

Projet annotation

Bacillus amyloliquefaciens

Maxime CHAZALVIEL

Introduction

L'annotation d'un génome consiste à analyser la séquence nucléotidique pour en extraire l'information biologique. Cette analyse a pour objectifs de localiser les gènes et les régions codantes et d'identifier ou de prédire leur fonction biologique. Ces deux étapes reposent initialement sur l'utilisation d'outils algorithmiques sophistiqués, comme le développement, une partie de la bio-informatique.

Bacillus amyloliquefaciens fait partie du groupe Bacillus subtilis.

Une annotation génomique se divise en deux parties, premièrement une annotation structurale où il est nécessaire d'utiliser les différentes approches vues en cours tel que ORFfinder, GeneMark, Patscan et Artemis. Puis une annotation fonctionnelle en utilisant plusieurs logiciels nous permettant de rechercher les fonctions et les fragments transmembranaire.

Annotation structurale

L'annotation structurale consiste donc à chercher les Open Reading Frame (ORF) via ORFfinder, les séquences codantes (CDS) via GeneMark et GeneMark HMM, les Ribosomal Binding Site (RBS) via Patscan et les promoteurs et les terminateurs également via Patscan.

Sens

Orf finder	Genemark 0,5	Genemark 0,4	Genemark HMM	Recherche RBS
2402..3313	2549/2810/2948/2993 ...3313 Par zone d'intérêt 2387..3313	2402/2417/2549/2810/2948/2993 ...3313 Par zone d'intérêt 2387..3313	2402...3313	2387,2404 → <u>2402</u>
3324..4067	Par zone d'intérêt 3288 4067	Par zone d'intérêt 3288 4067	3324...4067	3313,3326 3324
4057..4818	Par zone d'intérêt 4030..4818	Par zone d'intérêt 4030..4818	4084...4818	4071,4086 → <u>4084</u>
	5208..5309	5208/5103/5208..5309	5103...5309	5090,5105 → <u>5103</u>

RBS :

Motif :GGAGG[1,0,0] 5...12 DTG

Promoteurs :

Motif : {(-22,-29,-25,16),(-18,-25,-21,16),(-28,-17,18,-11),(12,-9,-19,-6), (-2,12,-15,-7),(9,-13,-

10,-1)} > 35 16...35 {(-32,-29,-45,17),(17,-45,-45,-28),(-3,-9,-21,11),(14,-7,-19,-21), (14,-7,-19,-20),(-28,-35,-45,17)} > 35

Termineurs :

Motif : r1={AU,UA,GC,CG,GU,UG,GA,AG} p1=3...5 p2=3...10 3...9 r1~p2~p1 TTTTTT[1,0,0]

Promoteurs	Termineur
2080,2111	

Antisens

Orf finder	Genemark 0,5	Genemark 0,4	Genemark HMM	Recherche RBS
1..1386	1..1386/ 1371/1362/131 7/1209/1104	1..1386/ 1371/1362/131 7/1209/1104	1...1440	1452,1438 → <u>1440</u>
1437..2159	Par zone d'intérêt 1437..2165	1437..1598/1562 /2159	1437...2159	2171..2157 → <u>2159</u>

Genemark 0,4 : pour le second gène (1437..2159) il est nécessaire de changer d'échelle pour pouvoir le trouver.

