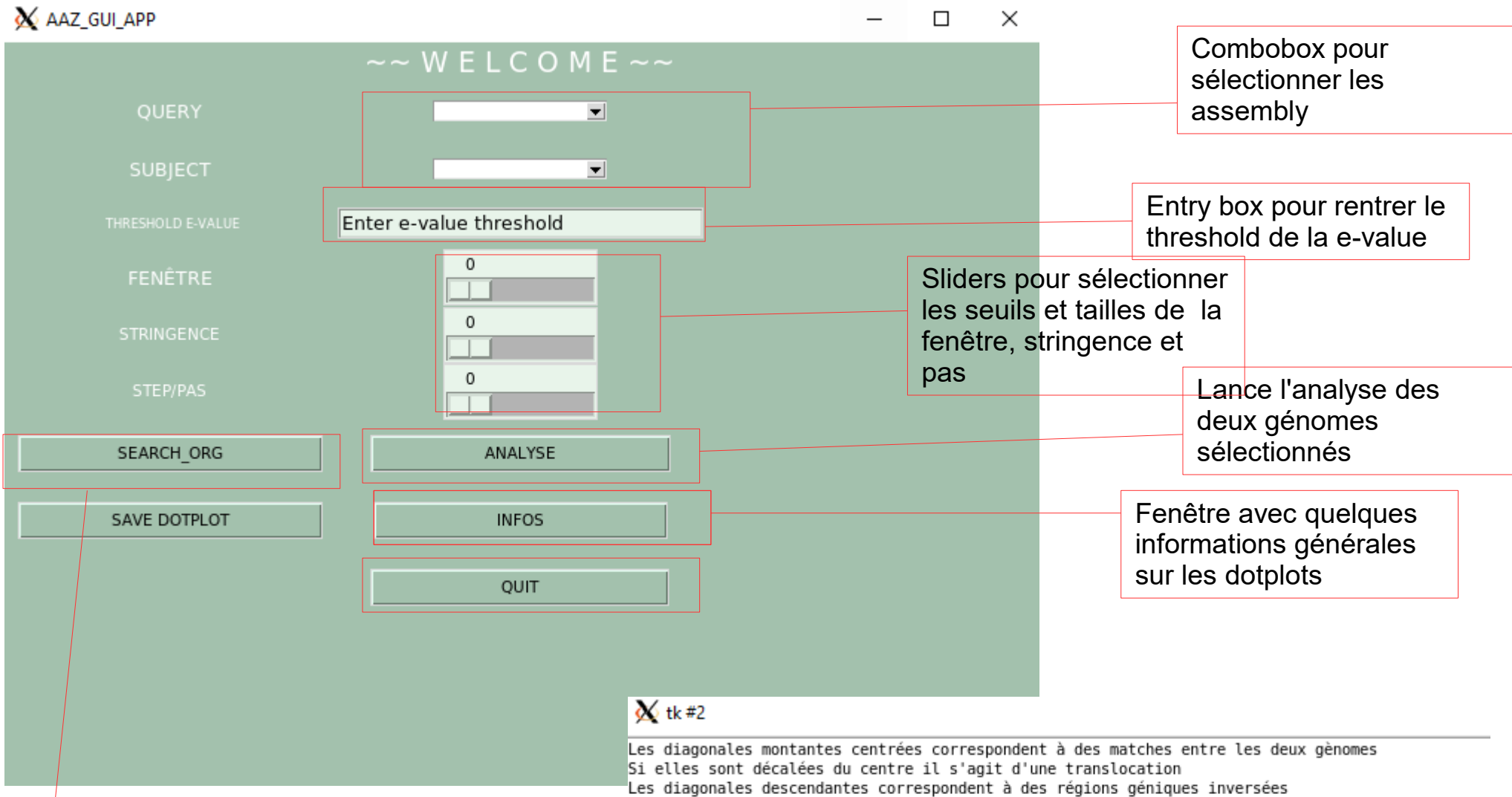


MU4BM748 : PROGRAMMATION AVANCE EN PYTHON – ALGORITHMES D'ANALYSE DE SEQUENCES

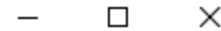
**PROJET : COMPARAISON DE PAIRES DE GENOMES
BACTERIENS PAR DOTPLOTS – FAMILLE DE GENES ET
BLOCS DE SYNTHENIE**

