Full MST algorithm also allows the visual exploration of how the Minimum Spanning Tree was constructed and how it could be subdivided by removing links

The **Level** box can be altered to only allow links between two nodes of distances equal or less than the number it indicates. This means that allows links to be drawn if the differences are 6 or less.



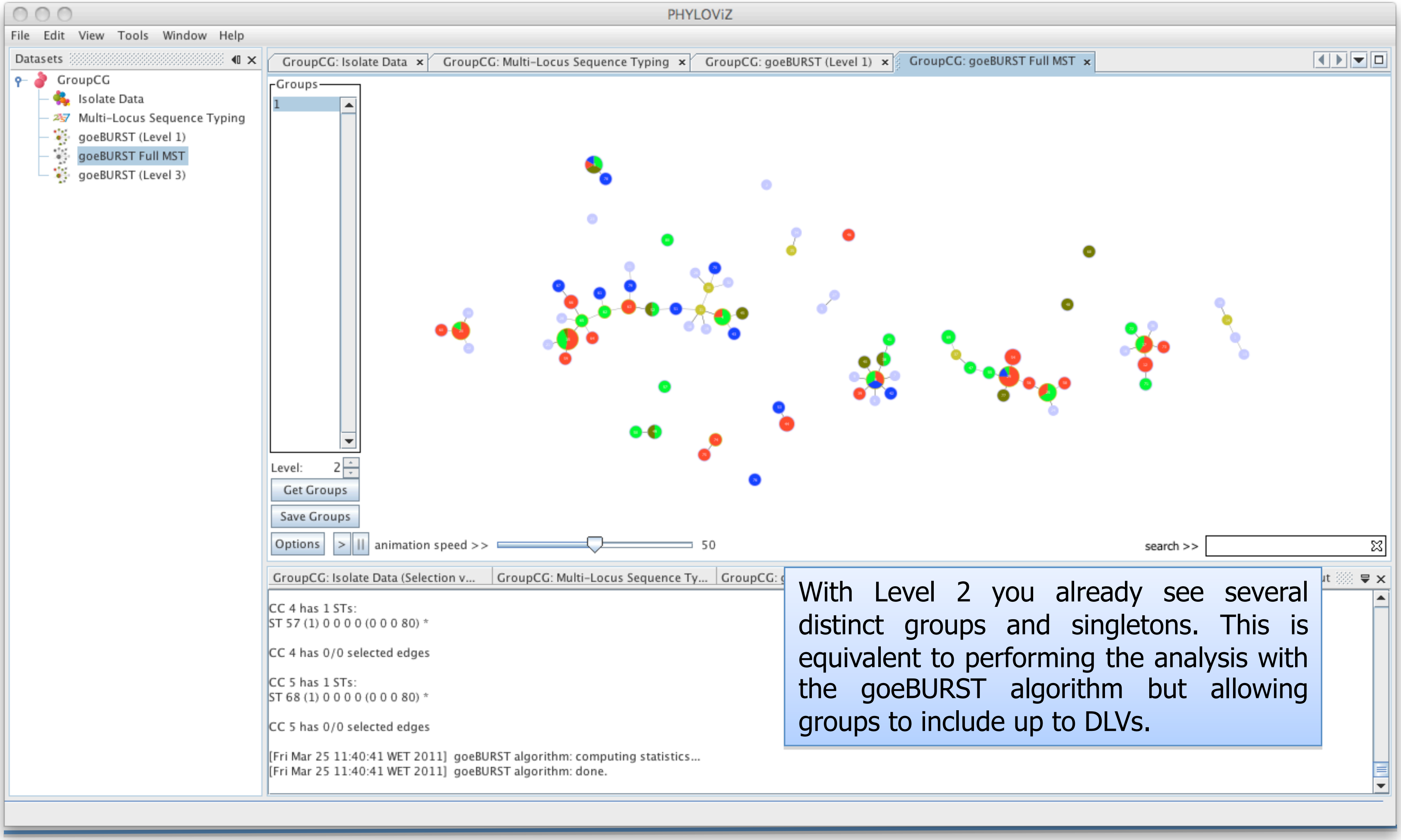
```
CC 4 has 1 STs:
ST 57 (1) 0 0 0 0 (0 0 0 80) *

CC 4 has 0/0 selected edges

CC 5 has 1 STs:
ST 68 (1) 0 0 0 0 (0 0 0 80) *

CC 5 has 0/0 selected edges

[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm:
[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: done.
```

In the dataset used in this example lowering the level to 5, you can see a group detaching. This means that the connection between the two nodes had at least 6 differences