Promoteurs	Termineur
2335,2290	

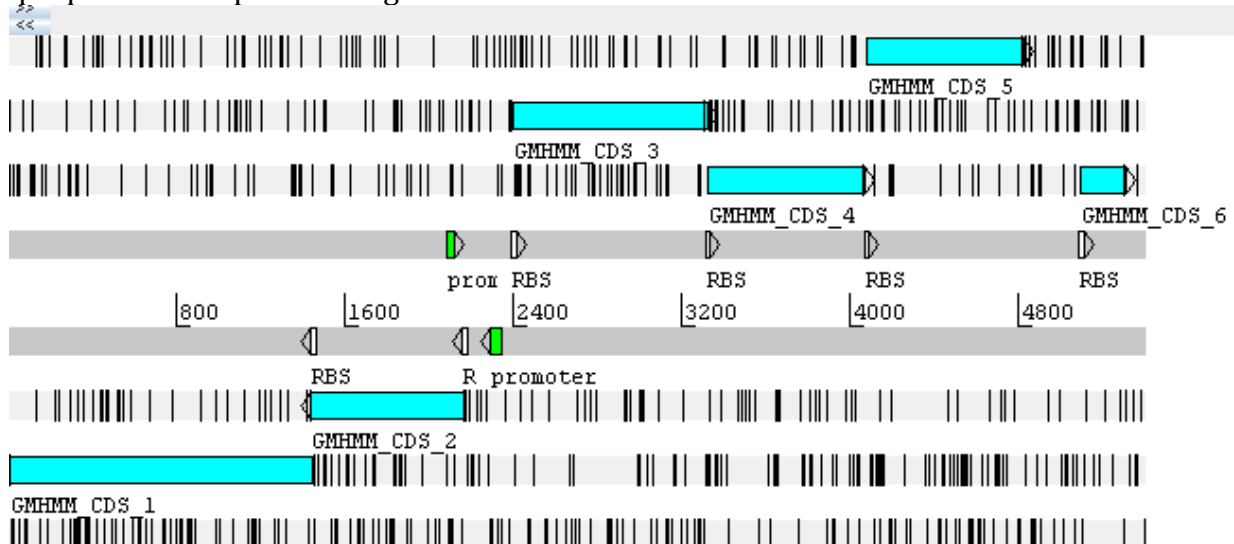
Après avoir récupéré 300 pb en plus à partir du génome entier, on peut prédire que le gène trouvé 1386-1 sur le brin indirect est peut être un gène coupé.

4081766 4087470

Orf finder trouve l'Orf 1686-301 le gène est donc complet.

Artémis

Artémis nécessite un format particulier pour annoter une séquence il a donc fallu créer des parsers pour transformer les résultats genemark, genemarkHMM et patscan en fichiers que pourra interpréter le logiciel.



Il est possible de distinguer un opérons sur chaque brin, mais il mais il n'est pas possible de dire s'ils sont complets ou s'ils continuent en dehors des limites de la séquence.

Celui sur le brin antisens pourrait être composé de deux gènes (de 2159 à 1). Patscan repère un promoteur dans la région 2335-2290.

Celui sur le brin sens serait composé de quatre gènes (de 2402 à 5103). Patscan repère un promoteur dans la région 2080,2111.

Prédiction fonctionnelle

Bacillus subtilis est une bactérie sans périplasme (gram +).

L'annotation fonctionnelle consiste donc à chercher la fonction des gènes prédits via le logiciel Blast par exemple, à savoir si le gène contient des Peptides signal via SignalP et à savoir si le gène contient des fragments transmembranaires.

Frame	Fonction	SignalP	Fragments Transmenbranaires
1_1440	Histidine kinase	Non	2
1437_2159	Histidine kinase	Non	0
2402_3313	ATP binding protein	Non	0
3324_4067	ABC-transporteur (perméase)	Non	6
4084_4818	antibiotic transport system permease	Non	<u>6</u>
5103_5309	Mercacidin precursor	Non	<u>1</u>

Conclusion

L'annotation obtenue nous permet donc de prédire six gènes, deux sur le brin antisens et quatre sur le brin sens.

Les deux premiers sont des Histidine kinase et se situent sur le brin antisens. Une s'étend de 1 à 1440 n'a pas de peptide signal et contient deux fragments transmembranaires. L'autre s'étend de 1437 à 2159 n'a pas de peptide signal et contient deux fragments transmembranaires.

Les quatre autres gènes se trouvent sur le brin sens. La première s'étend de 2402 à 3313, a pour fonction ATP binding protein n'a pas de signal peptide et ne contient pas de fragments transmembranaire. La suivante s'étend de 3324 à 4067, a pour fonction ABC-transporteur (perméase) n'a pas de signal peptide et contient six fragments transmembranaire. La troisième s'étend de 4084 à 4818, a pour fonction antibiotic transport system permease n'a pas de signal peptide et contient six fragments transmembranaire. La dernière s'étend de 5103 à 5309 a pour fonction Mercacidin precursor n'a pas de signal peptide et contient un fragment transmembranaire.