

# Mastering the Vidjil web application: Browsing and analysing clones

November 21, 2017



*The goal of this practical session is to learn common ways to visualize, filter, analyse and cluster clones on the Vidjil web application. These clones may have been computed by the Vidjil algorithm or by any other algorithm.*

0. Connect to the public server (<https://app.vidjil.org>), either with your account or the demo account ([demo@vidjil.org](mailto:demo@vidjil.org) / `demo`), select the *Demo LIL-L3 (tutorial)* patient, and click on the bottom right link, see results: `multi-inc-xxx`. Do not open the *Demo LIL-L3 (analyzed)* patient: this one contains the complete analysis. The Vidjil web application opens.

This patient (patient 063 from Lille study on the feasibility of MRD using HTS) suffering T-ALL has one diagnosis sample, with dominant clones both in IGH and TRG, and four follow-up samples, including a relapse.

1. In the settings menu, try the various options for `sample key`. The five samples can be labeled by their name, their date of sampling or by the number of days after the first sample.

In the following sections, we focus on the diagnosis sample. The section 4 will deal with the comparison of several samples.

## 1 Assessing the quality of the run and of the analysis

The Vidjil web application allows to run several “RepSeq” (immune repertoire analysis) algorithms. Each RepSeq algorithm has its own definition of what a clone is (or, more precisely a clonotype), how to output its sequence and how to assign a V(D)J designation. The number of analyzed reads will depend on the algorithm used. This sample has been processed using the Vidjil algorithm.

2. How many reads have been analyzed in the current sample with the embedded algorithm ?

Now we will try to assess the reason why some reads were not analyzed in our sample. This might reflect a problem during the sequencing protocol. . . or that could be normal. For that sake you will need to display the information box by clicking on the *i* in the upper left part.

3. What are the average read lengths on IGH? and on TRG? The lines starting with UNSEG display the reasons why some reads have not been analyzed.

You can see what those reasons mean in the online documentation of the algorithm: [vidjil.org/doc/algo.html#unsegmentation](http://vidjil.org/doc/algo.html#unsegmentation)

4. What are the major causes explaining the reads have not been analyzed? Also have a look at the average read lengths of these causes. Do you notice something regarding the average read lengths?

The percentage of analyzed reads can range from .01 % (for RNA-Seq or capture data) to 98-99 % (for very high-quality runs mostly on Mi-Seq).

## 2 Viewing and filtering clones

### 2.1 Looking to a clone

Each RepSeq algorithm has its own definition of what a clone is (or, more precisely a clonotype), how to output its sequence and how to assign a V(D)J designation.

In this file, the most abundant clone is IGHV3-9 7/CCCGGA/17 J6\*02.

5. Select this clone, either by clicking on the list or on the grid. How many reads do this clone represent? (see again the bottom part to the right)

There are several options to display the V(D)J designation.

6. In the **settings** menu, select **length** to show N zones by their length. Revert to the default **sequence (when short)** setting to show the full N on short sequences.
7. Try also the options **alleles in clone names** : by selecting **always**, the clone V gene is displayed as IGHV3-9\*01. Revert to the default when not \*01 to have more condensed V(D)J designations.

## 2.2 Showing more clones

By default Vidjil displays the 50 most abundant clones at each time point. With five time points, we may therefore have from 50 to 250 clones displayed depending if the top 50 are always the same or always different. This number can be increased to a maximum of 100 clones by going in the **filter** menu and by putting the slider to its right end.

8. Notice how the IGH smaller clones percentage changes. What was its initial value? What is it now? The *smaller clones* correspond to clones that are not displayed because they are never among the most abundant ones.

## 2.3 Tagging and filtering clones

Consider the most abundant clones in the list: IGHV3-9 7/CCCGGA/17 J6\*02 and TRGV10 13//5 JP1. Usually we may want to tag them in order to remember it later on.