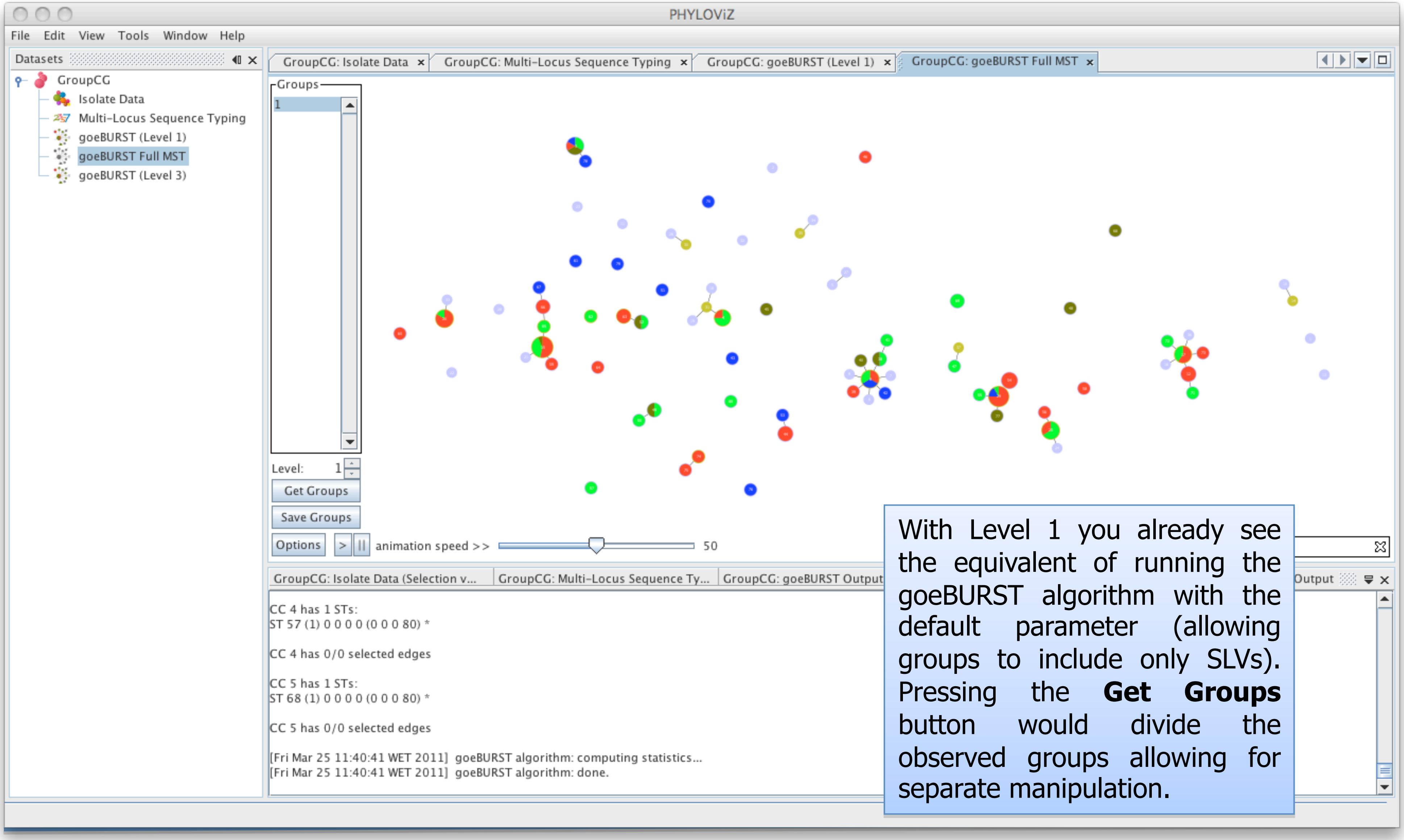
```
CC 4 has 1 STs:
ST 57 (1) 0 0 0 0 (0 0 0 80) *

CC 4 has 0/0 selected edges

CC 5 has 1 STs:
ST 68 (1) 0 0 0 0 (0 0 0 80) *

CC 5 has 0/0 selected edges

[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: computing statistics...
[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: done.
```

With Level 3 you already see 3 distinct groups and 3 singletons appearing. This is running goeBURST algorithm allowing groups to include up to TLVs.



```
CC 4 has 1 STs:
ST 57 (1) 0 0 0 0 (0 0 0 80) *

CC 4 has 0/0 selected edges

CC 5 has 1 STs:
ST 68 (1) 0 0 0 0 (0 0 0 80) *

CC 5 has 0/0 selected edges

[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: computing statistics...
[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: done.
```

With Level 2 you already see several distinct groups and singletons. This is equivalent to performing the analysis with the goeBURST algorithm but allowing groups to include up to DLVs.



```
CC 4 has 1 STs:
ST 57 (1) 0 0 0 0 (0 0 0 80) *

CC 4 has 0/0 selected edges

CC 5 has 1 STs:
ST 68 (1) 0 0 0 0 (0 0 0 80) *

CC 5 has 0/0 selected edges

[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: computing statistics...
[Fri Mar 25 11:40:41 WET 2011]  goeBURST algorithm: done.
```

With Level 1 you already see the equivalent of running the goeBURST algorithm with the default parameter (allowing groups to include only SLVs). Pressing the **Get Groups** button would divide the observed groups allowing for separate manipulation.