CATCH

USER MANUAL

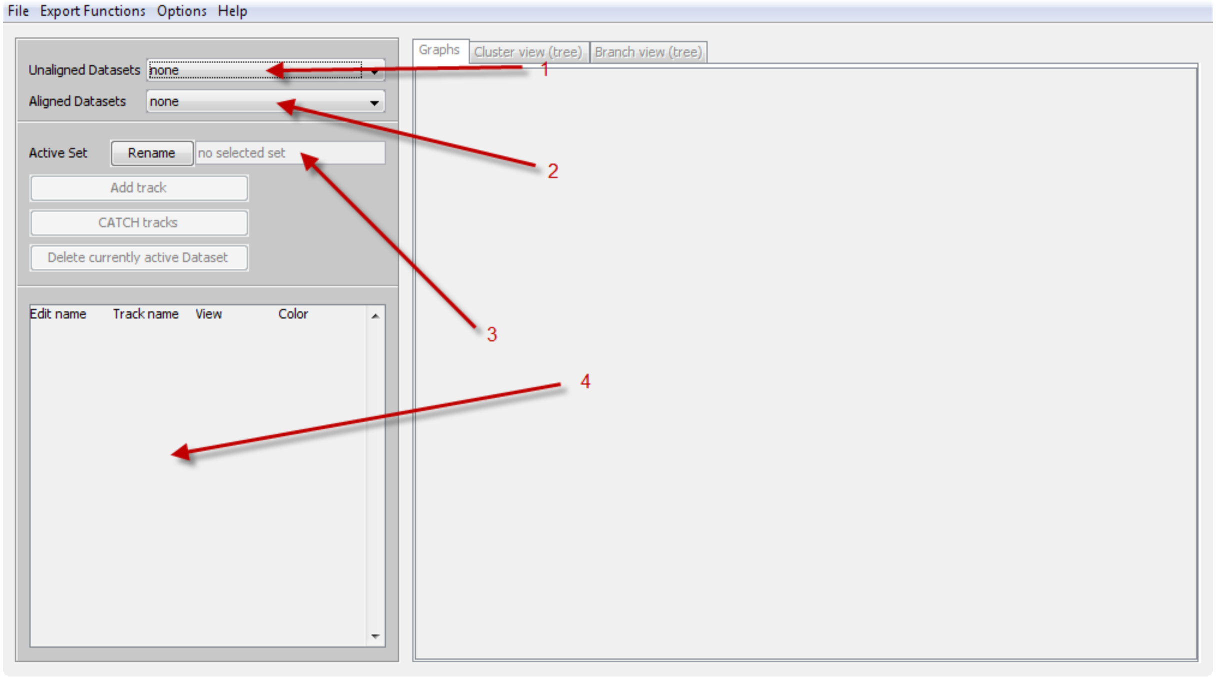
Author: Richard Willems, Fiona Nielsen

All comments and additions to this manual are more than welcome and can be sent to the address mentioned in the about section of the actual program. If you find any bugs, please report them by sending a mail to the CATCH- users mailing list. Please make sure to read the FAQ at the end of this manual to see if this bug is a known issue.

(All screenshots were created in a Windows Vista environment, other operating systems may show slightly different screens)

1. THE MAIN SCREEN

When you first run CATCH, you will be greeted what we call the main interface. In this main interface you will be able to perform all of CATCH's functions and view the results generated by the CATCH algorithm. The main interface looks like this:



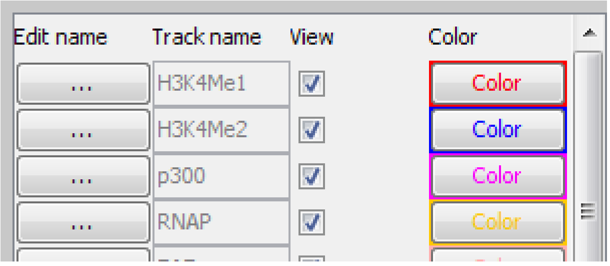
This is the main screen, when there are no datasets loaded (aligned or otherwise). There are four major aspects which are important to know right away. They are numbered one through four.

