MISIS Tutorial

Table of Contents

I. Introduction

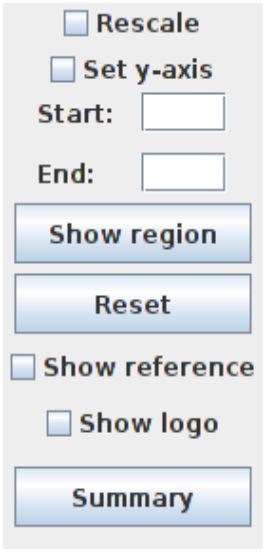
MISIS was originally developed to visualize small RNAs on the genome of plant viruses. MISIS is useful for any life sciences or medical researcher working with small RNAs. It is a tool to visualize, analyze and compare maps of small RNAs (sRNA ; typically such as siRNAs, miRNAs and piRNAs) on regions up to 300 kb. More specifically, sRNA cold and hotspots are displayed on both strands of a given reference sequence and any selected region of this sequence. This tool is easy and intuitive for the user. It is written in Java language, and therefore can be used on different Operating Systems such as Windows, Mac and Linux. Java must be installed on the user's computer. Otherwise, install a suitable Java Runtime Environment (JRE; see the instructions via this link http://www.java.com/).

II. Tool presentation

To run MISIS, double-click on the MISIS.jar file (or the MISIS icon). The following window will be displayed:

It contains four zones:

- the menu bar on the top: 

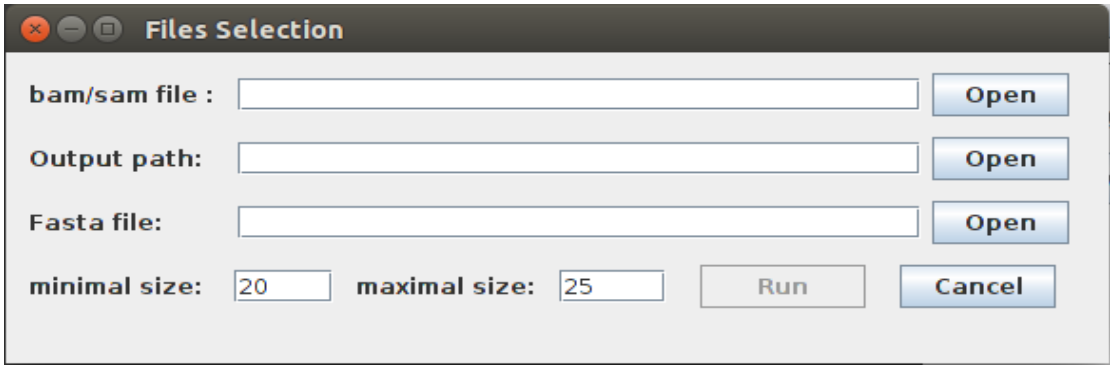
- the bar graphs zone in the middle (in grey)
- the visualization options on the right: 

- the slider zone on the bottom: 

III. Load files

a) Create a project by loading BAM files

MISIS creates a project by extracting the sRNA counts from a standard Sam/Bam file generated by mapping tools such as BWA. The user can load a Sam/Bam file and create a project via ' **File → load BAM/SAM** ' in the menu. A windows ' **Files Selection** ' appears.

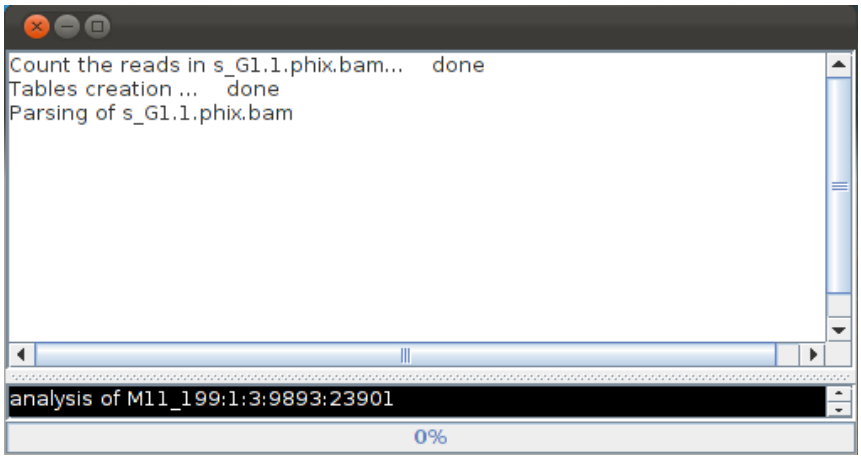


This window allow to indicate information for the project creation:

- '**bam/sam file**': the location of the Bam/Sam file which must be read
- '**Output path**': the output folder where the files created by MISIS will be saved
- '**Fasta file**' (optional) : the location of fasta file containing the reference sequence used for the mapping.
- '**minimal size**' : the minimal size of mapped reads which must be visualized with MISIS (20 by default).
- '**maximal size**' : the maximal size of mapped reads which must be visualized with MISIS (25 by default).