INTERFACE GRAPHIQUE



INTERFACE GRAPHIQUE

SEARCH.ORG



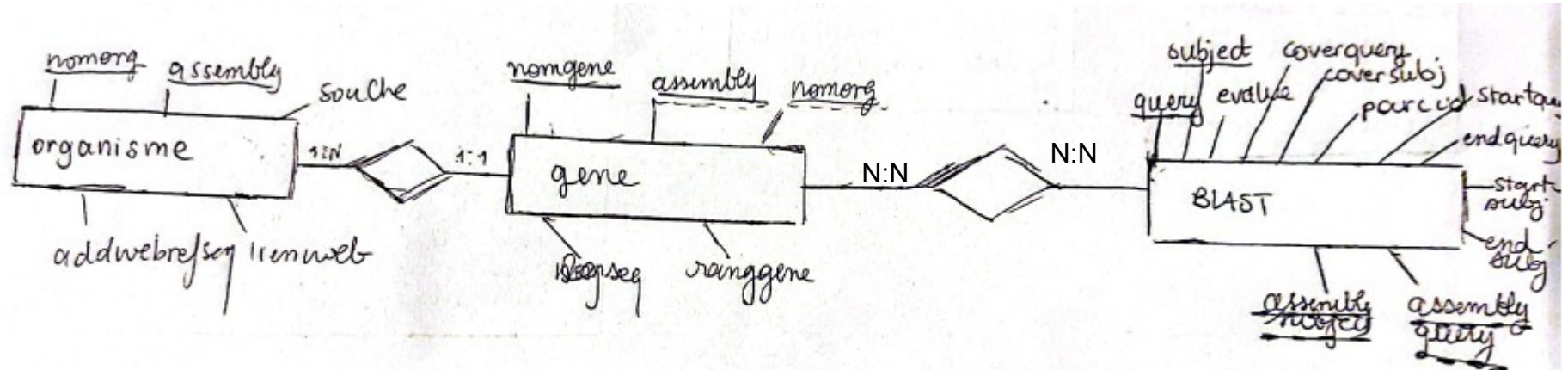
Search Organism by name

Escherichia coli

Search

```
('Escherichia coli', 'SQ171', 'GCA_000988445.1')
('Escherichia coli', 'JJ1897', 'GCA_001513655.1')
('Escherichia coli', 'ST540', 'GCA_000597845.1')
('Escherichia coli', 'ST540', 'GCA_000599625.1')
('Escherichia coli', 'ST540', 'GCA_000599645.1')
('Escherichia coli', 'ST2747', 'GCA_000599665.1')
('Escherichia coli', 'ST2747', 'GCA_000599685.1')
('Escherichia coli', 'ST2747', 'GCA_000599705.1')
('Escherichia coli', 'ECONIH1', 'GCA_000784925.1')
('Escherichia coli', 'RM9387', 'GCA_000801165.1')
('Escherichia coli', '94-3024', 'GCA_000801185.2')
('Escherichia coli', '6409', 'GCA_000814145.2')
('Escherichia coli', '789', 'GCA_000819645.1')
('Escherichia coli', 'C41(DE3)', 'GCA_000830035.1')
('Escherichia coli', 'BL21 (TaKaRa)', 'GCA_000833145.1')
('Escherichia coli', 'USML2', 'GCA_000833635.2')
('Escherichia coli', 'MNCRE44', 'GCA_000931565.1')
('Escherichia coli', 'K-12 substr. RV308', 'GCA_000952955.1')
('Escherichia coli', 'K-12 substr. HMS174', 'GCA_000953515.1')
('Escherichia coli', 'HUSEC2011', 'GCA_000967155.1')
('Escherichia coli', 'CI5', 'GCA_000971615.1')
('Escherichia coli', 'SEC470', 'GCA_000987875.1')
('Escherichia coli', 'SQ37', 'GCA_000988355.1')
('Escherichia coli', 'SQ88', 'GCA_000988385.1')
('Escherichia coli', 'SQ2203', 'GCA_000988465.1')
('Escherichia coli', 'CF5AN029787', 'GCA_001007915.1')
('Escherichia coli', 'NMEC 018', 'GCA_001021595.1')
('Escherichia coli', 'C43(DE3)', 'GCA_001039415.1')
('Escherichia coli', 'NCM3722', 'GCA_001043215.1')
('Escherichia coli', 'DH1Ec095', 'GCA_001183645.1')
('Escherichia coli', 'DH1Ec104', 'GCA_001183665.1')
('Escherichia coli', 'DH1Ec169', 'GCA_001183685.1')
('Escherichia coli', 'RR1', 'GCA_001276585.2')
('Escherichia coli', 'SF-088', 'GCA_001280325.1')
('Escherichia coli', 'SF-468', 'GCA_001280345.1')
('Escherichia coli', 'SF-166', 'GCA_001280385.1')
('Escherichia coli', 'SF-173', 'GCA_001280405.1')
('Escherichia coli', 'K-12 substr. MG1655_TMP32XR1', 'GCA_001308125.1')
('Escherichia coli', 'K-12 substr. MG1655_TMP32XR2', 'GCA_001308165.1')
```

BASE DE DONNEES



ORGANISME	
<u>nomorg</u>	varchar(500)
<u>assembly</u>	varchar(30)
<u>souche</u>	varchar(50)
<u>lienweb</u>	varchar(1100)
<u>addwebrefseq</u>	varchar(1000)