1. This list shows the unaligned datasets. When more than one set is loaded it will allow you to select the set you wish to use. An unaligned dataset is a dataset that has not yet been run through the CATCH algorithm and can contain one or more tracks. Once loaded, the tab marked 'graphs' will become active and will show all the tracks as graphs.

2. This list shows the aligned datasets. When more than one set is loaded it will allow you to select the set you wish to use. An aligned dataset is a dataset that has been run through the CATCH algorithm. Once an aligned dataset is loaded or created (more on that later) the tabs 'cluster view' and 'branch view' will become active and allow you to view the cluster tree created by the CATCH algorithm, as well as perform various functions on the result. These will be discussed in more detail in the appropriate chapter.

3. This box shows the full name of the loaded dataset if a dataset is loaded. The button to the left will allow you to make the box editable, and input a new name for the dataset. The change will be confirmed when you press the button again. This can be used for personal reference.

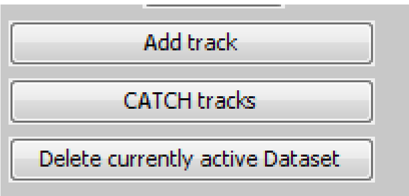
4. This list will show you the tracks you have loaded. An illustration is provided below.



Edit name	Track name	View	Color
...	H3K4Me1	☑	Color
...	H3K4Me2	☑	Color
...	p300	☑	Color
...	RNAP	☑	Color

The first column is a button which will allow you to change the name in the same way as described in point 3 of this chapter. The second column shows the name of the track. The checkbox in the third column will allow you to choose whether or not you want to make a track's graph visible in any of the view. Finally, the fourth column will allow you to choose the color which will be used to represent the track's graph in any of the views.

The buttons to the left (in the middle of the control panel) will be activated when relevant, they might look like this:



The first button allows you to add one or more tracks to the currently active unaligned dataset. If an active dataset is currently active, this function will not be available.

The second button starts the CATCH algorithm on the currently selected unaligned dataset, with the options selected in the options window. When the button is pressed a progress indicator will be shown for the current job. When this job is running other tasks may be performed, and it is also possible to configure and start a new job. Keep in mind that running several jobs at once can severely slow down your computer!

The final button deletes your currently selected dataset from the system, unaligned or otherwise. Remember that the dataset will be permanently gone, unless you saved to disk before deleting it!

2. THE TOOLBAR

The main toolbar will allow you to access a variety of CATCH's functions and tools from one location. The toolbar looks like this, and is found in the top of the main interface:

The first menu, the 'file menu', will allow you to perform various program and file related tasks. The top-most option you see will open the file manager which allows you to create new files for use in CATCH.

The second button, import dataset, will allow you to select one more processed wig files (*.pwig) to use in a new dataset.

The third button will allow you to select one more tracks (in the *.pwig format) to add to the currently active unaligned dataset. Below the first separator you will find the load and save CATCH cluster buttons. These will allow you to load a previously saved processed dataset which you have saved, or to save the currently selected aligned dataset.

Finally, at the bottom of the file menu is the exit button which does exactly what you would expect from such a button – close the program and return you to your OS.