9. Click on the star and choose colored tags for these two clones, such as **clone 1** or **clone 2**. Notice how the color applies throughout all the views.

Later you may want to filter clones depending on the tags you have chosen.

10. In the upper left part, click on the little gray square (at the right of the coloured squares). What happens? What if you click again?

This is a way of filtering some clones. This may be useful when we want to focus on some specific clones. Another way of doing so is to filter them by their gene names or by their DNA sequences.

11. In the search box, enter GGAGTCGGGG and validate with **Enter**. How many sequences are left? Note that the search is performed both on the forward and the reverse strand.

12. Check that by searching for the reverse complement of the sequence: CCCCGACTCC. Do you find the same results as previously?

13. How can you cancel this filter and view again all the clones?

Another solution to tag a specific clone is to rename it.

14. Double click on the name of a clone (in the list of clones) and choose another name (*e.g.* interesting clone) and validate using **Enter**.

After this rename, you can see that the clone is still selected.

15. Click on several clones by holding the **Ctrl** key to select more. Each time you add a new clone to the selection, its sequence is added in the bottom part.

16. How many clones are selected? How many reads do those clones represent?

17. When you want to focus on the selected clones, you can click on the focus link on the right, next to the number of selected clones. This feature is useful when you want to analyse some clones more thoroughly without being annoyed by other clones.

18. To remove this focus, click on the cross next to the search box, above the list.

19. To unselect them all, you can click in an empty area on the top or bottom plot.

Sometimes, one wants to hide noisy or unrelated clones.

20. Select a clone or several clones and click on the hide button, near the focus button. Show again these clones by clicking on the cross next to the search box.

## 3 Analysing clone populations

### 3.1 Clustering clones through inspection of their sequences

The first thing to be done is to see if some clones should be clustered (because of sequencing or PCR errors for instance). This step could be automatized but, in any case, the automatic clustering would need to be checked by an expert eye.

By default in the bottom plot (the *grid*), the clones are displayed according to their V and J genes (or more generally to their 5' and 3' genes).

**21.** Identify in the grid the clones with an *IGHV-3-13 IGHJ6* recombination and select them all. You can do so either by holding **Ctrl** or by drawing a rectangle around the clones while maintaining down the left button of the mouse.

The sequences of the clones now appear in the bottom part of the browser (the *segmenter*). If many clones are selected you can view more sequences by moving the mouse above the segmenter. The sequences in the segmenter can be visually compared but you can also align them to see more easily their similarities.

**22.** Click on the **align** button on the left-hand side. The differences are emphasized in bold.

Now it is the expertise of the user to determine if sequences are sufficiently similar, depending on the application. If some sequences don't appear to be similar enough, you can remove them from the segmenter by clicking on the cross in front of the sequence in the segmenter.

**23.** Remove all the sequences that are not similar enough with the first one.

Now all the sequences in the segmenter should be highly similar. All their differences should be due to sequencing or PCR errors. These artifacts (mutations, homopolymers, insertions, deletions) depend on the sequencer and the PCR technique.

**24.** Cluster all those clones in a single clone by clicking on the "cluster" button, next to the **align** button.

All the clustered sequences now appear within a same clone. That can be seen in the list: the clone which hosts the subclones appears with a + on its left. You can click on the + to see the subclones that have been clustered in the main one.

**25.** Click on the + and observe the changes in the grid.

As you may have noticed the subclones appear again in the grid. You can compare their sequences again if you'd like (for example to double check that you were right to cluster them). You can also remove some subclones from the cluster by clicking on the cross at their left in the list.

**26.** For the sake of the exercise, remove the last clone of the cluster.

**27.** Open the **cluster** menu, and choose **cluster by V/5**. What happened ? There are now two clones with TRGV2. Why ?

**28.** In the **cluster** menu, select **revert to previous clusters** to undo these clusterings.

### 3.2 Other metrics and analysis on the clones

As a proxy to sequence similarity we used the V and J genes, however there are other ways to assess sequence similarity that may be more pertinent. Moreover you may want to plot other metrics on the lymphocyte population. For instance we can choose to plot the V genes versus the length of the N insertions.

