

Résumé de projet : Sharane K.Murali (3701883)

I – Objectifs du projet, points de vue biologique et technique

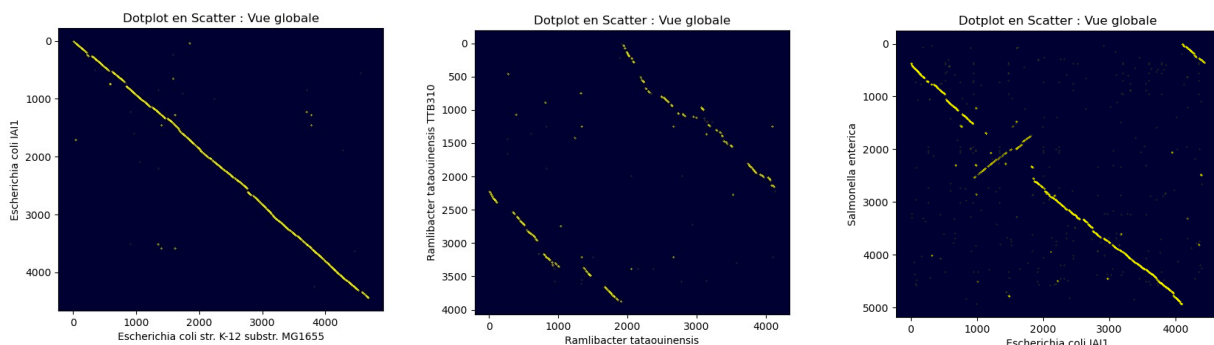
Le but du projet est de créer une interface graphique associée à une base de données de protéomes bactériens dans le but de visualiser la répartition des gènes homologues à travers un dotplot. Il est également censé identifier les blocs de synténie, la finalité étant d'observer la facette évolutive de la répartition des gènes.

D'un point de vue biologique, on devrait observer une diminution des blocs de synténie au fur et à mesure qu'on s'éloigne entre deux espèces au niveau taxonomique et du coup une diminution de la conservation des répertoires génomiques. Nous allons nous focaliser sur les protéomes initialement fournis et autres pour visualiser cette perte de conservation.

D'un point de vue technique et logistique, il est nécessaire de savoir les objets qui sont manipulés pour arriver aux résultats que nous souhaitons obtenir. Pour cela, une interface graphique codée sous Tkinter avec une base de données en PostgreSQL sont utilisées. Dans l'application, l'utilisateur peut exécuter les BLAST entre les protéomes bactériens présents dans la base avec un e-value à $1e-4$, ceux-ci sont injectés dans la base pour pouvoir être testés avec différentes conditions, valeurs de seuil et mode de représentation : apprécier globalement une synténie avec un scatter, visualiser à l'échelle près avec un imshow ou récupérer la liste des homologues dans un fichier texte pour plus de détails. L'affichage du dotplot se fait exclusivement avec matplotlib.

II – Résultats et interprétation

Voici les résultats obtenus suite à l'exécution de la création de dotplot pour les protéomes fournis



La figure de gauche représente le dotplot entre deux souches au sein d'une même espèce bactérienne intestinale (*Escherichia coli*) et on peut constater qu'il y a une grande conservation de l'ordre, une synténie parfois discontinue avec quelques duplications minimales qui peuvent être visibles avec le plot imshow. La taille de fenêtre était de 10, le seuil de stringence à 4 et le seuil d'e-value à $1e-30$.

La figure au centre représente le dotplot entre le protéome des bactéries *Ramlibacter sp. 5-10* et *Ramlibacter tataouinensis TTB310* avec une taille de fenêtre de 5, un seuil de 3 et un seuil d'e-value à $1e-30$, car il y avait beaucoup de bruit. Ces deux bactéries font partie du genre *Ramlibacter*. On peut observer quelques blocs de synténie contenant beaucoup de discontinuités. Ce qui est intéressant ici est la double relocalisation entre ces deux espèces différentes : cela nous laisserait supposer d'un point de vue évolutif qu'il y a eu une cassure dans la conservation et une relocalisation de la moitié du protéome chez les deux espèces. Cette conservation est faible lorsqu'on compare avec le dotplot précédent.

La figure de droite représente le dotplot entre le protéome des bactéries *Escherichia coli IAI1* et le protéome *Salmonella enterica* avec une taille de fenêtre de 10, un seuil de 2 et un seuil d'e-value à $1e-130$. Ces deux espèces appartiennent à la famille des *Enterobacteriaceae* mais sont issues de genres différents. On peut observer des grands blocs de synténie discontinus au début et à la fin. On peut également observer une relocalisation d'une partie du protéome ainsi qu'une inversion qui est approximativement de taille 1000 soit un quart du protéome d'*E.coli*. On peut constater une conservation plus faible que celle vue en premier lieu.

III – Conclusion et améliorations éventuelles

En terme de conclusion, on a pu observer qu'au sein d'une même espèce, l'ordre des gènes est généralement conservé, beaucoup moins au sein d'un genre et encore moins entre genres différents. Ici, le dernier plot a permis de pointer du doigt qu'il y a aussi des espèces avec des genres différents qui ont une grande conservation de l'ordre des gènes. Avec ce type d'approche, il est possible de retracer une histoire évolutive au sein des bactéries.

En guise d'amélioration, une interface moderne avec la librairie « customtkinter », l'insertion de l'annotation fonctionnelle COG dans la base et dans les critères ainsi qu'une méthode plus interactive pour cibler les blocs de synténie et extraire de manière plus pertinente les régions homologues seraient de bonnes pistes.