The '**Run**' button is visible only when the Bam/Sam file is selected and the output path is indicated.

After clicking on '**Run**' button, the user can follow the file processing steps in the following window:

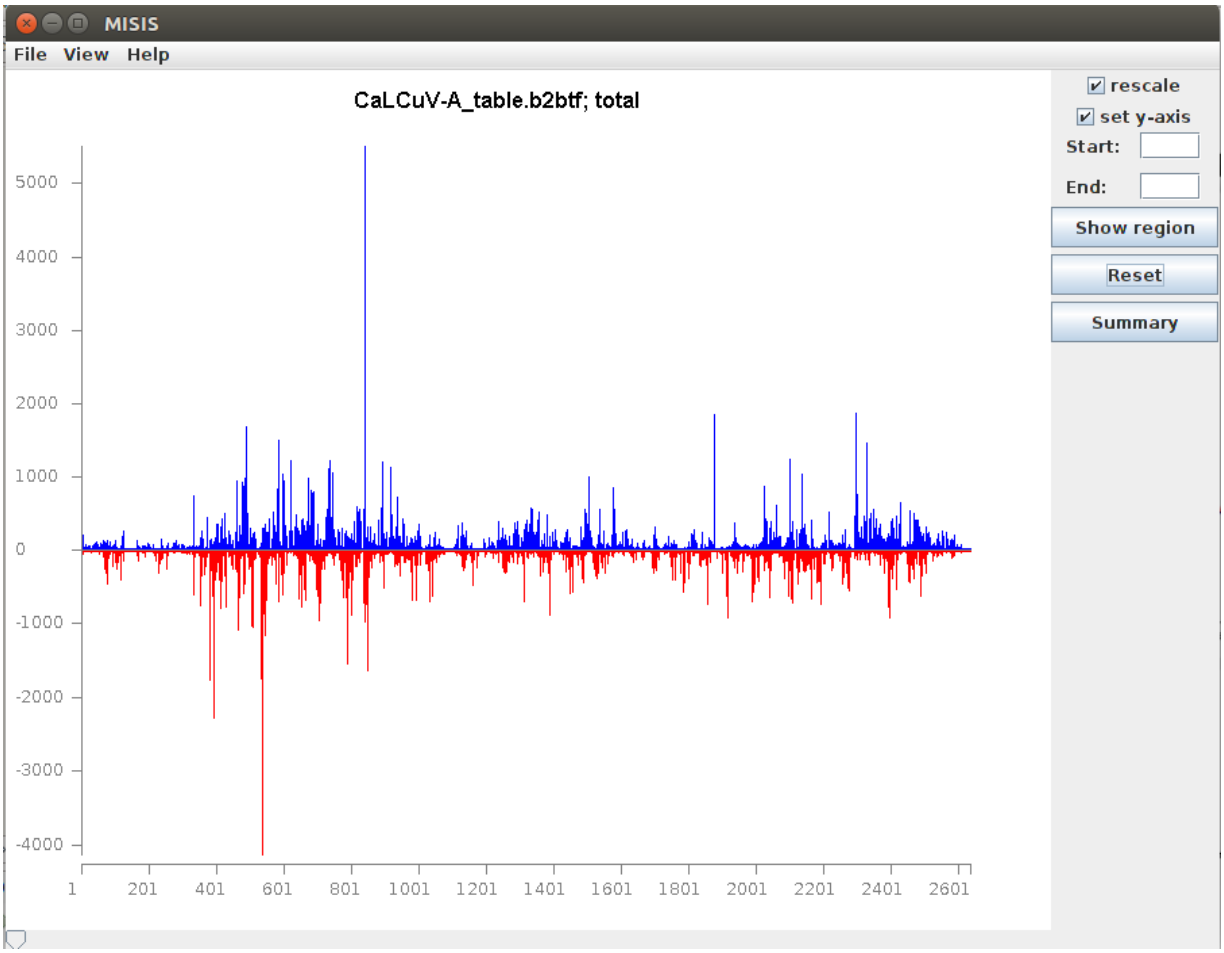


MISIS creates two types of files simultaneously: the first one with read counts for the sRNAs mapped to the reference sequence with zero mismatches ('**pm**' is present in its file name) and the second one with read counts of the sRNAs mapped to the reference sequence with up to 2 mismatches. In total, eight files are generated :

- two files which contain the count table and are extended by txt.
- two files which contain the MISIS projects. These projects contain the count table, the logo information and the sequence of reference if the fasta file was indicated. Their extensions are b2btf.
- four files which contain information to create the logos (two of which contain the nucleotide frequencies for the first mapped nucleotide of mapped reads).