**29.** Go to the **plot** menu (in the upper left corner of the grid), and in the preset box choose **V/N length**. Then you can continue aligning and clustering clones if necessary.

**30.** You can also try the preset **clone consensus length/GC content** which tends to separate quite nicely the distinct clones.

Note that you can choose any axis to be plotted: just go the **plot** menu and select any value you would like for the *x* axis and for the *y* axis. For bar charts, the box sizes always relates to the clone size, and the *y* axis selects the order of the boxes sharing a same *x*).

**31.** In the plot menu, switch between the “bubble plot” and the “bar plot”. In the bar plot mode, pass the mouse over the bars: What happens?

Another possibility is to request Vidjil to compute the similarity between clones.

**32.** Now select the preset **plot by similarity** or even **plot similarity by locus** to plot similarity for the current locus (beware: this may take some time). Now the most similar clones should be close together. However note that it is theoretically impossible to achieve such a representation in 2 dimensions. So it is possible that two dissimilar clones are close together or, conversely, that two similar clones are far apart.

**33.** Press the keys 0 to 9 on the numeric keypad. What happens ?

There is still a feature to help you analyse your data that we have not explored yet. You can change the colors to make it represent some variables of interest with the **color by** menu.

**34.** First choose the preset **plot by similarity and by locus** and then **color by N length** (in the box at the top of the screen). Clones that are close on the grid with similar colors are likely to be similar.

**35.** Choose now the preset **CDR3 length distribution** and then **color by productivity**. See that the color tiles in the info part (upper right) change to show the color key.

We apologize to color  
blinds: the colors are  
not yet color-blind  
friendly.

Using those different features you should be able to analyse how similar your sequences are, and potentially you could cluster them if you'd like or tag them.

*This part is specific to samples analyzed with the embedded algorithm of the Vidjil platform.*

Some clones may be less trustable than other ones... Let's see how to spot them.

**36.** In the clone list, search clones with an orange warning at the right side. Click on the warning. What are the warnings due to?

There may have two reasons:

- average coverage: in that case the clonal sequence displayed is short compared to the reads in the clone. This may be the case when too different sequences have been put in a clone. The value is generally  $\geq 80\%$ .
- e-value: It is a statistical value computed to ensure that recombinations have not been spot by chance. This value is generally much lower than 1 ( $< 10^{-5}$ ).

You can view those values for any clone by clicking the *i* icon on the right side, in the list of clones.

### 3.3 Analysing recombinations from several loci

If you want to focus on specific locus, you can click on the locus name in the upper left part. One click will make the locus disappear, another one will make it appear again. If you hold the **Shift** key (the one which is usually above the left **Ctrl** key) while clicking it will hide all the loci but the one you clicked on.

**37.** Click on **IGH**, while holding the **Shift** key. Now what is the number of analyzed reads? Why did it change?

**38.** Now click on **TRG**, to filter it in again.

**39.** Press on the **g** key. What happens? Now, press on the **h** key. Press on the **g** again (you can do that anytime you like :)). Let's stick to the **TRG** locus.

You can also change the current locus by clicking on the locus name in the right part of the grid.



## 4 Tracking clones on several samples

The *time graph* shows the evolution of the top clones of each sample into all the samples. Bear in mind that to ensure readability at most 50 curves are displayed in this graph.

40. Pass the mouse over the bubbles in the grid or over the lines in the time graph. Click on some clone. What happens ?

41. Click on the label of the time graph to select another sample. What happens to the number of analyzed reads ? to the size of the top clones ?

When switching the time point, the views dynamically update which allows to easily track the changes along time. Also note that the number of analyzed reads differ from the previous point. We can again analyse the reason why some reads were unsegmented.

We will look now at how the V gene distribution evolves along the time.

