

Introduction

- Les **lymphocytes B** sont des acteurs de l'**immunité adaptative**.
- Ils possèdent des immunoglobulines de surface appelées **récepteurs des cellules B** (BCR).
- Les BCR subissent plusieurs **mutations et sélections**, jusqu'à ne conserver que les BCR qui ont la plus **grande affinité pour l'antigène**.
- Grâce aux différentes mutations, il est possible de **reconstruire la lignée évolutive des cellules B** à partir de la séquence d'ADN non mutée (Abdollahi et al., 2022).

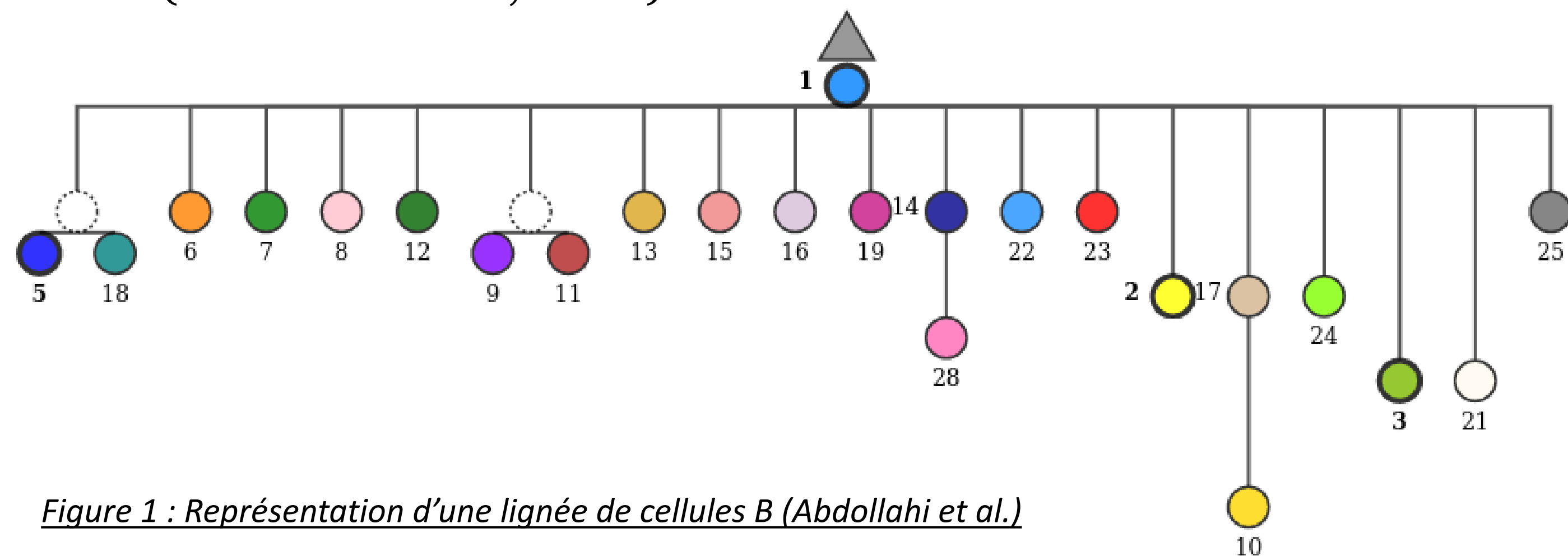


Figure 1 : Représentation d'une lignée de cellules B (Abdollahi et al.)

Une nouvelle métrique

La distance $LS(+)$ entre 2 arbres T_1 et T_2 est définie comme suit :

$$LS(+)(T_1, T_2) = p \sum_{i=1}^{n-1} |w(i) - \gamma(i)|^2 \sum_{j=i+1}^n |d(i, j) - \delta(i, j)|^2$$

p est une pénalité égale à $\frac{\text{nb nœuds en commun}}{\text{nb total nœuds}}$

n est le nombre de nœuds dans les arbres T_1 et T_2 .

On traduit l'abondance des séquences dans le répertoire par un poids : $w(i)$ est le poids du nœud i dans l'arbre T_1 et $\gamma(i)$ dans l'arbre T_2 .

On calcule la différence entre la distance entre les nœuds i et j dans l'arbre T_1 ($d(i, j)$) et dans l'arbre T_2 ($\delta(i, j)$).

Objectif et applications

Créer des clusters d'arbres de lignées de cellules B afin de :

contribuer à l'élaboration de nouvelles **immunothérapies**

favoriser l'avancement de la **médecine personnalisée**

mieux comprendre les **maladies lymphoprolifératives**

améliorer la **prédiction in silico de l'affinité des anticorps**

Implémentation

Programme de clustering d'arbres phylogénétiques **KMeansSuperTreeClustering** (Tahiri et al., 2022) en C/C++

→ intègre les distances : RF, LS, Quartet, Dissimilarité de Bipartition

1. Ajout de la lecture d'arbres de lignées + ajout métrique $LS(+)$

Par la suite...

2. Tests sur des données **simulées**

3. Validation sur des données **réelles**

Problématique

Pour créer un cluster il faut **mesurer la distance entre les arbres** avec une **métrique**. Il existe des métriques pour les **arbres phylogénétiques** (Tahiri et al., 2018, Tahiri et al., 2022).

Or les arbres de lignées de cellules B présentent des particularités :

nœuds internes de **degré 2**

multifurcations

racine connue (séquence non mutée)

Séquences **intermédiaires connues**

abondance de la séquence



Notre laboratoire analyse l'évolution des espèces et évalue leurs impacts sur l'environnement en combinant, entre autres, des informations issues de la génétique des espèces et des paramètres climatiques (<https://tahirinadia.github.io>).

Cette étude est soutenue par :



RÉFÉRENCES

- Nika Abdollahi, Lucile Jeusset, Anne de Septenville, Frédéric Davi, and Juliana Silva Bernardes. Reconstructing the evolutionary history of a b cell lineage with minimum spanning tree and genotype abundances. 2022.
- Nadia Tahiri, Bernard Fichet, and Vladimir Makarenkov. Building alternative consensus trees and supertrees using k-means and robinson and foulds distance. *Bioinformatics*, 38(13):3367–3376, 2022.
- Nadia Tahiri, Matthieu Willems, and Vladimir Makarenkov. A new fast method for inferring multiple consensus trees using k-medoids. *BMC evolutionary biology*, 18:1–12, 2018.