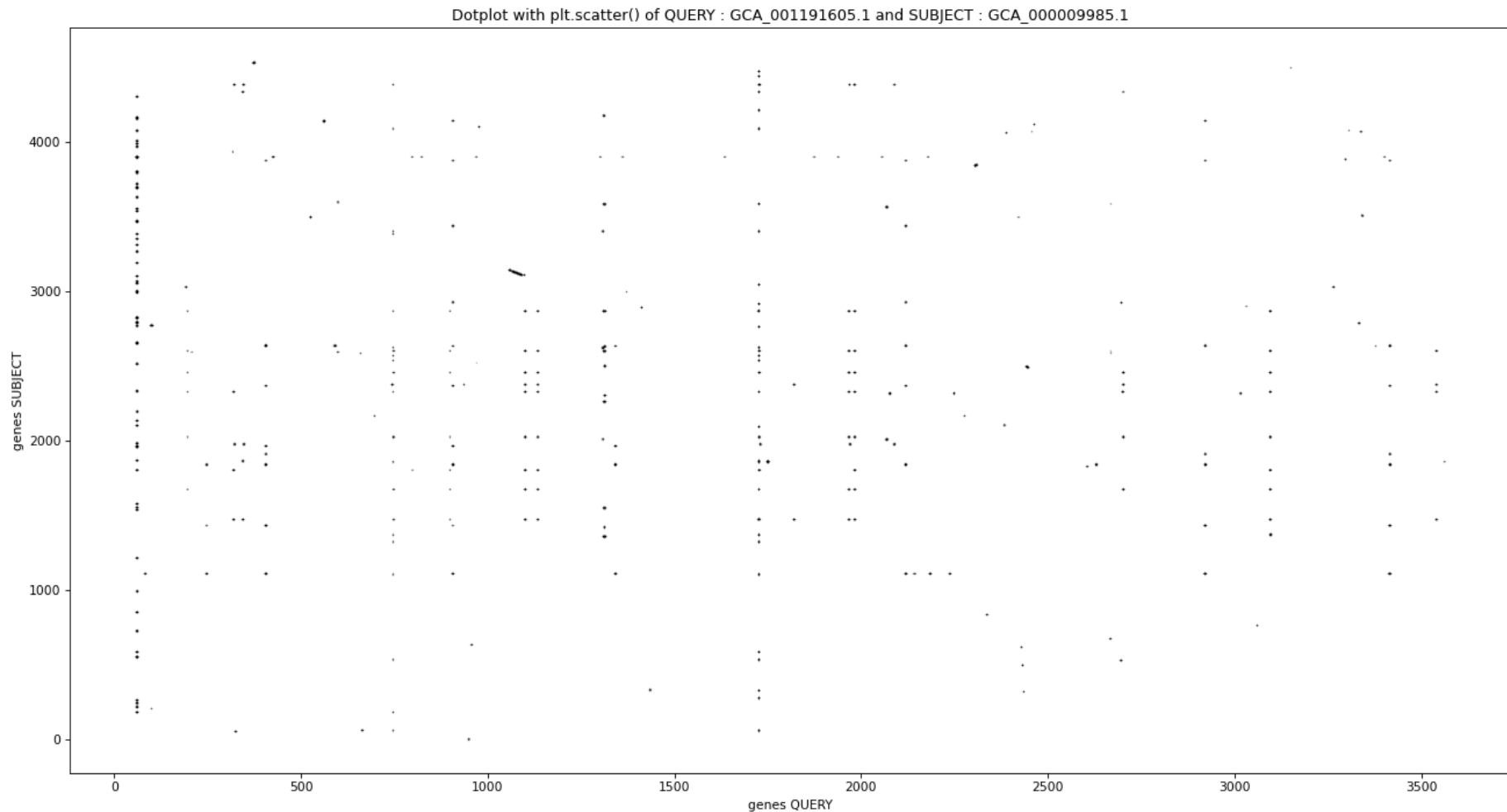
GENE	
<u>nomgene</u>	varchar(300)
<u>assembly</u>	varchar(50)
<u>nomorg</u>	varchar(500)
<u>ranggene</u>	smallint
<u>lg seq</u>	smallint

BLAST	
<u>query</u>	varchar(300)
<u>subject</u>	varchar(300)
<u>evalua</u>	float / double precision
<u>coverquery</u>	float
<u>cover subject</u>	float
<u>pourcid</u>	float
<u>start query</u>	smallint
<u>end query</u>	smallint
<u>start subj</u>	smallint
<u>end subj</u>	smallint
<u>assembly query</u>	varchar(50)
<u>assembly subject</u>	varchar(50)

Tables intermédiaires utilisées pour aider au remplissage

QUELQUES ANALYSES DE DOTPLOTS ...