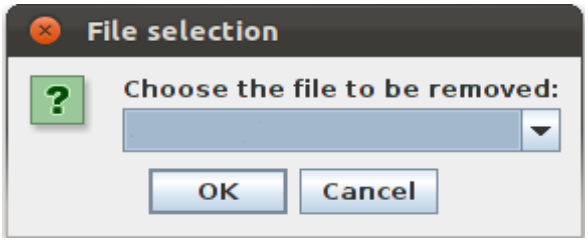
b) Load the Project

To load the project, the user selects '**File → Load Project**' in the menu bar. Then, MISIS draws the bar graph corresponding to the selected Input Table.



c) Remove the project

MISIS keeps the table files in the memory during the work. The user can add or remove different bar graphs from the window. The user can unload any project from the memory via '**File → Unload table**'. A pop-up window will allow to select the file to be unloaded.



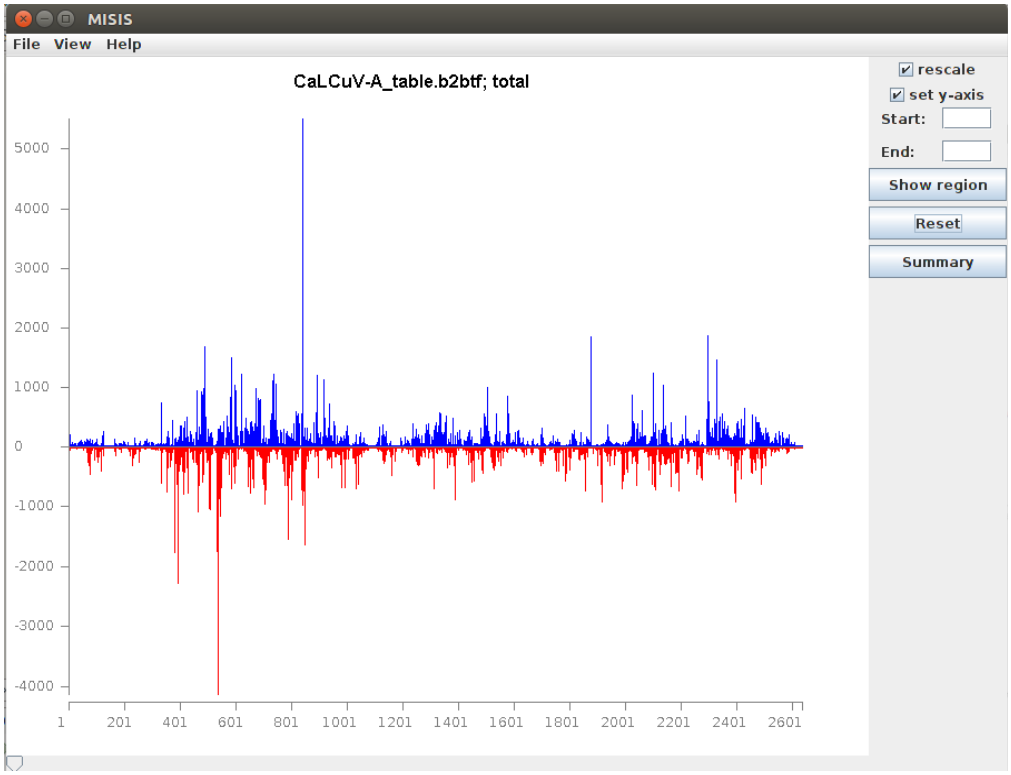
d) Load the demo

When MISIS runs for the first time, it propose to display a demo. Even if the user refuses to display the demo, he can display it later via '**File → Load demo data**'.

IV. Viewing bar graphs

a) Description of bar graphs

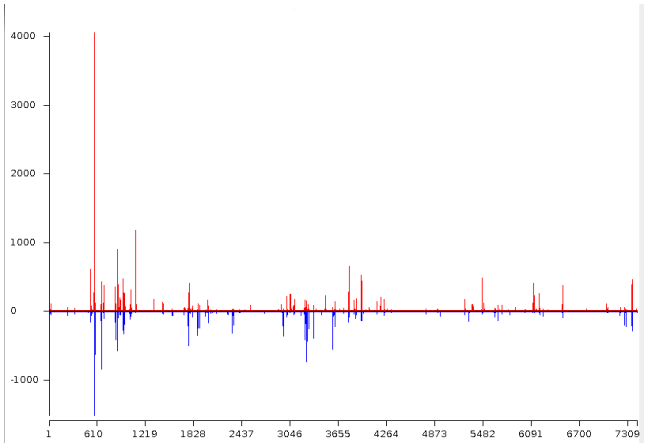
The name of the selected file with the selected columns is indicated as a title on the top. The *y* axis indicates the counts of sRNA reads. The positive values are counts of sRNAs mapped to the forward strand, and the negative values to the reverse strand. The *x* axis indicates the nucleotide positions along the sequence. Bars indicate the positions of 5'- or the 3'-end of each sRNA mapped to the reference sequence forward and reverse strands, respectively. Size of each bar corresponds to the read count for each sRNA starting or ending at the respective position.



