

# Maîtriser l'application web Vidjil: Analyser et interagir avec les clones

21 novembre 2017



*Le but de cette session pratique est d'apprendre comment visualiser, filtrer, analyser et regrouper les clones de l'application web Vidjil. Ces clones ont pu être calculés avec l'algorithme de la plateforme Vidjil ou par n'importe quel autre algorithme.*

0. Connectez-vous au serveur public (<https://app.vidjil.org>), soit avec votre compte ou avec le compte de démonstration ([demo@vidjil.org](mailto:demo@vidjil.org) / `demo`), sélectionnez le patient *Demo LIL-L3 (tutorial)* patient, et cliquez sur le lien en bas à droite "see results : multi". L'application web Vidjil s'ouvre.

Ce patient (patient 063 de Étude de Lille sur la faisabilité de la MRD en utilisant le séquençage à haut débit) souffre d'une T-ALL. Nous avons un échantillon de diagnostic, avec des clones dominants en IGH et TRG, et quatre échantillons de suivi, incluant une rechute.

1. Dans le menu `settings`, essayez les différentes options de `sample key`. Les cinq échantillons peuvent être annotés par leur nom, leur date d'échantillonnage ou par le nombre de jours depuis la date du premier échantillon.

Dans la section suivante, nous nous focalisons sur l'échantillon de diagnostic. La section 4 concernera la comparaison de plusieurs échantillons.

## 1 Évaluation de la qualité d'un run et de l'analyse

L'application web Vidjil permet d'exécuter plusieurs algorithmes "RepSeq" (analyse de répertoires immunologiques). Chaque algorithme ReqSeq a sa propre définition de ce qu'est un clone (ou, plus précisément, un clonotype), et de comment analyser chaque read et leur assigner une désignation V(D)J. Le nombre de reads analysées dépend de l'algorithme utilisé. Cet échantillon a été traité par l'algorithme de Vidjil.

2. Combien de reads ont été analysées dans l'échantillon en cours ?

Nous allons essayer de comprendre pourquoi certains reads n'ont pas été analysées dans notre échantillon. Cela pourrait refléter un problème pendant le protocole de séquençage... ou être normal. Nous cliquons sur le « i » dans la partie en haut à gauche pour accéder à la boîte d'information sur l'échantillon en cours.

3. Quelle est la longueur moyenne des reads sur les IGH ? et sur les TRG ? Les lignes commençant par UNSEG donnent les raisons pour lesquelles certains reads n'ont pas été analysées. Vous pouvez voir ce que signifient ces raisons dans la documentation en ligne de l'algorithme : [vidjil.org/doc/algo.html#unsegmentation](http://vidjil.org/doc/algo.html#unsegmentation)

4. Quelles sont les principales causes expliquant que les reads n'ont pas été analysées ? Regardez aussi la longueur moyenne des reads non analysées. Observez-vous quelque chose de particulier au niveau de la longueur moyenne de ces reads ?

Le pourcentage de reads analysées peut varier de 0,01% (voire moins, pour du RNA-Seq ou de la capture) à 98-99% (pour des séquençages de très bonne qualité, généralement sur séquenceurs Illumina)

## 2 Visualiser et filtrer les clones

### 2.1 Visualiser les clones

Dan ce fichier, le clone le plus abondant est IGHV3-9 7/CCCGGA/17 J6\*02.

5. Sélectionnez ce clone, en cliquant sur la liste ou la grille/ Combien de reads ce clone représente-t-il ? (voir en bas à droite)

Il y a plusieurs manières d'afficher la désignation V(D)J.

6. Dans le menu *settings*, sélectionner *length* pour afficher les zones N en fonction de leur taille. Revenez à la situation initiale grâce au réglage *sequence (when short)* pour afficher en entier les séquences N (quand elles sont courtes).

7. Essayez aussi les options *alleles in clone names* : en sélectionnant *always*, le gène V est affiché comme IGHV3-9\*01. Revenez à la situation initiale avec *when not \*01* pour avoir des désignations V(D)J plus compactes.