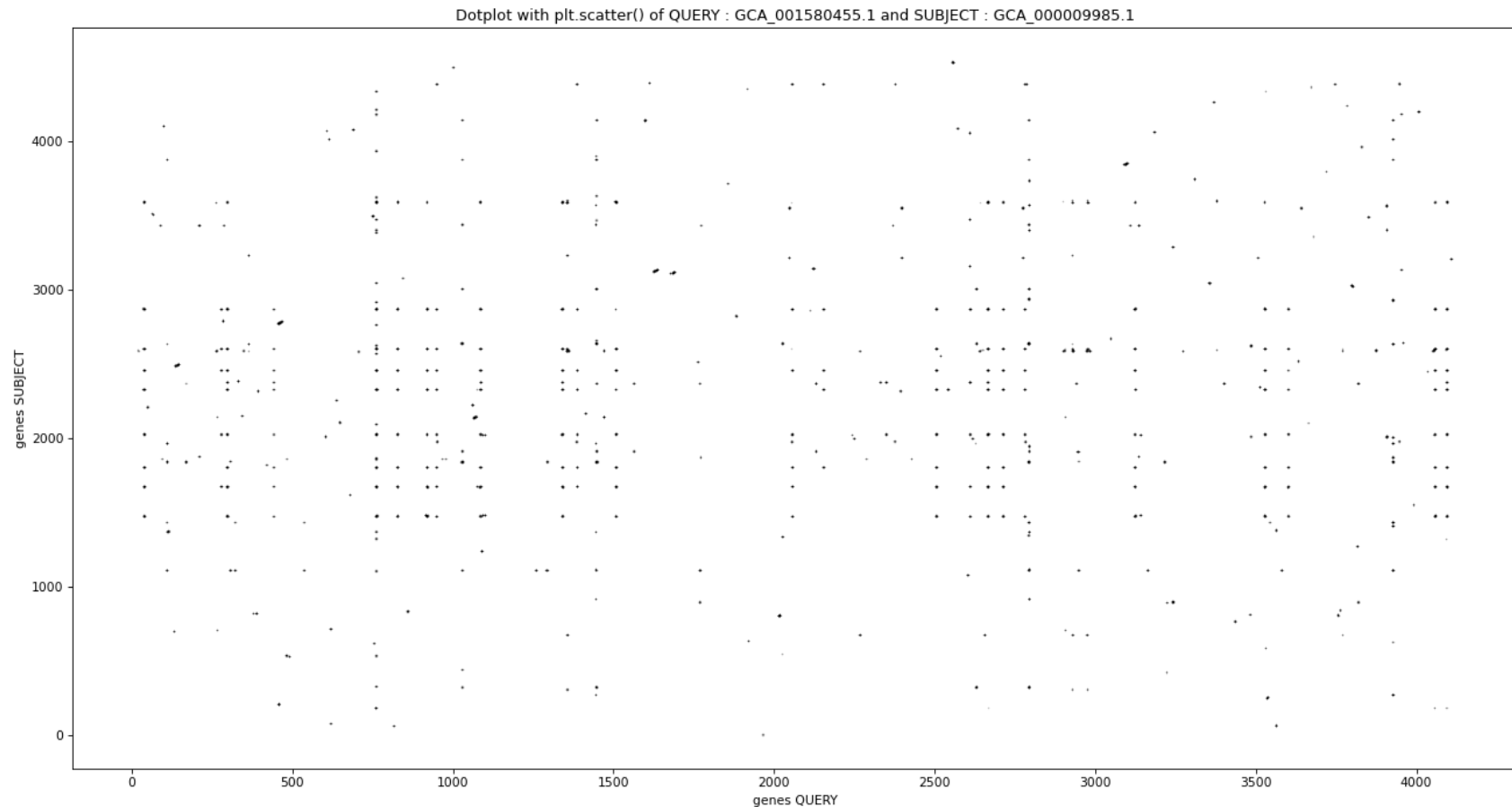
ORGANISMES APPARTENANT AU MÊME REGNE



Bacillus pumilus GR-8 et *Magnetospirillum magneticum* AMB-1
(*Bacteria*; *Terrabacteria* group; *Firmicutes* et *Bacteria*; *Proteobacteria*; *Alphaproteobacteria*)

QUELQUES ANALYSES DE DOTPLOTS ...

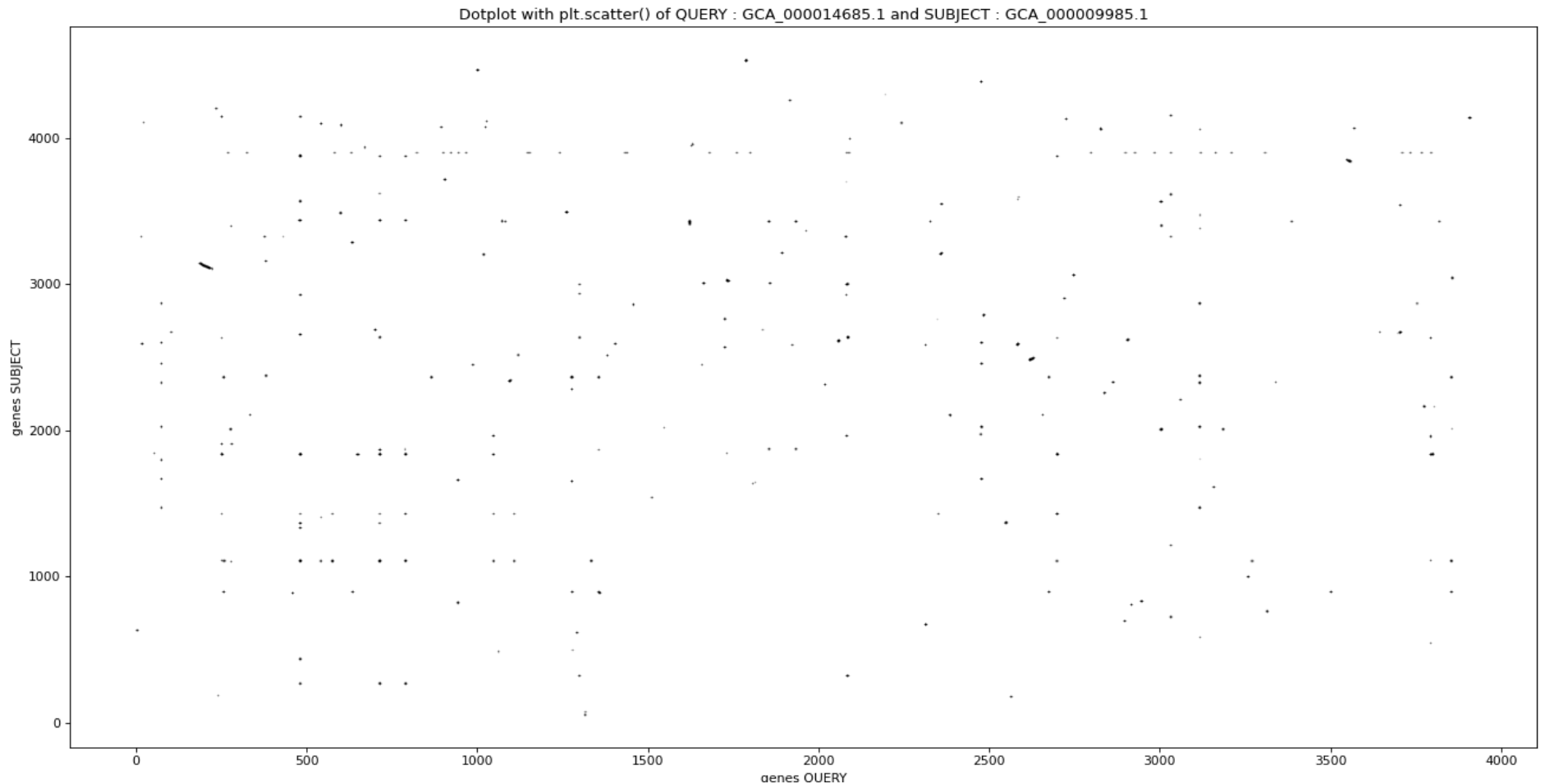
ORGANISMES APPARTENANT AU MÊME EMBRANCHEMENT



Ramlibacter tataouinensis et *Magnetospirillum magneticum* AMB-1
(*Bacteria;Proteobacteria;Betaproteobacteria* et *Bacteria;Proteobacteria;Alphaproteobacteria*)

QUELQUES ANALYSES DE DOTPLOTS ...

