

OveDB : A database integrating experimental and computational data of Corynebacterium pseudotuberculosis

Rapport final

Par

BENKACEM Soukaina

EL HADDAD Sana

Sous l’encadrement de :

Mme Juliana BERNARDEZ

**Table des matières**

# Contexte général du projet

## Présentation du projet

La Corynebacterium pseudotuberculosis est une bactérie pathogène qui cause plusieurs maladies chroniques infectieuses et contagieuses.

Ce pathogène Infecte principalement deux types de biovar : les Ovis et les Equis et peut conduire à la morte des animaux et à la réduction de la production de laine pour les ovins et les caprins, et il peut aussi diminuer la production du lait et de la viande pour les équidés et les bovins ce qui provoque des pertes économiques très importantes pour l’industrie agricole.

Récemment, une comparaison génomique, de 15 souches de C. pseudotuberculosis, a été réalisée [Soares et, al]. Cependant, les données expérimentales et computationnelles de cette étude n’ont pas été rassemblés et aucune base de donnée n’a été crée.

Notre projet consiste à élargir l’analyse réalisé dans [Soares et, al] pour considérer les génomes de 55 souches disponibles, et à construire une base de données qui permet de rassembler toutes les informations expérimentales et computationnelles de C. pseudotuberculosis qui va nous permettre par la suite de les interpréter et les analyser afin de

* Comprendre les différences entre les bactéries qui provoquent des maladie chez les différents biovars : Ovis et Equi
* Identifier des gènes orthologues et spécifique de chaque espèce
* Retrouver les gènes présentent dans des voie métaboliques associées à la maladie, etc.

Resultats :

A ce stade, notre base de données est déjà prête, elle contient toutes les informations nécessaires pour pouvoir les interpréter et aborder les analyses.

**Les singletons** :

Les singletons sont les gènes qui se trouvent dans une seule souche et qui sont absentes dans toutes les autres souches.

Pour calculer le nombre de singletons pour chaque souche, nous avons utilisé les requêtes suivantes :

Premièrement, nous avons crée une table temporaire groupe\_sing qui contient tous les groupes orthologues singletons avec la requête:

SELECT COUNT(genes\_edgar), groupe\_orthologue) FROM genes GROUP BY groupe\_orthologue HAVING COUNT(genes\_edgar)=1

Ensuite, nous avons exécuté une requête SELECT COUNT pour calculer le nombre de groupes orthologues par souche :

SELECT COUNT(DISTINCT groupe\_sing.groupe), genes.strainID) FROM genes, groupe\_sing, WHERE groupe\_sing.groupe=genes.groupe\_orthologue GROUP BY genes.strainID

Les résultats de cette requête sont représentés sur le plot numéro 1.

Et pour calculer le nombre des singletons que pour les Ovis la requête est la suivante :

Et pour les Equis :

SELECT COUNT(DISTINCT groupe\_sing.groupe), genes.strainID) FROM genes, groupe\_sing, strains WHERE groupe\_sing.groupe=genes.groupe\_orthologue AND genes.strainID=strains.strainID AND strains.biovar= « Ovis » GROUP BY genes.strainID

Nous avons interprété les résultats obtenus sur les plots suivants :

SELECT COUNT(DISTINCT groupe\_sing.groupe), genes.strainID) FROM genes, groupe\_sing, strains WHERE groupe\_sing.groupe=genes.groupe\_orthologue AND genes.strainID=strains.strainID AND strains.biovar= « Equi » GROUP BY genes.strainID

Venn Diagram :

Pour calculer le nombre des genes qui se trouvent dans toutes les souches nous avons utilisé la requete ci-dessous qui nous donne comme resultat 1139 genes :

SELECT COUNT(genes\_edgar), groupe\_orthologue FROM genes GROUP BY groupe\_orthologue HAVING COUNT(genes\_edgar)=55

SELECT COUNT(DISTINCT groupe\_sing.groupe), genes.strainID) FROM genes, groupe\_sing, strains WHERE groupe\_sing.groupe=genes.groupe\_orthologue AND genes.strainID=strains.strainID AND strains.biovar= « Equi » GROUP BY genes.strainID

SELECT COUNT(genes\_edgar), groupe\_orthologue) FROM genes

SELECT COUNT(DISTINCT groupe\_sing.groupe), genes.strainID) FROM genes, groupe\_sing, strains WHERE groupe\_sing.groupe=genes.groupe\_orthologue AND genes.strainID=strains.strainID AND strains.biovar= « Equi » GROUP BY genes.strainID