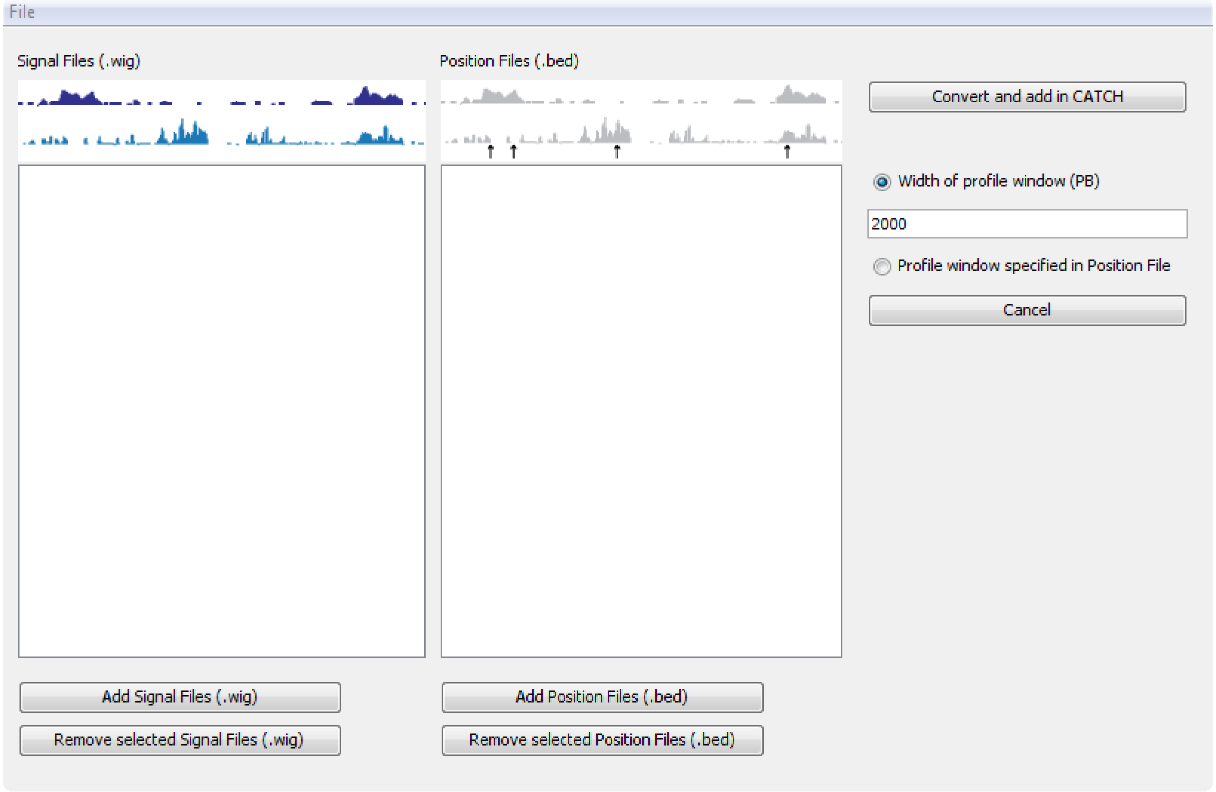
The second menu, the one called 'export functions' contains various commands to export CATCH's data into formats to be used in other analytical programs. The first one, 'export dataset' will store the processed wig files with a prefix you can specify. This will allow you to organize *.pwig for future reference. The other buttons will allow you to store the currently selected aligned dataset in a variety of formats; in this case NEWICK, BED and CSV. The scope of these file types will not be discussed in this manual. Free reference material is widely available on the internet.

The third menu is called 'options' and allows you to change CATCH's options. More information on the options menu is available in a later chapter.

The last menu is the 'help' menu. In further version of CATCH there will be an interactive help function, but in this version only the 'about' window can be accessed here. This window will allow you to marvel at the names of the wonderful people who brought you CATCH.

3. THE FILE MANAGER

When you click the 'import file' button in the 'file menu' (as explained in chapter 2) you will open the CATCH file manager, which allows you to import and convert files. More specifically: you will be able to select one or more wiggle/signal/*.wig files, combine these with a peak file *.bed and generate what we call 'Processed Wig Files', or *.pwig (to be pronounced as [pee-wig], not [pwig] which when you mispronounce it might cause you to be hit in the face, try it if you do not believe us). These files are downsized wiggle files from which we only take the parts around the relevant peaks as described in your peak file. You can add signal files and peak files by using the buttons at the bottom of the window as illustrated here:



● Width of profile window (PB)