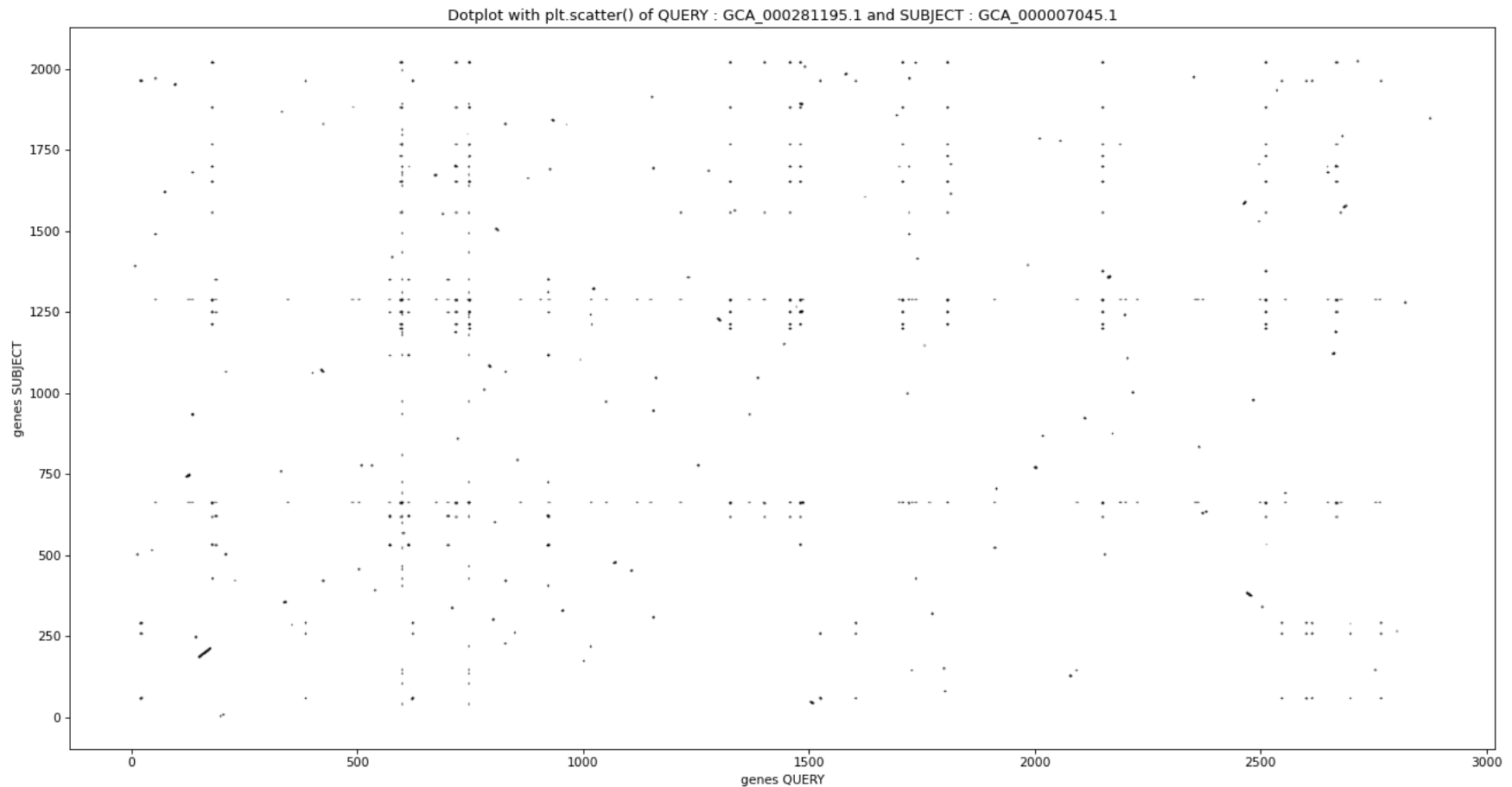
ORGANISMES APPARTENANT A LA MÊME CLASSE



Magnetococcus marinus MC-1 et *Magnetospirillum magneticum* AMB-1
(Bacteria; Proteobacteria; Alphaproteobacteria)

QUELQUES ANALYSES DE DOTPLOTS ...