## 2.2 Montrer plus de clones

Par défaut, Vidjil affiche les 50 clones les plus abondants (*top 50*) pour chaque échantillon ou point de MRD. Avec 5 points de suivi, nous pouvons donc avoir entre 50 et 250 clones observables, tout dépend si le *top 50* reste le même pour tous les points ou s'il varie. Ce nombre minimum peut être augmenté jusqu'à 100 en allant dans le menu *filter* et en poussant vers la droite le curseur.

8. Observez comment change la taille cumulée des petits clones IGH (« smaller clones »). Quelle était la valeur initiale ? Quelle est-elle maintenant ? Les « smaller clones » correspondent aux clones qui n'ont pas été séparément analysés car ils ne sont jamais parmi les plus abondants.

## 2.3 Taguer et filtrer les clones

Considérons les clones les plus abondants de la liste : IGHV3-9 7/CCCGGA/17 J6\*02 et TRGV10 13//5 JP1. Il serait intéressant de pouvoir les étiqueter pour s'en souvenir plus tard.

9. Cliquez sur l'étoile à côté de ces clones et choisissez des catégories telles que *clone 1* ou *clone 2*. Observez comment la couleur s'applique dans toutes les parties de l'application.

Lorsque des clones ont été étiquetés, il est ensuite possible de les filtrer selon les couleurs choisies.

10. Dans la partie en haut à gauche, cliquez sur le carré gris (à droite de la liste des carrés de couleur). Que se passe-t-il ? Et si vous recliquez encore ?

C'est un moyen de filtrer certains clones. Cela peut être utile si vous voulez vous focaliser sur quelques clones spécifiques. Il est également possible de filtrer ces clones, soit par leur nom, soit par leur séquence d'ADN.

11. Dans le champ de recherche, entrez la séquence GGAGTCGGGG et validez la recherche par la touche Entrée. Combien de séquences ont été retirées ? À noter, la recherche se fait à la fois sur les brins sens et antisens.

12. Vérifiez si cela est vrai en recherchant maintenant la séquence réverse complémentaire : CCCCGACTCC. Trouvez-vous le même résultat que précédemment ?

13. Comment pouvez-vous annuler ce filtre et revoir l'ensemble des clones ?

Une autre façon de signaler un clone spécifique est de le renommer.

14. Double-cliquez sur le nom d'un clone (parmi la liste des clones) et lui choisir un autre nom (comme par exemple « clone principal »). Validez cette action par la touche Entrée.

Après ce renommage, vous pouvez voir que ce clone est déjà sélectionné.

15. Cliquez sur plusieurs clones en maintenant la touche *Ctrl* enfoncée pour conserver la sélection. A chaque ajout d'un clone, sa séquence apparaît dans l'encart du bas.

16. Combien de clones ont été sélectionnés ? Combien de reads cela représente-t-il ? (Regarder à droite dans l'encart du bas.)

17. Quand vous voulez vous focaliser sur des clones sélectionnés, vous pouvez cliquer sur le lien (*focus*) sur la droite de l'encart du bas (à droite du nombre de clones sélectionnés). Cela est pratique lorsque vous souhaitez analyser quelques clones parmi de nombreux autres sans être gêné par ces autres clones.

18. Pour retirer ce focus, cliquez sur la croix située à droite du champ de recherche.

19. Pour tout désélectionner, vous pouvez cliquer sur une zone vide de la grille des clones.

Parfois, on souhaite cacher des clones provenant d'un bruit de fond.

**20.** Sélectionnez un clone ou plusieurs clones et cliquez sur `hide`, à côté de `focus`. Rendez ces clones à nouveau visible en cliquant sur la croix à droite du champ de recherche.

