

748 - Projet : fiche résumé

I. Stratégie adoptée :

La réalisation de ce projet s'est faite en plusieurs étapes. En premier j'ai configuré ma base de donnée, ensuite j'ai créé des scripts permettant de remplir cette base grâce aux fichiers donnés sur moodle. Lorsque j'ai obtenu une base remplie avec quelques BLAST j'ai commencé à coder mon interface graphique permettant de lire les données et de ressortir un dotplot. Enfin, j'ai fusionné tous mes scripts et ai ajusté quelques éléments afin que mon application fonctionne chez tous les utilisateurs et que ma base de donnée puisse être mise à jour simplement.

Dans les choix que j'ai effectué, il y a celui d'utiliser plt.scatter(), en effet j'ai rencontré de nombreux problèmes de lisibilité avec plt.imshow() ce qui me conduit à utiliser un scatter, moins joli mais plus lisible.

Aussi, ce dont je suis le plus fier dans mon code, c'est de m'être servis, dans mes gros algorithmes, uniquement de matrices numpy, que ce soit pour appliquer les critères d'homologie, ou pour mettre à jour le dotplot avec la fenêtre et la stringence. Ceci m'a permis d'optimiser la vitesse de mon code.

II. Résultats Obtenus (description plus détaillée dans l'annexe)

Afin de m'interroger sur la conservation des répertoires génomiques, j'ai effectué des BLAST sur plusieurs couples d'organismes. Pour tous les résultats présentés, les paramètres seront les mêmes : **E-value = 1e-30, P-ident = 80, Q-cover = 0.5, S-cover = 0.5**. J'ai séparé ces résultats en 3 catégories : même espèce, même genre et espèces plus éloignées.

Les souches K-12 (substr. MG1655) et IAI1 de l'espèce **Escherichia coli** (**fig1**), ainsi que les souches 5-10 et TTB310 de l'espèce **Ramlibacter tataouinensis** (**fig2**), m'ont servis d'exemple pour étudier la synthénie entre deux souches de la même espèce.

Trois organismes du genre **Bartonella** m'ont servis d'exemple pour des organismes du même genre mais pas de la même espèce. On remarque une grande différence entre les deux dotplots, pour le premier on voit un nombre de hit considérablement bas (205) alors que pour le deuxième, on peut voir un nombre considérable de hit (920) ce qui est, plus que pour le dotplot entre les souches de **Ramlibacter tataouinensis**.

Enfin les organismes **Magnetospirillum magneticum** AMB-1 et **Magnetococcus marinus** MC-1 sont deux alphaprotéobactéries qui n'appartiennent pas au même genre. Leur dotplot a révélé 0 hits.

III. Conclusion

Des études complémentaires auraient permis d'établir une vraie conclusion, avec des algorithmes concrets permettant de mesurer la synthénie (autrement que visuellement). Aussi, il est difficile en prenant un seul exemple par catégorie de tirer une véritable conclusion sur la conservation du répertoire génomique au sein d'espèces, ou de genres. Il aurait fallu faire une moyenne entre plusieurs souches d'une même espèce ou du même genre. Cependant nous pouvons quand même réagir aux résultats que nous avons obtenus, en gardant en tête qu'il ne s'agit pas de vérité générale.

On a donc vu qu'au sein d'une même espèce il peut y avoir une très forte densité d'homologie avec l'exemple des souches d'**Escherichia coli** qui présentent une forte homologie. Aussi, entre deux espèces du même genre on a vu qu'on pouvait reconnaître une certaine conservation mais que celle-ci était beaucoup moins importante qu'entre deux individus de la même espèce. Enfin entre deux espèces plus lointaines taxonomiquement, on n'a observé aucune synthénie comme on pourrait s'y attendre à priori.

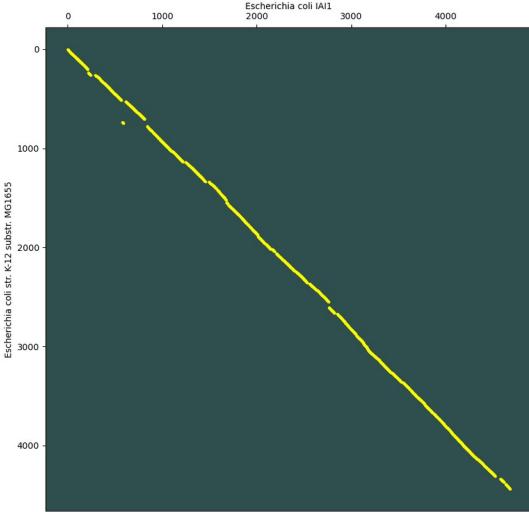


fig1 : Dotplot réalisé entre les souches K-12 (substr. MG1655) et IAI1 de l'espèce *Escherichia coli*. Il y a en tout 4510 hits. On peut voir très nettement que les points forment une belle diagonale ce qui indique une forte conservation du répertoire génomique entre les deux souches et donc une forte synthénie.

