



Projet python : Observer la synténie des espèces

Ihsân Grichi

2023



Plan

I - Objectif du projet et question biologique

II - Méthode pour y répondre (aspect technique)

III - Résultats et conclusions

IV - Possibles améliorations et questionnement

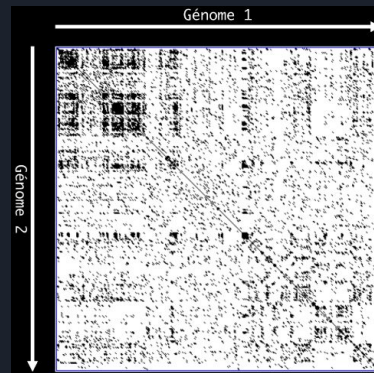
Question biologique

La synténie est-elle plus ou moins conservée chez deux souches différentes d'une même espèce ?

Synténie ? Ordre des gènes/protéines chez un être vivant

Parenté ? Au niveau évolutif

Stratégie pour évaluer synténie de deux espèces => Dot plot

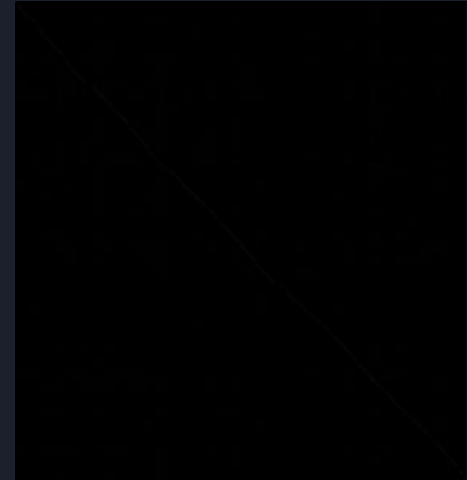


Comment faire des dot plot ?

À partir des hits de BLASTp entre deux fichiers .faa (protéomes)

Comparer deux protéomes :

- Protéines considérés homologues selon plusieurs critères
- Protéines homologues = points sur le dot plot
- Synténie observée à l'oeil nu

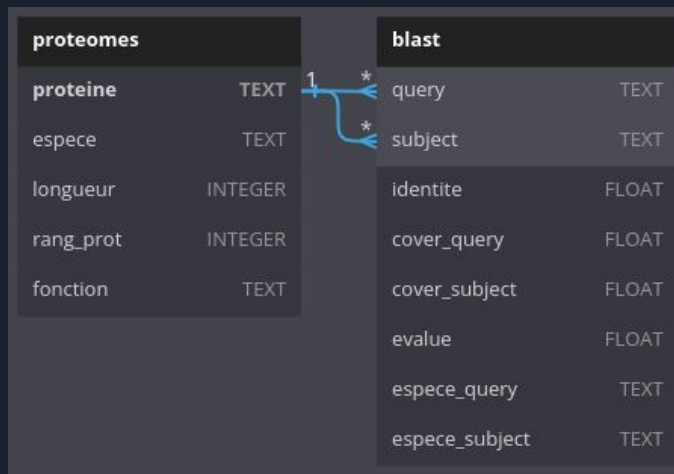


Dot plot entre deux E.coli (K-12 et IA1)

À partir de quelles données ?

proteine	espece	longueur	rang_prot	fonction
CP000471.1_prot_ABK42530.1_1	M. marinus MC-1	447	1	Chromosomal replication initiation ATPase DnaA
CP000471.1_prot_ABK42531.1_2	M. marinus MC-1	376	2	DNA polymerase III sliding clamp (beta) subunit, PCNA homolog

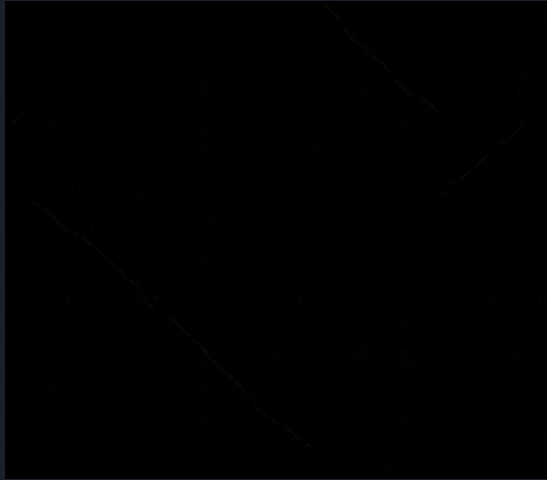
query	subject	evalue	cover_query	cover_subject	identite	espece_query	espece_subject
CP000471.1_prot_ABK42530.1_1	AP007255.1_prot_BAE49440.1_636	8e-124	0.987	0.962	41.31	M. marinus MC-1	M. magneticum AMB-1
CP000471.1_prot_ABK42530.1_1	AP007255.1_prot_BAE49256.1_452	7e-13	0.208	0.576	31.25	M. marinus MC-1	M. magneticum AMB-1





Démonstration

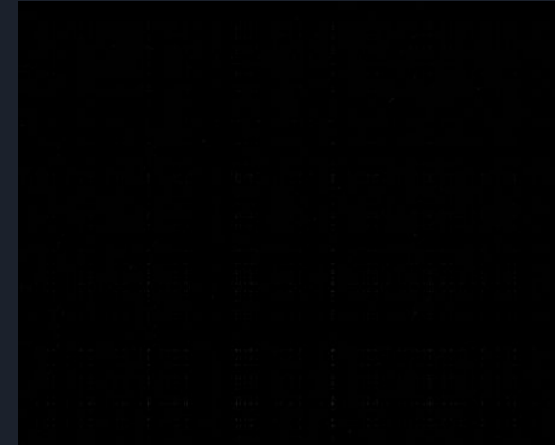
Résultats



Dot plot entre 2 E.coli



Dot plot entre deux Ramlibacter



Dot plot entre deux espèces différentes



Conclusions

La synténie est-elle plus ou moins conservée chez deux souches différentes d'une même espèce ?

Réponse : OUI

MAIS étude portant sur peu de cas => pas de conclusions définitives à apporter

critères trop arbitraires et pas assez sérieux pour être viables (contexte biologique ?)



Problèmes et améliorations possibles

Outil fenêtre et seuil de stringence trop lent => réfléchir à un autre algo (numba ?)

Utilisation de imshow => utilisation de scatter en parallèle de l'amélioration précédente

Manque d'infos et de contexte bio pour chaque espèce => un onglet infos sommaires sur chaque espèce ?

Absence du critère COG => l'ajouter par un meilleur traitement de fichier

GUI trop lente => télécharger fichier .out au lieu de le créer ?

=> Implémenter une fonctionnalité permettant de bien visualiser les blocs de synténie



Librairies utilisées

- re, os, gzip, subprocess, glob, tkinter, numpy, matplotlib lib et numba

=> bibliothèques natives de python

- ftplib et psycopg2

=> bibliothèques à installer

```
import psycopg2, re, os, gzip, subprocess, glob
import tkinter as tk
from tkinter import ttk
import numpy as np
import matplotlib.pyplot as plt
from ftplib import FTP
from numba import njit
```



Merci pour votre attention