### 3 Analyse de populations clonales

#### 3.1 Regroupement de clones (« cluster ») par l'inspection de leur séquence

La première chose à faire est de voir si certains clones peuvent et doivent être « regroupés » (du fait d'erreurs de PCR ou de séquençage). Cette étape peut être automatisée mais cela n'enlève pas la nécessité d'une vérification de ce regroupement par un œil expert.

Par défaut, dans l'encart de visualisation graphique des clones (« grille »), les clones apparaissent classés selon leurs gènes V et J ou plus généralement selon les gènes 5' et 3'.

**21.** Identifiez dans la grille les clones avec la recombinaison *IGHV-3-11-IGHJ6* et sélectionnez-les. Vous pouvez utiliser la touche `Ctrl`, mais aussi les sélectionner en dessinant un rectangle autour des clones souhaités en maintenant le bouton gauche de la souris appuyé.

Les séquences des clones apparaissent désormais dans l'encart du bas (« segmenteur »). Si beaucoup de clones ont été sélectionnés, il est possible de voir toutes les séquences en faisant glisser la souris dans le segmenteur (l'encart s'agrandit). Les séquences dans ce segmenteur peuvent être comparées visuellement mais il est aussi possible de les aligner pour voir plus facilement les similarités.

**22.** Cliquez sur le bouton `align` dans la partie en bas à gauche. Les différences sont mises en exergue par un changement de style (gras et couleur brique).

Il en va de l'expertise de l'utilisateur et du cas d'utilisation pour déterminer si le degré de similarité entre les clones du segmenteur est suffisant pour un éventuel regroupement. Si certaines séquences ne semblent pas suffisamment homologues, vous pouvez les retirer du segmenteur en cliquant sur la croix devant la séquence (partie à gauche du segmenteur).

**23.** Retirez toutes les séquences qui ne sont pas assez proches de la première.

Maintenant toutes les séquences du segmenteur doivent être hautement similaires. Les différences observées ne doivent être dues qu'à des erreurs de séquençage ou de PCR. Ces artefacts (mutations, homopolymères, insertions, délétions) sont dépendants du séquenceur utilisé et de la technique de PCR.

**24.** Regroupez tous les clones en un seul en cliquant sur le bouton `cluster` puis sur le bouton `align`.

Toutes les séquences ainsi regroupées apparaissent sous un seul et même clone. Dans la liste des clones, c'est repérable : le clone qui contient les sous-clones apparaît avec un `+` sur la gauche. Vous pouvez cliquer sur ce `+` pour avoir la liste des sous-clones qui compose le clone issu du regroupement.

**25.** Cliquez sur le `+` et observez les changements dans la grille.

Comme vous avez pu le constater, les sous-clones apparaissent encore sur le graphe. Vous pouvez encore les comparer si vous le voulez (par exemple pour vérifier si le regroupement est correct ou non) et si nécessaire vous pouvez toujours retirer certains sous-clones du regroupement par la croix à gauche de la liste.

**26.** Juste pour l'exercice, retirez le dernier clone de la liste du segmenteur.

**27.** Ouvrez le menu `cluster`, et choisissez l'option `cluster by V/5`. Que s'est-il passé ? Il y a maintenant dix clones avec TRGV2. Pourquoi ?

**28.** Dans le menu `cluster`, sélectionnez `revert to previous cluster` pour annuler le groupe.

#### 3.2 Autres paramètres et analyses des clones

Pour déterminer les homologies de séquences, nous avons utilisé les gènes V et J. Toutefois, il existe d'autres façons de rechercher les similitudes entre les séquences, parfois plus pertinentes. De même, vous pourriez vouloir afficher d'autres paramètres de visualisation de la population lymphocytaire. Pour l'exemple, nous allons afficher les clones selon les gènes V versus la longueur des N insérés.

**29.** Dans le menu `plot` (coin haut gauche de la grille), dans le champ `preset`, sélectionnez `V/N length`. Vous pouvez continuer de regrouper les clones en les alignant puis en cliquant sur `cluster` si besoin.