2000

○ Profile window specified in Position File

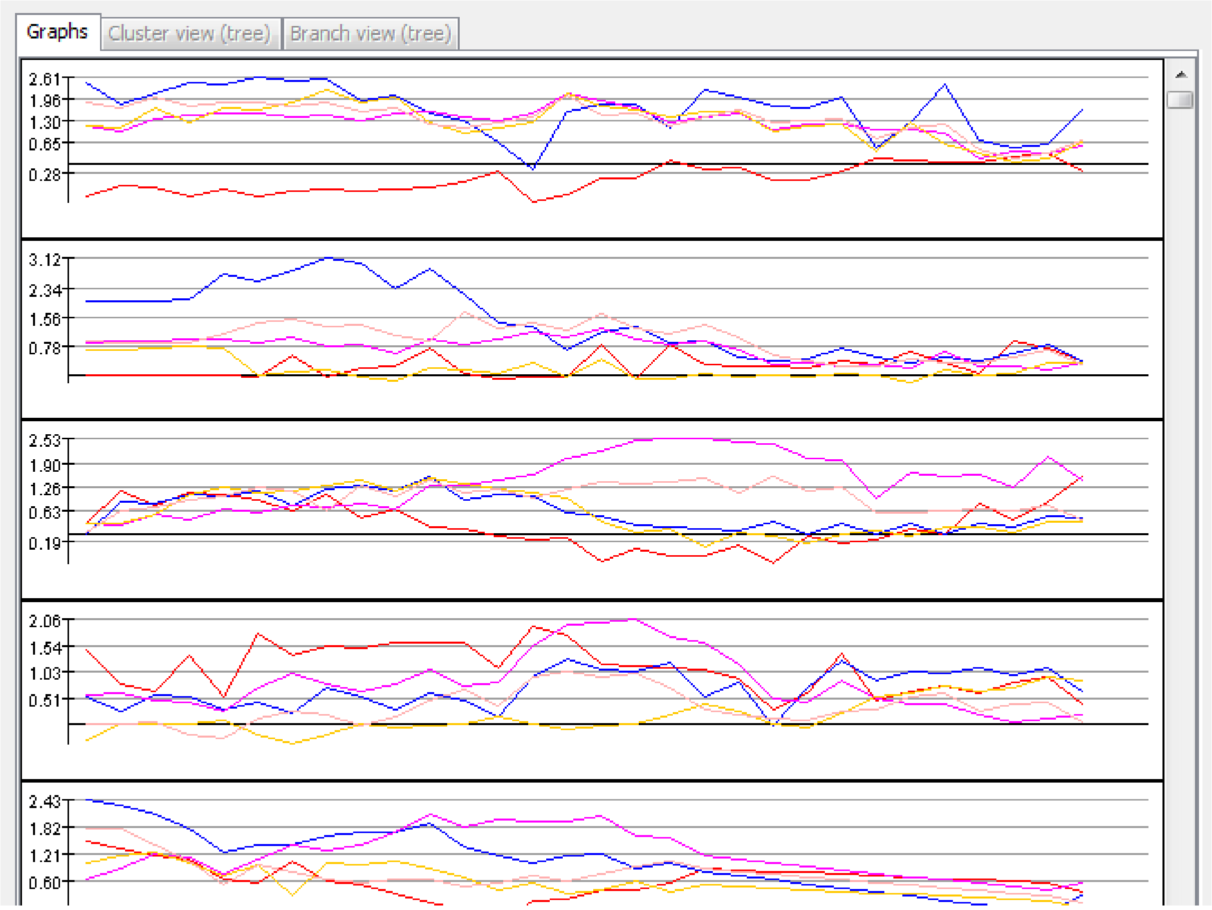
The loaded files will be shown in the lists above the buttons. You may load as many peak and signal files as you want. However, in the lists you can only select as many files as you want when working with signal files. For every amount of signal files you may only select one peak file. This is fairly straightforward since every processed wig file may only have one set of peaks.

