

## Second monthly report

J'ai regroupé les deux visualisations faite jusqu'ici pour obtenir une représentation de la totalité d'un répertoire immunitaire. A gauche de la page il y a la visualisation du répertoire avec tous ses clones sous forme de cercle et à droite de la page il y a un diagramme à barre représentant les différents clones et leurs abondances au sein du répertoire. L'utilisateur peut modifier le seuil de clonalité du diagramme de barre à l'aide d'un formulaire. Lorsque l'on clique sur l'un des clones de la représentation en cercle on peut observer les différents clonotypes qu'il contient et une visualisation des clonotypes sous forme d'arbre apparaît sur la droite. Pour passer de la visualisation en cercle des clonotypes à celle en arbre il a fallut modifier le format des données.

En parallèle, j'ai commencé à récupérer les données des fichiers de sortie de GCtree pour obtenir un arbre des clonotypes. Après traitement des données, GCtree fournit plusieurs fichiers contenant les résultats. J'ai donc implémenté un script python permettant de récupérer les informations (abondance, hiérarchie, séquence...) contenues dans ces fichiers pour les mettre au format JSON. Puis à partir de ces informations j'ai construit un arbre avec des branches linéaires de différentes longueurs en fonction de la distance entre chaque séquences. Cette représentation pose néanmoins un problème, en effet la modification de la longueur des branches fait que certaines branches se superposent. Il y a également un problème au niveau des nœuds car la fixation des branches se fait au centre du cercle représentant le nœud donc lorsque l'on fait varier la taille du cercle cela fait également varier la longueur de la branche.

Il faut maintenant adapter la page représentant un répertoire de clones pour permettre la visualisation des données récupérées via GCtree. Il faudra alors modifier le script permettant de récupérer les données issues de GCtree pour pouvoir avoir des informations sur un ensemble de clones et pas uniquement un.

Plusieurs questions se posent pour la suite de la visualisation des données, notamment comment doit apparaître la distance des branches et de manière générale comment organiser le pipeline pour pouvoir traiter un grand nombre de données.