**30.** Vous pouvez aussi utiliser le preset `clone consensus length/GC content` qui a tendance à bien séparer les clones distincts.

Notez que vous pouvez modifier directement les axes du graphe, en ouvrant le menu `plot` et en sélectionnant les axes  $x$  et  $y$ . Dans la visualisation en histogramme, la taille des rectangle dépend toujours de la taille des clones, et l'axe  $y$  règle l'ordre des rectangles pour chaque même  $x$ .

**31.** Dans le menu `plot`, changez la représentation des clones en bulles par celle en histogramme. Que se passe-t-il en mode histogramme lorsque vous passez la souris sur les bâtons ?

Une autre façon de visualiser les clones est de demander à Vidjil de classer les clones selon leur degré de similitude (distance).

**32.** Dans le menu `plot`, sélectionnez maintenant l'option `plot by similarity` ou `plot by similarity and by locus` pour afficher les clones d'un locus donné selon leur degré de similitude (attention : cela peut prendre un peu de temps). Ainsi, les clones présentant une très forte homologie sont à proximité les uns des autres. Il est théoriquement impossible d'avoir ce type de représentation en deux dimensions. Il est donc possible que deux clones non similaires soient très proches ou inversement que deux clones similaires soient très éloignés l'un de l'autre.

**33.** Appuyez sur les touches 0 à 9 du pavé numérique. Que se passe-t-il ?

Il existe encore une autre façon de vous aider dans l'analyse de vos données. Vous pouvez changer les couleurs à l'aide du menu `color by` pour mettre en exergue certains paramètres.

**34.** Choisissez d'abord dans le menu `plot` de la grille la visualisation `plot by similarity and by locus`. Puis, dans le menu `color by`, sélectionnez `N` (dans la fenêtre en haut de l'écran). Les clones qui sont proches sur la grille avec une même couleur sont probablement similaires.

Nos excuses aux daltoniens, puisque les couleurs ne sont pas encore bien différenciables.

**35.** Choisissez maintenant le réglage `CDR3 length distribution` puis colorez par `productivity`. Vous pouvez observer que les carrés de couleurs dans la partie information (en haut à gauche) changent pour montrer la légende des couleurs.

En utilisant ces différentes fonctionnalités, vous devez être capables d'évaluer le degré d'homologie de vos séquences et potentiellement de regrouper des clones ou les colorer si vous le souhaitez.

*Cette partie est spécifique aux échantillons analysés avec l'algorithme inclus dans la plateforme Vidjil.*

Certains clones peuvent être moins fiables que d'autres... Voyons comment les repérer.

**36.** Dans la liste des clones, recherchez des clones qui ont une alerte orange à droite. Cliquez sur l'alerte. À quoi sont dues ces alertes ?

À l'heure actuelle, il peut y avoir deux raisons :

- La moyenne de couverture ou *average coverage* : dans ce cas, la séquence du clone obtenue a une longueur plus courte que les reads qui composent le clone. Ce peut être le cas quand trop de séquences différentes ont été associées à un clone unique. Cette valeur doit normalement être supérieure à 0.8 (80%).
- La *e-value* : C'est une valeur statistique qui permet d'évaluer la fiabilité des recombinaisons et de s'assurer que les clones n'ont pas été détectés par hasard. Cette valeur doit normalement être très inférieure à 1 ( $< 10^{-5}$ ).

Vous pouvez voir ces valeurs sont visibles pour chaque clone en cliquant sur le  $i$  à droite des séquences dans la liste des clones.

### 3.3 Analyse des recombinaisons de plusieurs locus

Si vous voulez vous concentrer sur un locus spécifiquement, vous pouvez cliquer sur le nom du locus dans l'encart en haut à gauche. Un clic fera disparaître le locus (en grisé), un second le fera réapparaître (en couleur). Si vous maintenez la touche `Shift` (généralement au-dessus de la touche `Ctrl` de gauche) pendant que vous cliquez sur le nom d'un locus, cela cache les autres locus.

