

# Debrief du vendredi 20/01

## 1 Récapitulatif de la journée

Nous avons dégagé plusieurs pistes de recherche avec G. Fertin et G. Jean :

- il serait intéressant d’analyser les performances d’une première version de la méthode séquentielle décrite dans le debrief du vendredi dernier. Cette première version sera naïve et écartera tout baitModel trop ambigu. Éventuellement, cette version préliminaire ne renverra pas de solution (si tous les baitModels considérés sont ambigus alors ils seront tous ignorés). Cette implémentation sera à tester sur les 10 000 baits fournis dans le fichier de données. Nous cherchons surtout à savoir pour combien de baits nous arrivons à obtenir un baitModel “fusionné”. Lorsqu’un tel baitModel est obtenu nous quantifierons sa proximité avec le bait.
- un premier état de l’art sur les méthodes d’alignement de séquences multiples (ou MSA pour “Multiple Sequence Alignment”) doit être dressé pour discuter de la viabilité de la méthode de fusion de baitModels utilisant l’alignement de séquences (voir debrief du vendredi 13/01). De plus, des tests utilisant au moins une méthode de la littérature sont attendus pour déterminer comment étendre ces méthodes à notre problème.

Dans ce debrief, nous présenterons tout d’abord un état de l’art des méthodes MSA pour ensuite interpréter les résultats obtenus avec les implémentations de quelques méthodes de fusion de baitModels.

## 2 Alignement de séquences

Pour déterminer la similarité entre des séquences, les biologistes utilisent des méthodes d’alignement de séquences. Nous verrons qu’il existe plusieurs types d’alignements et plusieurs approches pour résoudre le problème d’alignement. Cependant, ce problème est NP-difficile et la complexité des algorithmes pour résoudre ce problème à l’optimal est souvent exponentielle. Dans notre cas, les baitModels sur lesquels nous travaillons sont souvent courts et peu nombreux. Par conséquent, on peut envisager d’utiliser les méthodes plus lourdes en termes de calculs mais dont la qualité des solutions est meilleures.

### 2.1 Alignement pair-à-pair

Dans un premier temps, nous présenterons les méthodes d’alignement de séquences pair-à-pair. Ces méthodes sont utilisées pour évaluer les similarités entre deux séquences. Nous serons amenés à fusionner plus de deux baitModels mais les méthodes pair-à-pair sont utilisées par certaines méthodes d’alignement multiple et sont, historiquement, les premières méthodes qui ont été étudiées pour le problème d’alignement.

!! SOURCES!!

-> sous-section sur les types d'alignement -> présentation des algorithmes

## **2.2 Alignement multiple**

-> sous-sections sur les différents types d'algos

==> section sur les algos implémentés ==> pq l'alignement local semble être un meilleur choix que l'alignement global