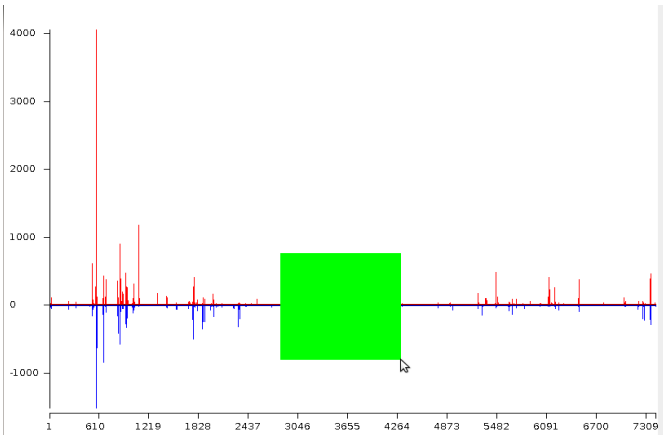
b) Region selection

The user can zoom in any region of the reference sequence by pressing on the mouse left button, selecting the region (visualized by green rectangle), and, when the region is selected, releasing the left button.

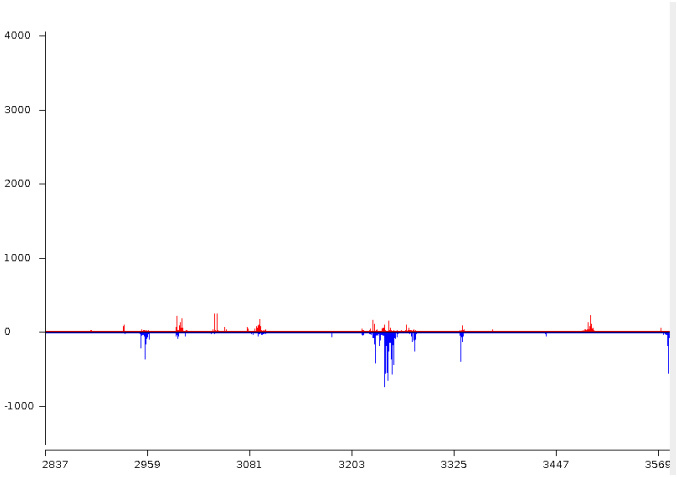
Before the selection:



During the selection:



After the selection:



To cancel the last zoomed selection, the user can click on the mouse right button. The '**reset**' button (in the visualization options zone on the right) cancels all zoomed selections.

Otherwise, the start and the end positions of a new region can be indicated in the visualization options zone: the selected region will appear when the user clicks on the '**show region**' button.

When the bar graph shows a selected region, the user can move along the sequence by using the slider on the bottom.