SELECT COUNT(genes\_edgar), groupe\_orthologue) FROM genes GROUP BY groupe\_orthologue HAVING COUNT(genes\_edgar)=1

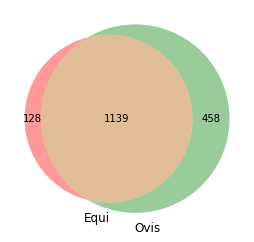
Pour calculer les nombres de gênes qui se trouvent dans toutes les souches Ovis et qui sont absentes dans toutes les souches Equis :

SELECT COUNT(genes\_edgar), groupe\_orthologue FROM genes, strains WHERE genes.strainID=strains.strainID AND strains.biovar= « Ovis » GROUP BY groupe\_orthologue HAVING COUNT(genes\_edgar)=55

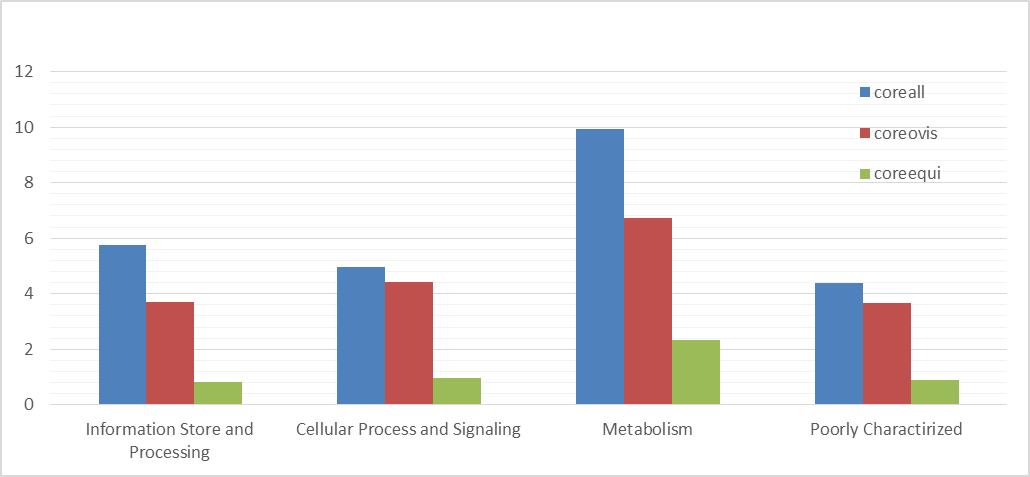
Pour calculer les nombres de gênes qui se trouvent dans toutes les souches Equi et qui sont absentes dans toutes les souches Ovis :

SELECT COUNT(genes\_edgar), groupe\_orthologue FROM genes, strains WHERE genes.strainID=strains.strainID AND strains.biovar= « Equi » GROUP BY groupe\_orthologue HAVING COUNT(genes\_edgar)=55

Nous avons représenté les résultats obtenus sur le venn diagramme ci-dessous :



Classification par les catégories fonctionnelles COG :



select count(distinct g.groupe\_orth) from groupe\_cog\_ovis g, cogs c where g.cogID=c.CogId and c.categorie IN ('General function prediction only','Function unknown') select count(distinct g.groupe\_orth) from groupe\_cog\_ovis g, cogs c where g.cogID=c.CogId and c.categorie IN ('Energy production and conversion','Amino acid transport and metabolism','Nucleotide transport and metabolism','Carbohydrate transport and metabolism','Coenzyme transport and metabolism','Lipid transport and metabolism','Inorganic ion transport and metabolism','Secondary metabolites biosynthesis, transport and catabolism')

select count(distinct g.groupe\_orth) from groupe\_cog\_ovis g, cogs c where g.cogID=c.CogId and c.categorie IN ('Cell cycle control, cell division, chromosome partitioning','Cell wall/membrane/envelope biogenesis','Cell motility','Posttranslational modification, protein turnover, chaperones','Signal transduction mechanisms','Intracellular trafficking, secretion, and vesicular transport','Defense mechanisms','Extracellular structures','Nuclear structure','Cytoskeleton') select count(distinct g.groupe\_orth) from groupe\_cog\_ovis g, cogs c where g.cogID=c.CogId and c.categorie IN ('RNA processing and modification','Chromatin structure and dynamics','Translation, ribosomal structure and biogenesis','Transcription','Replication, recombination and repair')