To the right of the window you can select the width of the profile window; which is how large the area around the peaks should be when cutting them from the signal files and putting them into the processed wig file. When you press the button to the top-right, 'convert and add in CATCH' you will tell the program to convert the selected files to a processed wig file and load them into the main CATCH program where you can work with them as usual.

Finally, the 'cancel' button will close the file manager without making any changes.

4. THE DATA TABS

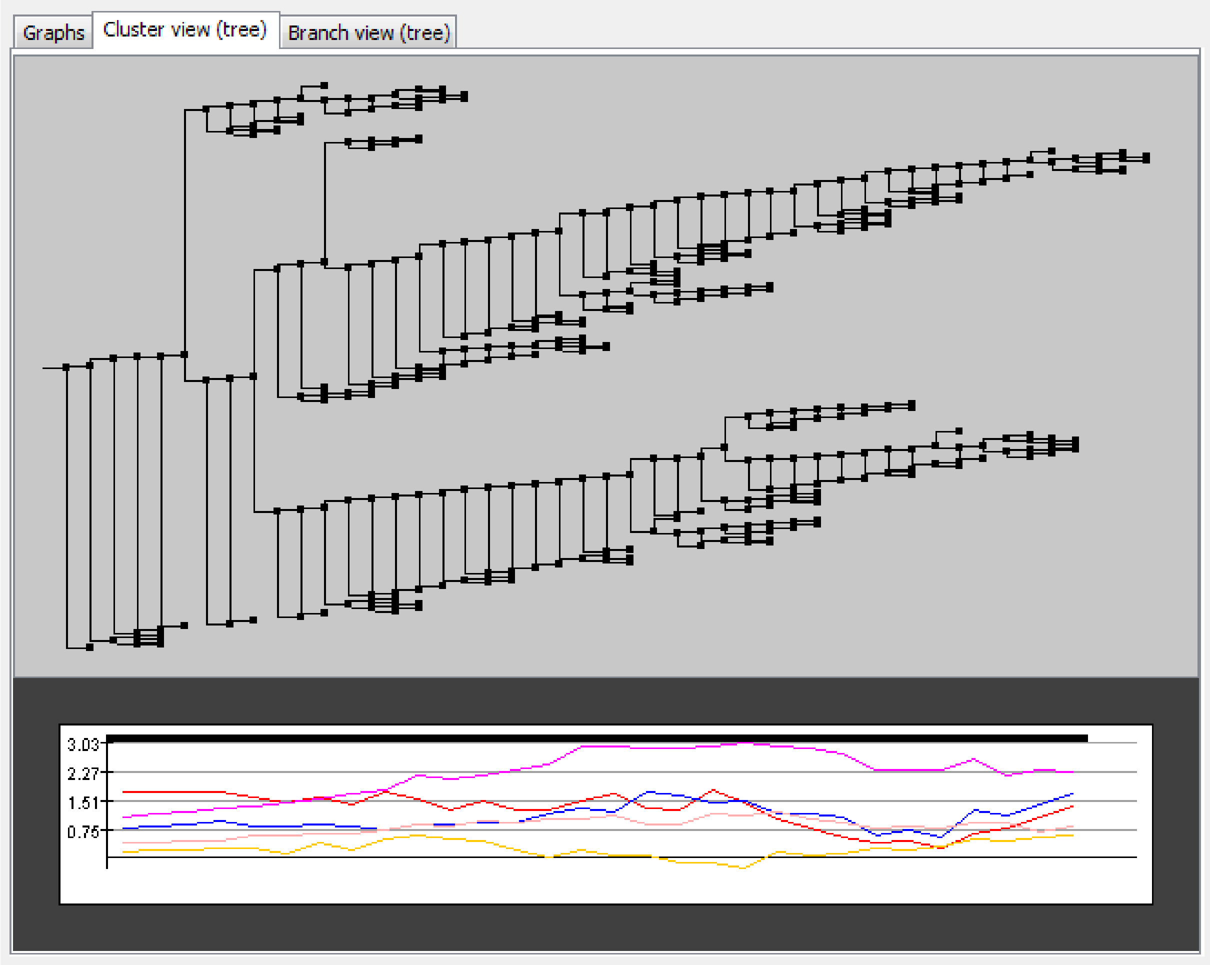
To the right of the main interface you will find the data tabs. These data tabs contain all of the profiles contained in the datasets you have loaded. When you have an unaligned dataset selected with one or more tracks, it will look like this:



You will notice that 'cluster view' and 'branch view' are currently disabled. This is because this data is not available in an unaligned dataset. To access this data you will first need to run the CATCH algorithm as described in chapter 1.

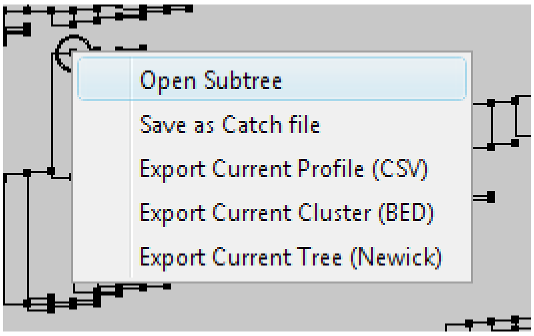
This graph view itself shows you a graphical representation of the selected profiles from your dataset. The intention of this screen is to present the profiles and allow the user to verify these profiles before running the CATCH algorithm.

When you run the CATCH algorithm the second and third tab become available and the first tab will be disabled. The cluster view is a tree which will allow you to visualize the entire cluster tree. It could look like this:



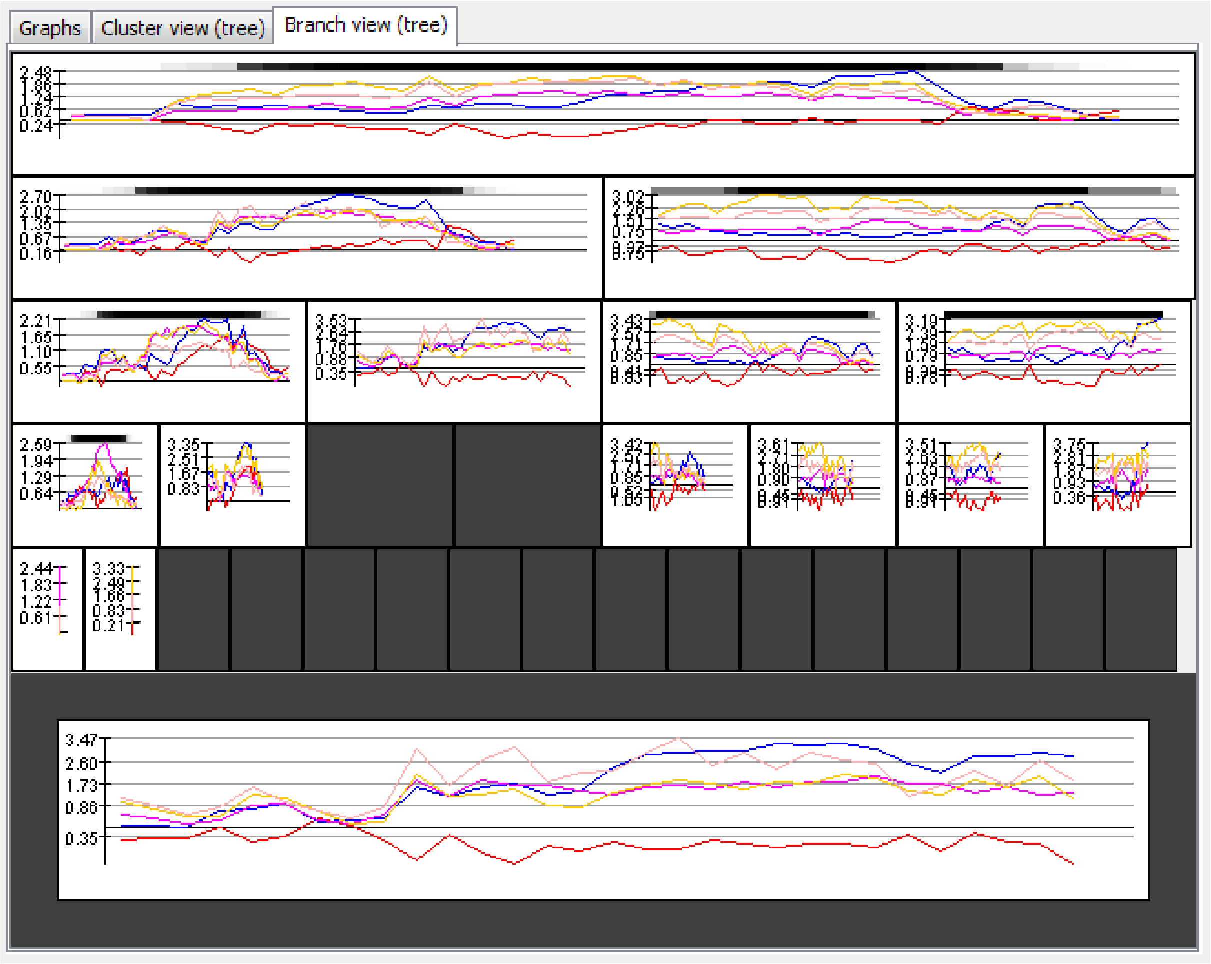
In this tree you can hover the mouse over any node to visualize the profile pattern in this node. When the node is clicked it is 'marked as interesting' and will turn blue for future reference. Zooming can be done with either a mouse wheel or using the '+' and '—' keys on your keyboard. When you press the 'r' button the view will return to the default zoom level to once again allow you to have an overview of the entire tree.