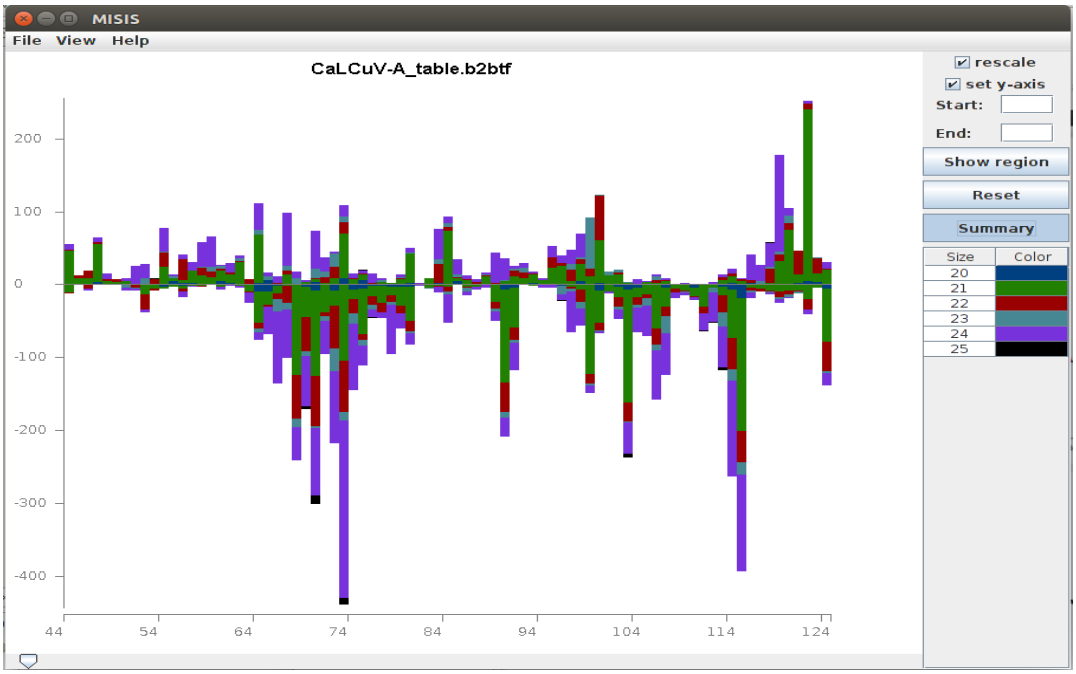
c) Modification of the y axis

The *y* axis is calculated according to the maximal value found in the Input Table. To adapt the bar graph to the observed minimal and maximal values, the user can select the checkbox called '**rescale**' in the visualization option zone.

When two or more bar graphs are displayed, the checkbox '**set *y*-axis**' will adjust the scale of all the bar graphs to the biggest maximal value found. It allows to adapt all bar graphs according to the best observed. Defining the scale allows to easily compare the read counts between the bar graphs.

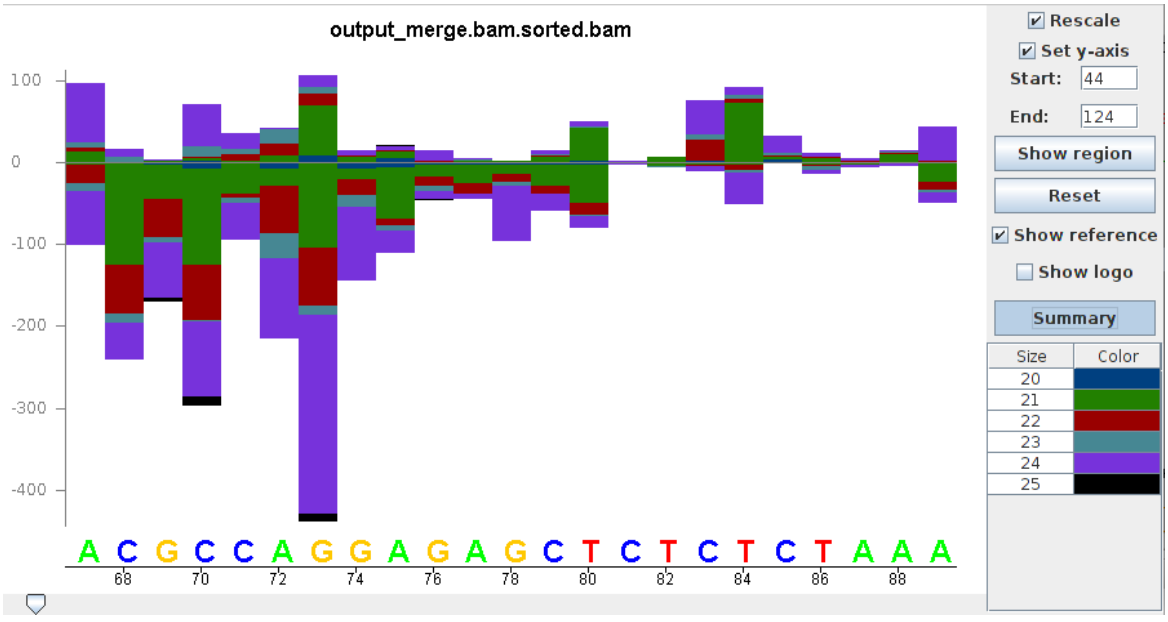
d) bar graphs summary

The bar graphs can be concatenated within a summary bar graph. Each bar contains different colors according to the size of mapped read per position. To generate this figure, the user clicks on the '**Summary**' button in the visualization option zone. The bar color can be changed by clicking on the corresponding color of the table in the visualization option zone. To return to the previous visualization, the user just clicks on the '**Summary**' button again.