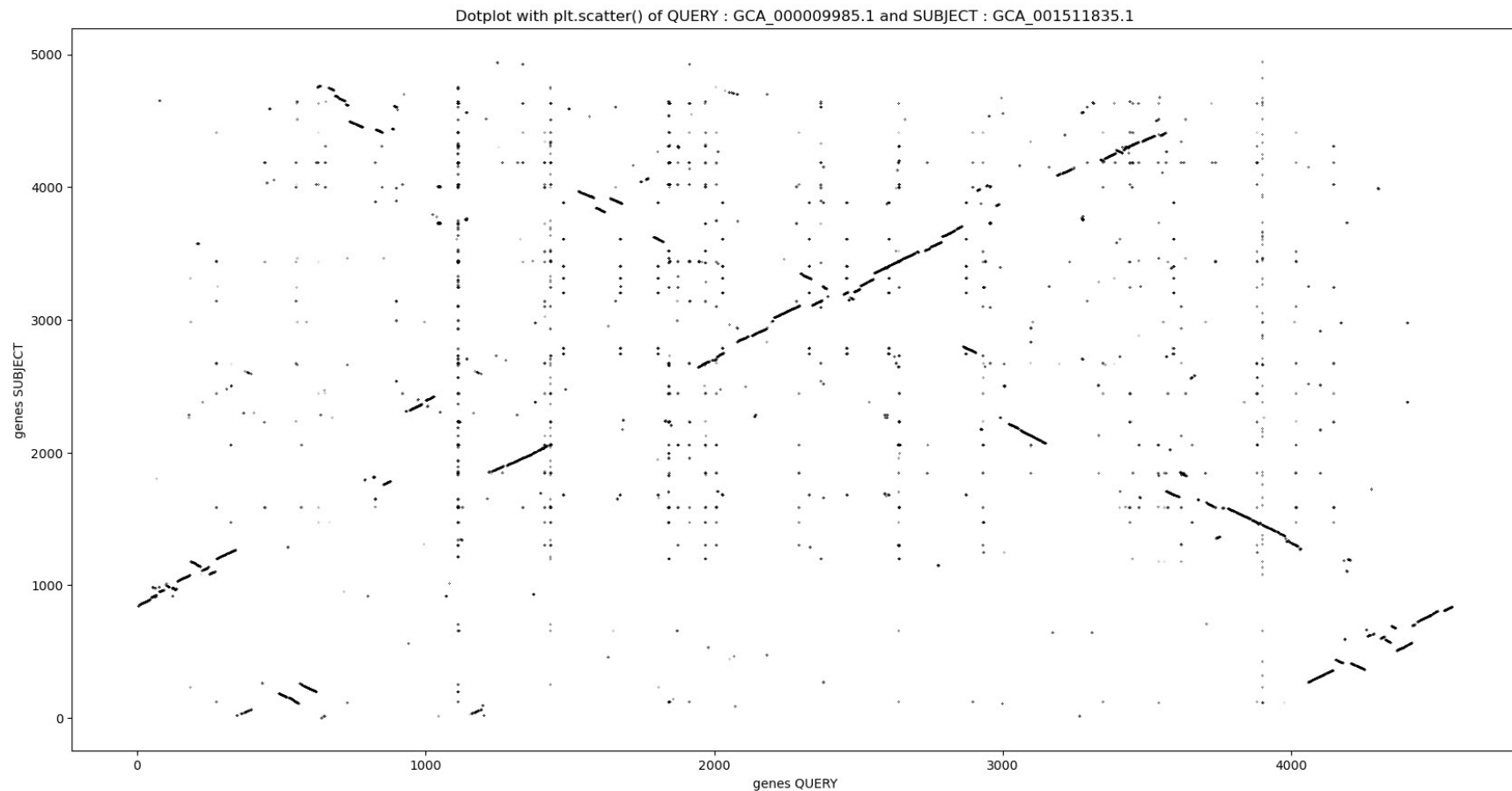
ORGANISMES APPARTENANT AU MÊME ORDRE



Magnetococcus marinus MC-1 et *Magnetospirillum magneticum* AMB-1
(*Bacteria*; *Proteobacteria*; *Alphaproteobacteria*)

QUELQUES ANALYSES DE DOTPLOTS ...

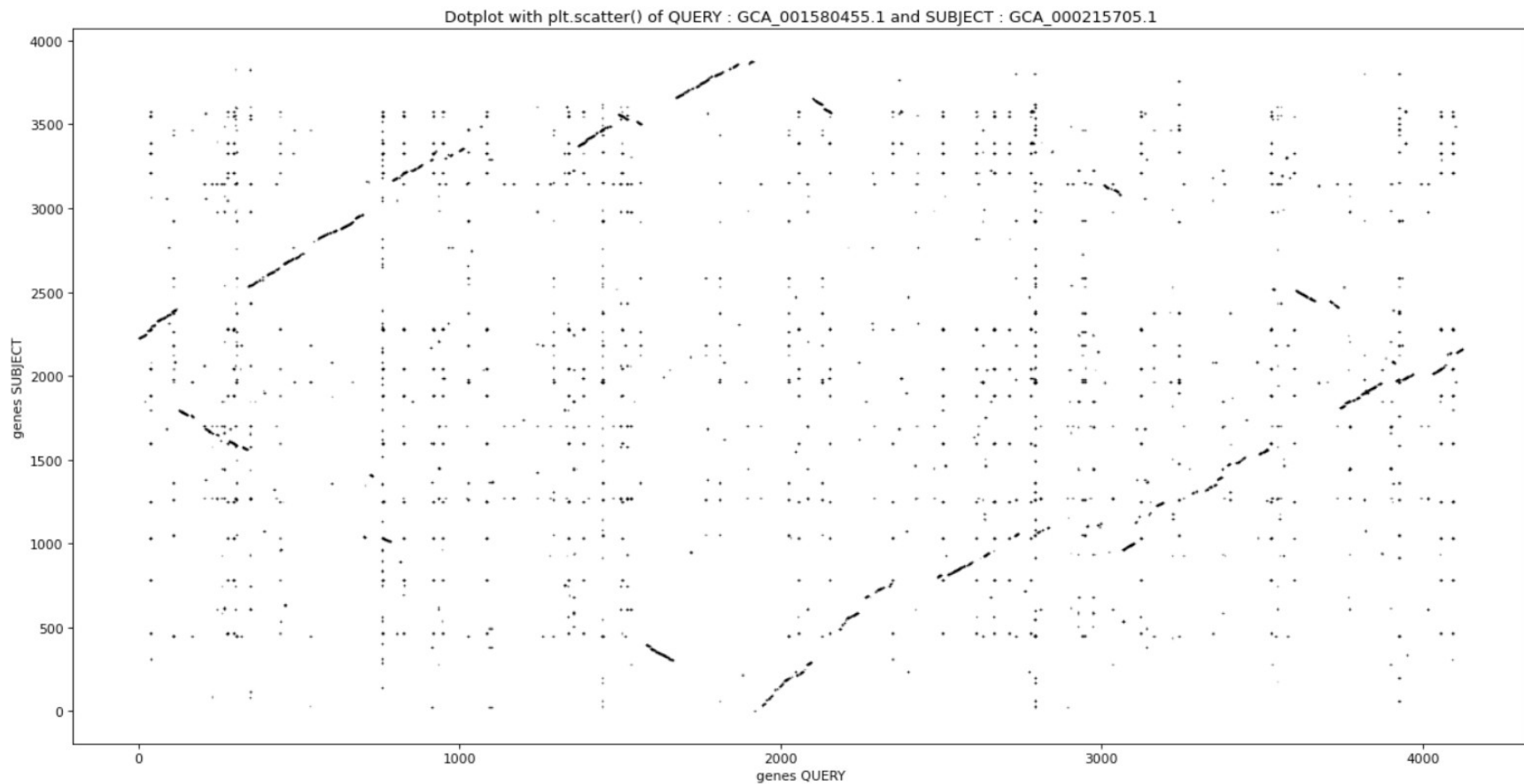
ORGANISMES APPARTENANT AU MÊME GENRE



Magnetospirillum magneticum AMB-1 et *Magnetospirillum* sp. XM-1
(Bacteria; Proteobacteria; Alphaproteobacteria)

QUELQUES ANALYSES DE DOTPLOTS ...