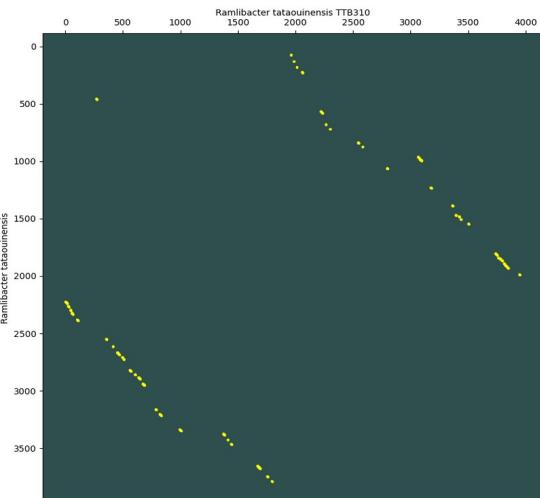


fig2 : Dotplot réalisé entre les souches 5-10 et TTB310 de l'espèce *Ralibacter tataouinensis*. Il y a en tout 469 hits ce qui montre une synthénie bien plus faible que pour les souches d'***Escherichia coli***. Cependant on peut remarquer une certaine similitude. Deux diagonales séparées avec une densité d'homologie plus élevée que le reste se forment. Ceci est sûrement un indicateur qu'il y a eu une inversion dans le génome d'un des organismes par rapport à leur ancêtre commun.

NB : Je ne suis pas sûr que ces deux organismes appartiennent à la même espèce, car dans l'énoncé on nous dit qu'il s'agit d'espèces différentes alors que dans le fichier 'prokaryotes_complete-genomes.csv' on nous dit que ce sont deux souches d'une même espèce.

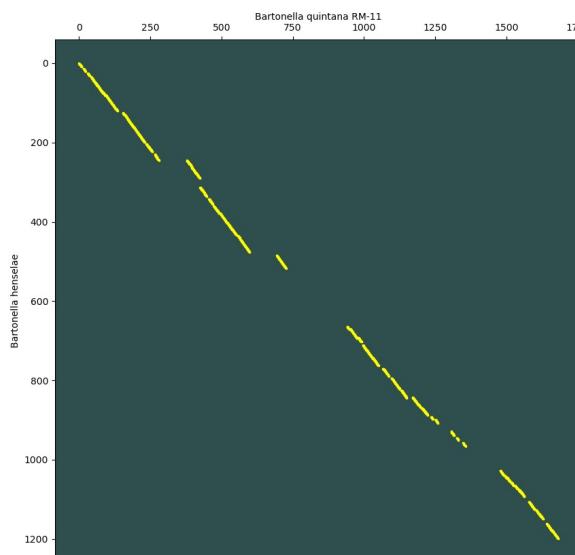
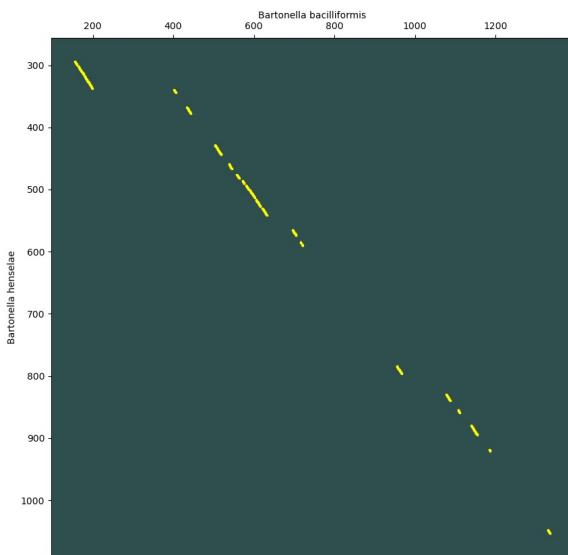


fig3 : Dotplot entre plusieurs espèces du genre *Bartonella*. On peut voir qu'appartenir à un même genre ne veut pas obligatoirement dire avoir une synthénie élevée. Pour les deux organismes à gauche (***Bartonella henselae*** et ***Bartonella bacilliformis***) il y a une synthénie beaucoup plus faible que pour les deux organismes à droite (***Bartonella henselae*** et ***Bartonella quintana*** RM-11), il y a 205 hits contre 920 respectivement.