e) Display the sequence of reference

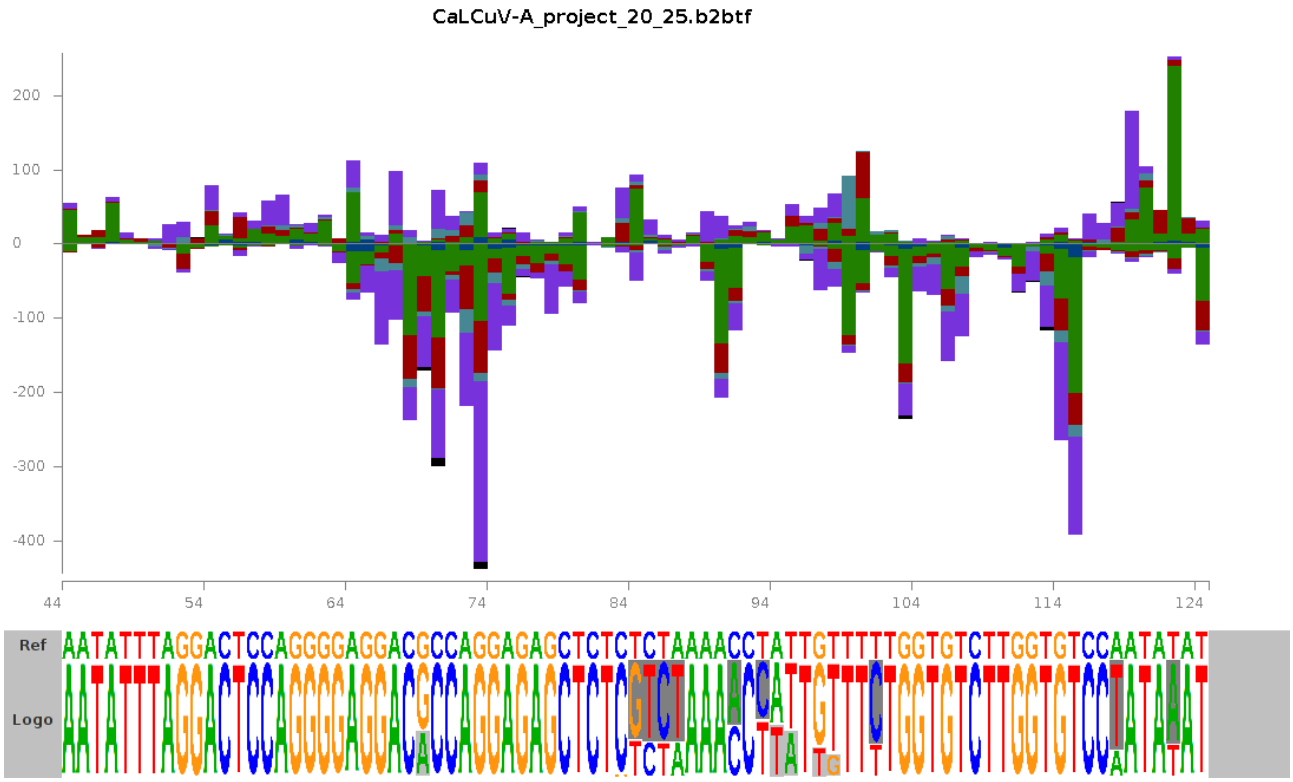
The sequence of the reference can be displayed by selecting the checkBox '**Show reference**' in the visualization option. Each nucleotide is represented with a specific color : A in green, C in blue, G in orange and T in red. The sequence is visible only in a zoomed region. The nucleotides are absent as long as the selected region is too big.



If the reference is not present, it is possible to load a fasta file containing this reference by selecting '**View → Load the reference**'. The sequence of reference is already saved within the project files only if the user provided the fasta file during the project creation (see section III.a).

f) Display the logo

In order to visualize the frequencies of nucleotides present on mapped reads, MISIS can display a logo which summarizes this information. The user must select the checkBox named '**Show logo**' to visualize the nucleotides frequency along the genomic sequence. The logo is visible according to the selected region : as the reference, the user must define a short region of mapped genome and then move along its in order to visualize the logo in both directions. Two types of logo can be visualized thanks to the combo box under the checkbox '**Show logo**' : the first displays the logo of all mapped nucleotide while the second displays only the frequency of the first mapped nucleotide of reads.



The background color of the logo is different according to the reference:

- Dark grey color: the nucleotide is different from the reference and is present in more than 50% of mapped reads.
- Light grey color: the nucleotide is different from the reference and is present in more than 10% of mapped reads.
- White color: the nucleotide is present in the reference or is present in less than 10% of mapped reads.

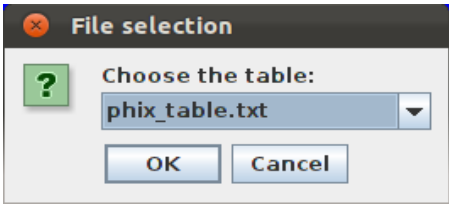
The 10 and 50% thresholds can be changed by the user via '**View → change the threshold for logos**' in the menu bar. Moreover, the exact count of each nucleotide can be displayed directly by a mouse right-click on the corresponding position on the logo panel within a pop-up window.

V. Export the consensus sequence

When a project is loaded, it is possible to export the consensus sequence of displayed histogram. This consensus sequence is determined according to the logo of all mapped nucleotides. It is created with the selection of the most represented nucleotide per position along the reference. The user must select '**File → Export Consensus Sequence**' in the menu bar, and then indicates the location of the fasta file which will contain the consensus sequence. The name of fasta file is the bar graph title.