When the right mouse button is pressed on a node, it will show you the 'node menu' which looks like this:



The first option will allow you to open a sub tree. This is a new dataset which will start from the node you have selected. The tree viewer will then reset itself to show a new tree with the selected node as the root of the tree. All the other functions in the 'node menu' work the same as the export functions explained in chapter 2, with the difference that these will only save the tree from the node you have selected.

The third and final tab will show you the branch view. This is another version of the tree viewer which looks like this:



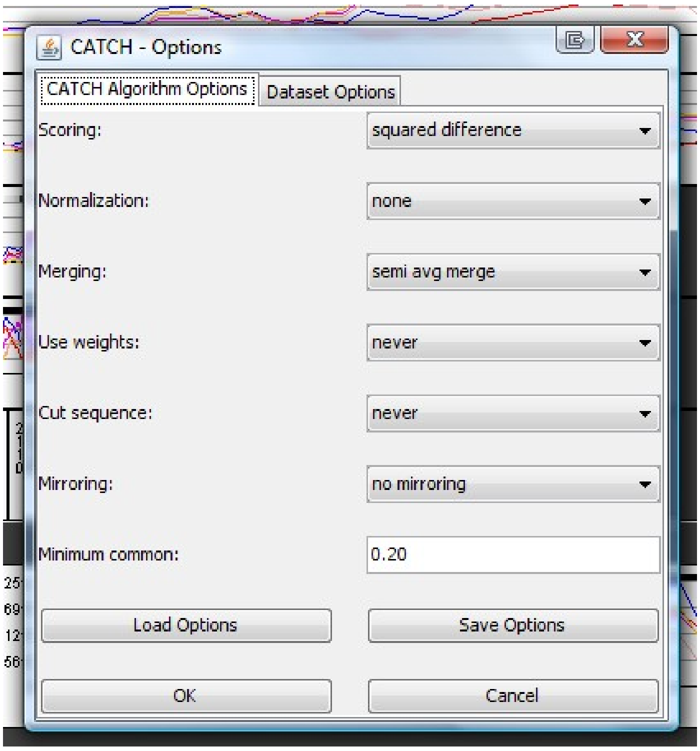
The main difference from the other tree viewer is that this one does not show the entire tree, but only the first four levels downwards from the node. What it does do, is show you the content of all the nodes which are visible in these levels.