42. In the grid, select the preset V distribution. Then click on the play icon in the upper left part (below the *i* icon).

By doing so you can look at how the V distribution changes along the time. Of course you can also change the data displayed in the grid to look at the evolution of another information.

We remind that by default at most 50 clones are displayed on the time graph. However the remaining of the application usually displays the 50 *most abundant clones* at each sample (which can account to hundreds of clones, when having several samples).

43. Select a sample, order the list by size, and pass the mouse through the list of top 50 clones. What happens in the graph when hovering clones that are not in the top 50 ?

If you have many samples, you may wish to reorder the samples.

44. Drag the label of one sample to reorder the samples.

45. Drag one label to the box with the pin icon to hide this sample.

You may also want to compare two samples, either to check a replicate, to check for possible contaminations, or to compare different research or medical situations.

46. In the color by menu, choose by abundance. Select a different sample. What happens ? Are there some clones with a significant different concentration in both samples ? Revert the color by choosing by tag.

Another option is to directly plot a log-log curve comparing two samples.

47. In the plot menu, choose the preset compare two samples. Click successively on two labels in the time graph to select the samples to be compared. Are there again some clones with a significant different concentration in both samples ?

## 5 Working with external software and exporting data

### 5.1 Checking VDJ designations with other software

For some studies, VDJ designations are very important. In the list and in the segmenter, those designations are written in their short form.

48. Put the mouse cursor over a clone. In the status bar (between the grid and the segmenter), the complete designation appears.

We can double check this designation with other popular software.

49. Select a few clones.

50. Click on the down triangle, which is right to IMGT/V-QUEST. The clone sequences are sent to IMGT/V-QUEST.

When loading data with only one sample, the time graph is replaced by a second bar/grid plot.

This requires an internet connection.

**51.** Then tick the checkbox 5'V/D/3'J. In the segmenter the boundaries of the V(D)J genes as computed by IMGT/V-QUEST are underlined.

Note that data returned by IMGT/V-QUEST is available by clicking on the *i* icon of analyzed clones, enabling you to compare the annotations made by the original software and by IMGT/V-QUEST.

**52.** You can also directly send the sequences to IMGT/V-QUEST or IgBlast by clicking the corresponding buttons. This opens a new page with the corresponding websites.

It may happen the software makes a mistake in the VDJ designation. In such a case you're very welcome to report us the problem and we will try to improve the designation algorithm.

**53.** Go in the **Help** menu and click on **get support**. It opens your mailer with a pre-composed email describing the data you are on as well as the clones you selected..

Even if you do not use the **get support** button, it's a good practise to send the complete address showing in your web browser, such as `http://app.vidjil.org/?set=3241&config=39&plot=v,size,bar`, when you want to discuss with colleagues or with us your data or your analyses.

Suppose that you would like to change the VDJ designation shown on the web application.

**54.** Click on the *i* icon in the list of clones for the clone you want to change the designation. In the segmentation part, click the edit button. Choose what you would like to modify.

Beware: the modifications you made (name changes, clusters, clone tagging, sample reordering...) will **not** be automatically saved. You have to save your changes by yourself either by clicking on **save patient** in the top left menu (where the "patient" name is written) or by using the **Ctrl+S** keyboard shortcut. For this demonstration data, you cannot save your changes as you do not have the rights to modify this patient.

## 5.2 Exporting data

**55.** In the export menu, generate printable reports by clicking on both entries starting with **export report**. What differs between both?

**56.** Select some clones and then, in the export menu, choose **export fasta**. What happens?

**57.** Open the **import/export** menu, and click on **export csv**. The resulting file describes all visible clones (V(D)J designation, size for each sample). It can be opened by any spreadsheet software such as LibreOffice Calc or Excel for further analysis.

**58.** Open again **import/export** menu, and click on one of the **export SVG** buttons. This exports the current view of the plot or the graph. The resulting file can be opened and edited from any image drawing software such as Inkscape.