VI. Export the Input Table

The user can create one Input Table corresponding to the observed region via '**File → Export table**' in the menu bar. Then a pop-up window allow to select the table.



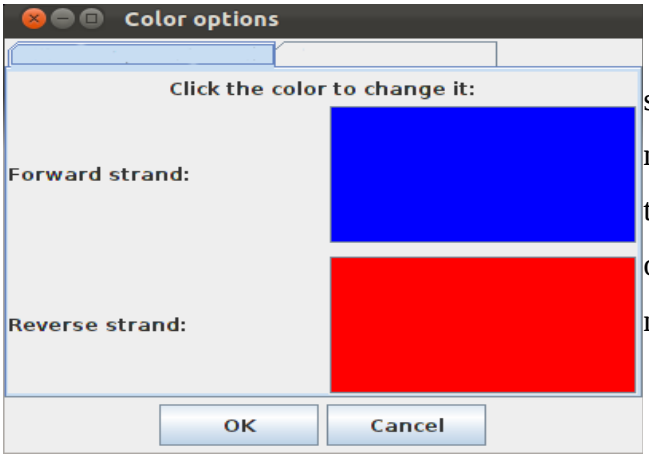
After the selection of the output path, the new Input Table is saved within the corresponding folder.

VII. Save image file

The user can save the drawn bar graphs in one image file in gif, png or jpeg format via '**File → Save image**' in the menu bar. The saved image contains only the displayed bar graphs (or the zoomed regions of the bar graphs).

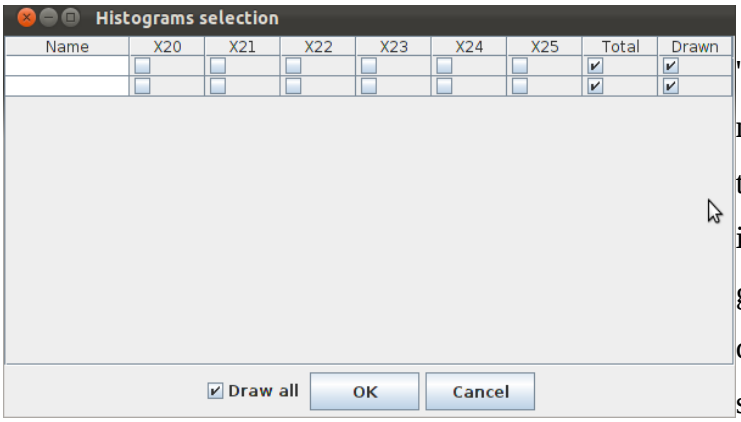
VIII. Options

a) Color option



The user can change the colors of bar graphs by selecting '**View → Change the colors**' in the menu bar. In the pop-up window, the tab on the top corresponds to the observed histogram, and one mouse click on the color allows to choose a new color among the color palette.

b) bar graphs selection option

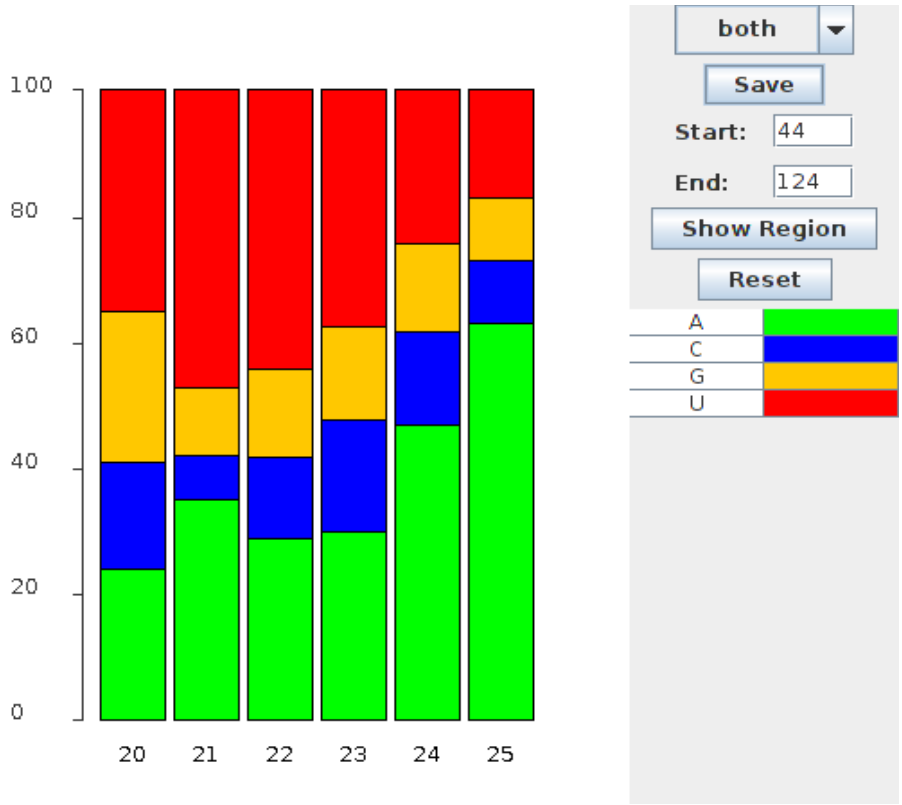


The user can select the bar graphs via '**View → Select the bar graphs**' in the menu bar. The pop-up window contains a table where the user can select the columns in each Input Table to be visualized as bar graphs. By clicking on '**drawn**', the user can indicate if the corresponding bar graph should be shown.