The profile at the bottom is a more detailed view of the node you hover the mouse over.

When you click or right-click a node, it will reset the tree to show a deeper level of the tree with the clicked node as the new root of the tree. This will also be reflected in the tree viewer in the cluster view.

5. THE CATCH OPTION PANEL

The CATCH options are tools for you to further customize the results of your CATCH algorithm. It can be accessed through the 'options menu' as described in chapter 2. When you first open it, it may look like this:



Scoring: squared difference
Normalization: none
Merging: semi avg merge
Use weights: never
Cut sequence: never
Mirroring: no mirroring
Minimum common: 0.20

The first tab will show you all the options relevant for the CATCH algorithm, while the second tab will show you the options relevant for the dataset loading. Now before we continue we will first explain the way CATCH options work.

The options are stored in *.ini files, which allow you to quickly save and load your options for future use. This means that when you have fine tuned the options to your liking you can simply save these to a new file and load them the next time you run the program. To do this, use the buttons at the bottom of the screen labeled 'load options' and 'save options'.

The option screen itself is fairly straightforward. Some options allow you to choose from a pre-defined list of the options, while others allow you to enter any number you want.

Please keep in mind that some options drastically increase or decrease the time the algorithm takes to finish so if a certain calculation takes very long and you only wish to view a quick and dirty result you can play around with the options a bit to see which combination is best suited for your dataset. The time the algorithm actually takes is visible when it finishes, so this can be used as a reference.

The second tab contains the 'reference point interval' which should be set to the average distance between datapoints.

Further options might be added in future versions of CATCH.

6. FREQUENTLY ASKED QUESTION

Q. Why does CATCH pop up a file chooser when I load the program?

A. This is when the program cannot find the default options file, which is known to occur on certain versions of Linux. In the file chooser you are supposed to navigate to where you have installed CATCH and select the setup.ini file. This is a known bug that will be addressed in future version of CATCH but does not occur very often.

Q. Can I use the program when CATCH is calculating?

A. Yes you can. When you have initiated a very big task, you can simply put the progress window aside and continue using the program. You can even start more tasks, but keep in mind that performance will decrease when you have multiple tasks running at once. This is especially true if your computer is not equipped with a multi-core processor.

Q. Why am I able to load any sort of file I want?

A. Currently we have not yet perfected our file filter. Right now it is the responsibility of the user to make sure that only valid files are selected for use within CATCH. Also, file errors are minimal so if results are very unexpected or the program crashes, verify that the files you use are valid *.wig, *.pwig or *.catch files, depending on where you attempt to load these files.

Q. I own a Mac and don’t have a right-mouse button, what do I do?

A. Buy a decent mouse. Seriously, all programs these days require a right mouse button to function. Why Apple does still supply one-button mice is beyond us but we do not support them in CATCH.

Q. Why am I not able to see the relative alignment of the tracks?

A. We are working on this and it will be added in future version of CATCH.

Q. The CATCH algorithm is slow. Can’t you make it faster?

A. Yes we can, and we are in fact already working a version of CATCH that will work with parallel processing. This means that by using a still secret and mysterious process we will eventually be able to increase the speed by a factor of 40 or more. Stay tuned for this.

7. FINAL THANKS

As a last word of thanks we wish to thank the people of the ATV tree viewer for allowing us to make use of their work while construction the tree viewer we use. We also wish to thank the people who helped us test the application and who we did not stop bugging until we were sure we had milked their brains dry. Without them, CATCH would not be the magnificent piece of software you see before you.

We would also like to thank the main instigator of the project, Fiona Nielsen, for allowing us to work with her and gain some valuable experience with software development and multi-threaded application development. She could have said no ☺.

The main programming team,

Berry Lijklema
Mark Zandstra
Richard Willems