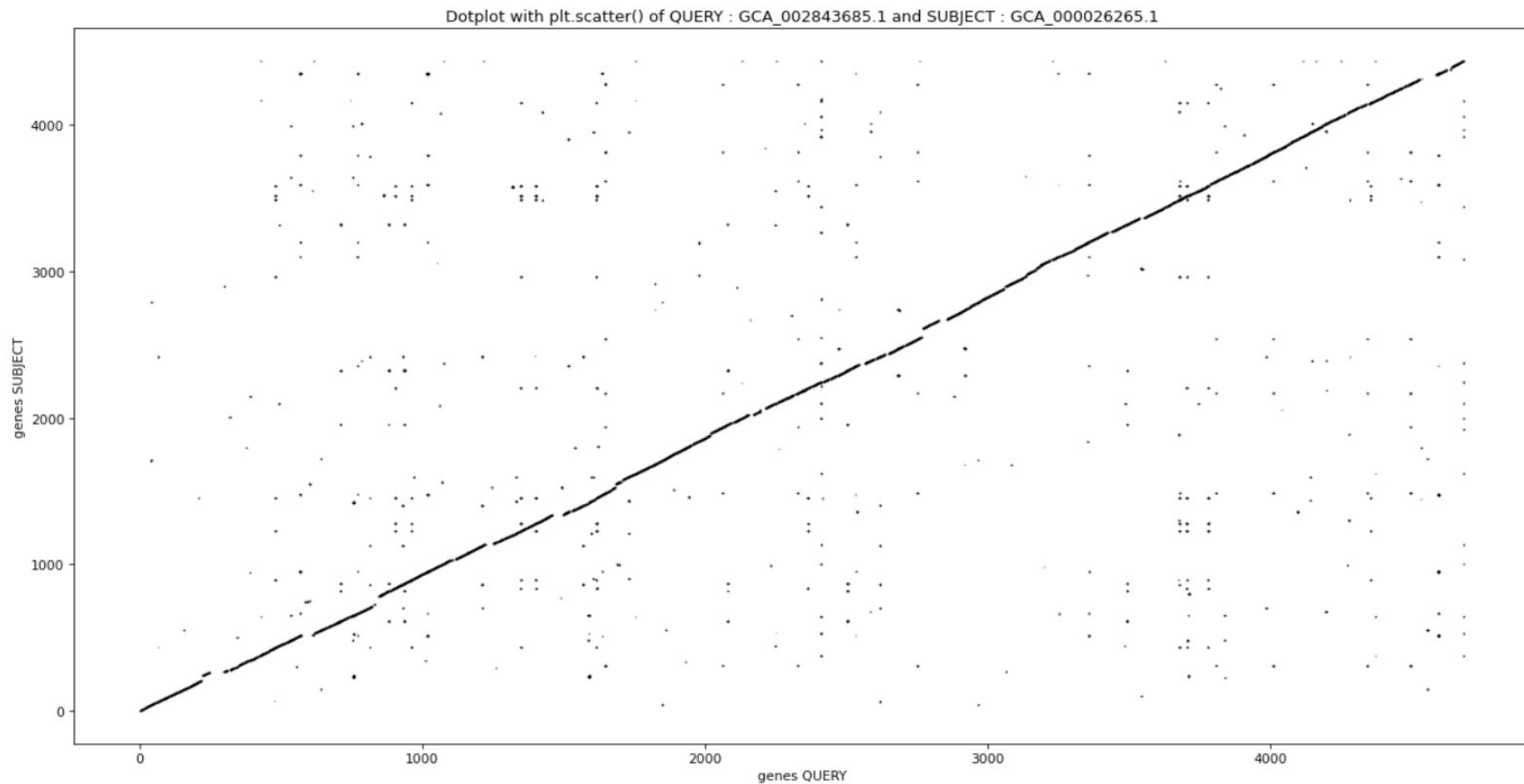
ORGANISMES APPARTENANT A LA MÊME ESPECE



Ramlibacter tataouinensis et *Ramlibacter tataouinensis* TTB310
Bacteria;Proteobacteria;Betaproteobacteria

QUELQUES ANALYSES DE DOTPLOTS ...