c) bar graph of 5' frequency

The 5' nucleotide frequency of mapped reads can by displayed in one bar graph where each bar represents the frequency for one specific size-classes. This graph is displayed by selecting '**View → 5' nucleotide distribution**' in the menu.

Different options on right allow to change the bar graph profile. The frequencies can displayed according to the mapped strand of the genomic reference: a combo box on the top allows to select the forward, reverse or both strands. By default, the bar graph represent the frequencies for the observed region. Another region can be selected by indicating its start and end position, and clicking on '**Show region**' button; the button '**Reset**' allow to select the complete sequence. The bar graph can be saved by clicking on the '**Save**' button: the image format are png, jpg or gif. The colors can be changed by clicking on the corresponding color associated with the nucleotide.



Appendix

The Input Table is a text files which contain columns separated by tabulation. The first column contains a number corresponding to the nucleotide positions of a given reference sequence starting from 1. The following columns contain the counts of sRNA reads which start at the given position on the forward strand and, after the forward read columns, the columns containing the counts of sRNA reads which end at the given position on the reverse strand. The number and the order of the reverse read columns is identical to the number and the order of the forward columns. For example, if the five forward read columns contain the counts of 20-, 21-, 22-, 23-, 24- and 25-nt sRNA reads mapped to the forward strand, these columns are followed by the five reverse read columns with the counts of 20-, 21-, 22-, 23-, 24- and 25-nt sRNA reads mapped to the reverse strand. After the reverse read columns, the next two columns contain the total counts of sRNA reads mapped on the forward strand and the reverse strand. The last column contains the total count of sRNA reads mapped on both strands. **Important : the first line of this table file must contain a header.**

Example of the Input Table:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	position	X20	X21	X22	X23	X24	X25	X20_rev	X21_rev	X22_rev	X23_rev	X24_rev	X25_rev	total_forward	total_reverse	total
2	1	1	3	2	1	2	0	-2	-2	-1	0	0	0	9	-5	14
3	2	0	0	0	0	1	0	-5	-9	0	-1	-8	0	1	-23	24
4	3	2	0	0	1	0	0	0	-2	0	-1	-2	0	3	-5	8
5	4	3	5	2	1	2	0	-2	-1	-1	0	0	0	13	-4	17
6	5	1	2	0	1	0	0	-1	0	-3	0	-5	0	4	-9	13
7	6	3	3	1	0	0	0	0	-17	-1	-1	-3	0	7	-22	29
8	7	1	0	1	0	10	0	0	-3	-1	-1	0	0	12	-5	17
9	8	0	0	0	0	0	0	-1	-5	-1	0	-1	0	0	-8	8
10	9	0	1	0	0	0	0	0	-3	0	0	0	0	1	-3	4
11	10	0	6	4	0	1	0	-1	-2	-2	0	0	0	11	-5	16
12	11	4	93	0	2	1	0	0	-5	0	-1	0	0	100	-6	106
13	12	10	0	1	0	1	0	0	0	0	-1	0	0	12	-1	13
14	13	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	1
15	14	0	0	0	0	0	0	0	-1	0	0	0	0	0	-1	1
16	15	0	0	0	0	1	0	0	-1	0	0	0	0	1	-1	2
17	16	0	2	0	2	0	0	0	0	0	0	0	0	4	0	4
18	17	0	3	0	0	0	2	0	0	0	-1	0	0	5	-1	6
19	18	0	3	0	0	5	0	-4	-34	0	-1	-6	0	8	-45	53
20	19	1	8	3	7	13	0	0	0	0	0	0	0	32	0	32
21	20	0	5	3	7	1	0	0	0	0	0	0	0	16	0	16
22	21	0	2	1	0	0	0	0	0	0	0	0	0	3	0	3
23	22	0	1	0	1	0	0	0	0	0	0	0	0	2	0	2
24	23	1	1	0	0	0	0	-1	-1	-1	0	0	0	2	-3	5
25	24	0	1	1	0	0	0	0	0	0	0	0	0	2	0	2
26	25	1	0	0	0	0	0	-5	-2	-1	0	0	0	1	-8	9
27	26	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
28	27	0	0	0	0	0	0	0	-3	0	0	0	0	0	-3	3
29	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0