**37.** Cliquez sur `IGH` tout en appuyant sur la touche `Shift`. Quel est maintenant le nombre de reads analysées ? Pourquoi a-t-il changé ?

**38.** A présent, cliquez sur `TRG` pour les montrer à nouveau.

**39.** Appuyez sur la touche `G`, que se passe-t-il ? Appuyez maintenant sur la touche `H` puis à nouveau

sur la  $\mathbf{G}$  (vous pouvez faire ces alternances de touches autant de fois que vous le souhaitez). Continuons avec le locus TRG.

Vous pouvez aussi changer le locus actuel en cliquant sur le nom d'un autre locus à droite de la grille.

## 4 Suivi temporel des clones sur plusieurs échantillons

Le graphique en haut à droite, le graphique temporel, montre l'évolution des clones les plus abondants (figurant au moins dans le *top 10*) de chacun des échantillons (ou points) de l'analyse. Notons que par souci de lisibilité seules 50 courbes sont affichées au plus.

Lorsque un seul échantillon est présent, le graphique est remplacé par une seconde grille.

40. Passez la souris sur les bulles dans la grille ou au niveau des lignes du graphique temporel. Que se passe-t-il ?

41. Dans le graphique temporel, cliquez sur le titre d'un échantillon pour le sélectionner. Que se passe-t-il au niveau du nombre de reads analysées ? et sur la taille des clones principaux ?

Quand vous changez d'échantillon, les visualisations se mettent à jour dynamiquement pour faciliter le suivi. À noter, le nombre de reads analysées diffèrent à chaque point. Nous pouvons encore regarder la raison pour laquelle certains reads n'ont pas été segmentées.

Nous allons maintenant regarder comment la distribution des gènes *V* évolue au cours du temps.

42. Dans la grille, sélectionnez le preset *V distribution*. Ensuite cliquez sur l'icône play dans l'encart gauche haut (sous le bouton « i »).

Vous pouvez ainsi observer comment la distribution des gènes *V* évolue au cours du temps (à chaque point). Bien sûr, vous pouvez aussi changer les données (axes) de la grille et suivre l'évolution d'autres paramètres.

Rappelons que par défaut au plus 50 clones sont visualisés sur le graphique temporel, alors que les 50 clones *les plus abondants* de chaque échantillon sont affichés dans le reste de l'application.

43. Sélectionnez un échantillon, classez la liste par taille (*size*) et passez la souris sur la liste du top 50 (50 clones les plus abondants). Que se passe-t-il dans le graphique lorsque vous passez la souris au dessus de clones qui ne sont pas dans le top 50 ?

Dans le cas où vous avez plusieurs échantillons, il est possible de les réorganiser.

44. Cliquez sur le titre d'un échantillon et maintenez le bouton de la souris enfoncé pour le faire glisser et le déplacer.

45. Faites la même chose sur un autre échantillon, en le faisant glisser au niveau de l'icône en forme d'épingle. Cela permet de cacher l'échantillon.

Vous pourriez aussi vouloir comparer deux échantillons, par exemple pour vérifier un réplikat, rechercher des contaminations, ou comparer différentes recherches ou situations médicales.

46. Dans le menu *color by*, choisissez *by abundance*. Sélectionnez un autre échantillon. Que se passe-t-il ? Certains clones présentent-ils une concentration significativement différente entre les deux échantillons ? Modifiez la couleur en choisissant *by tag*.

Une autre option pour comparer les échantillons est de se placer directement en représentation log-log.

47. Dans le menu *plot*, sélectionnez le preset *compare two samples*. Cliquez successivement sur deux titres dans le graphique temporel pour sélectionner les échantillons à comparer. Existe-t-il encore des clones avec une différence de concentration significative entre les deux échantillons ?