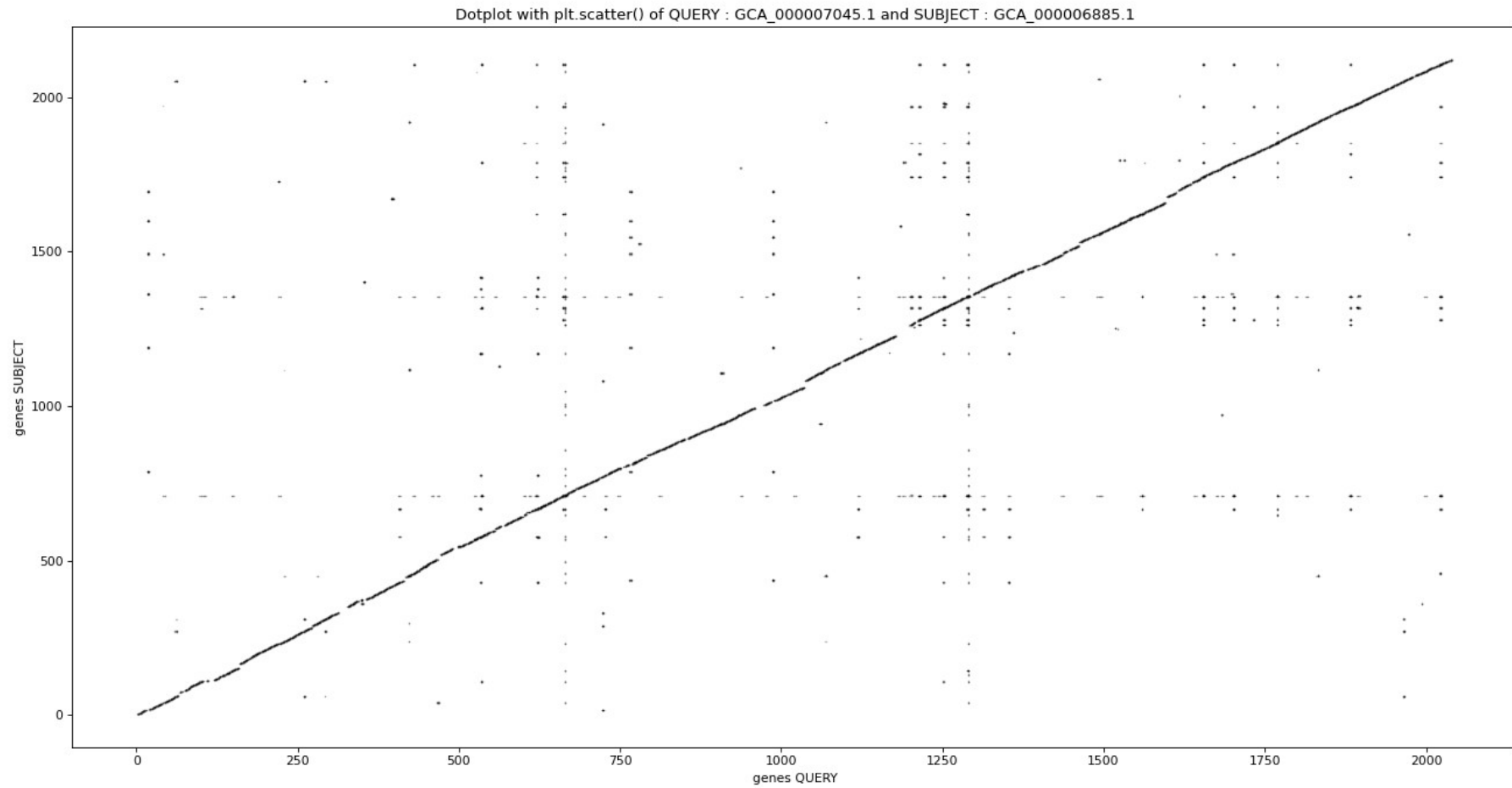
ORGANISMES APPARTENANT A LA MÊME ESPECE



Escherichia coli str. K-12 substr. MG1655 et *Escherichia coli* IA11
Bacteria;Proteobacteria;Gammaproteobacteria

QUELQUES ANALYSES DE DOTPLOTS ...

ORGANISMES APPARTENANT A LA MÊME ESPECE



Streptococcus pneumoniae R6 et *Streptococcus pneumoniae* TIGR4
Bacteria; Terrabacteria group; Firmicutes

CONCLUSION ET DIFFICULTES ET AMELIORATION

CONCLUSION

-Régions de synténie plus définie observées pour des organismes qui appartiennent au moins au **même genre**.

Selon les organismes → des résultats très différents.

→ si on teste deux organismes au hasard, on peut faire des hypothèses sur leur niveau de parenté.

DIFFICULTES

-Difficulté sur la compréhension de comment fonctionne l'algorithme avec l'histoire de fenêtre

-Difficulté sur la gestion du temps

-Difficulté sur certaines commande SQL dans le script python

-Difficulté sur l'analyse des dotplots (car je ne trouvais que des informations sur un dotplot de deux séquences)

AMELIORATION

-L'esthétique de l'interface graphique

-Bouton et fonction permettant pour pouvoir mettre à jour la base de données si on veut rajouter un organisme dans le base de données

-Faire les tests avec les cog-names

-Les informations pour aider la personne à analyser le dotplot

-Corriger des erreurs que je n'avais pas remarqué

-Rajouter la condition de filtrage avec pident (oubli)

MERCI DE VOTRE ATTENTION

BASE DE DONNEES

Tables intermédiaires utilisées pour aider au remplissage

```
aa_proj=> \d blastinterm
```

Table "public.blastinterm"				
Column	Type	Collation	Nullable	Default
qseqid	character varying(500)		not null	
sseqid	character varying(500)		not null	
pident	double precision		not null	
length	smallint		not null	
mismatch	smallint		not null	
gapopen	smallint		not null	
qstart	smallint		not null	
qend	smallint		not null	
sstart	smallint		not null	
send	smallint		not null	
eval	double precision		not null	
bitscore	smallint		not null	

```
aa_proj=> \d
```

List of relations			
Schema	Name	Type	Owner
public	blast	table	anastasia
public	blastinterm	table	anastasia
public	gene	table	anastasia
public	interm	table	anastasia
public	organisme	table	anastasia
public	temp	table	anastasia
public	test	table	anastasia
(7 rows)			

```
aa_proj=> \d temp
```

Table "public.temp"				
Column	Type	Collation	Nullable	Default
q	character varying(300)			
s	character varying(300)			
ev	double precision			
pid	smallint			
sq	smallint			
eq	smallint			
ssb	smallint			
esb	smallint			
qlg	smallint			
slg	integer			

```
aa_proj=> \d test
```

Table "public.test"				
Column	Type	Collation	Nullable	Default
q	character varying(300)			
s	character varying(300)			
ev	double precision			
cq	double precision			
cs	double precision			
pid	double precision			
sq	smallint			
eq	smallint			
ssb	smallint			
esb	smallint			
qlg	smallint			
slg	smallint			