## 5 Travailler avec d'autres logiciels et exporter ses données

### 5.1 Vérifier les désignations VDJ avec d'autres logiciels

Pour certaines études, il est très important d'avoir les bonnes désignations VDJ. Dans la liste des clones et dans le segmenteur, ces désignations sont écrites de manière concise afin de gagner de la place.

**48.** Positionnez le curseur de la souris sur un des clones. La désignation complète de la séquence apparaît dans la barre de status, entre la grille et le segmenteur.

Nous pouvons vérifier cette assignation par d'autres logiciels.

**49.** Sélectionnez quelques clones.

**50.** Cliquez sur le triangle à droite du bouton **IMGT/V-QUEST** dans la partie en bas à gauche. Les séquences des clones sélectionnés sont envoyées à **IMGT/V-QUEST**.

Cela nécessite une connexion internet

**51.** Ensuite, cochez la case **5'V/D/3'J**. Dans le segmenteur, les jonctions **V(D)J** qui ont été définies par **IMGT/V-QUEST** sont soulignées.

**52.** Vous pouvez aussi envoyer directement les séquences sur **IMGT/V-QUEST** ou **IgBlast** en cliquant sur le bouton portant leur nom respectif. vous pouvez remarquer que les données renvoyées par **IMGT/V-QUEST** sont disponible en cliquant sur sur le textiti bouton ce qui vous permet de comparer les annotations du logiciel originel et celle d'**IMGT/ V-QUEST**.

Il arrive parfois que **Vidjil** fasse des erreurs dans la désignation **VDJ**. Dans ce cas, vous êtes fortement invités à nous en faire part et nous essaierons d'améliorer l'algorithme de désignation.

**53.** Utilisez le bouton **get support** situé dans le menu **help**. Un mail pré-composé s'ouvre avec les références des données que vous visualisez tout comme des clones sélectionnés.

Même sans utiliser le bouton **get support**, une bonne habitude est d'envoyer l'adresse complète se trouvant dans votre navigateur web, telle que <http://app.vidjil.org/?set=3241&config=39&plot=v,size,bar>, quand vous voulez discuter avec des collègues ou avec nous de vos données ou de vos analyses.

Supposons que vous souhaitiez changer la désignation **V(D)J**.

**54.** Dans la liste des clones, cliquez sur le « *i* » à droite du clone dont vous voulez changer le nom. Dans la partie **Segmentation**, cliquez sur le bouton **edit**. Faites-les modifications souhaitées.

Attention : aucune des modifications effectuées (changement de nom, regroupement de clones, apposition de tags) ne sont sauvegardées automatiquement. Il faut le faire manuellement soit en cliquant sur **save patient** dans le menu tout en haut à gauche (là où le nom du « patient » est écrit), soit avec le raccourci clavier **Ctrl+S**. Cependant, pour ce jeu de démonstration, il n'est pas possible de faire de sauvegarde. Vous pourrez en revanche sauvegarder les échantillons qui vous appartiennent !

### 5.2 Export de données

**55.** Pour générer des rapports imprimables, dans le menu **import/export**, cliquez sur les deux entrées commençant par **export report**. Quelles sont les différences entre les deux ?

**56.** Sélectionnez quelques clones. Ensuite, dans le menu **import/export**, choisissez **export fasta**. Que se passe-t-il ?

**57.** Ouvrez le menu **import/export** et cliquez sur **export svg**. Le fichier ainsi généré décrit tous les clones (désignation **V(D)J**, abondance à chaque échantillon) et peut être ouvert par n'importe quel tableur pour de plus amples analyses.

**58.** Ouvrez de nouveau le menu **import/export** et cliquez sur **export svg**. Ceci exporte la vue du graphe ou de la grille. La fichier ainsi généré peut être ouvert et édité par tout logiciel d'image vectoriel comme **Inkscape**.

*Aurélien Béliard, Aurélie Caillault, Mathieu Giraud, Tatiana Rocher, Mikaël Salson, Florian Thonier,  
contact@vidjil.org*